new mathematical monographs: 38

Potential Theory and Geometry on Lie Groups

N. Th. Varopoulos

Potential Theory and Geometry on Lie Groups

This book provides a complete and reasonably self-contained account of a new classification of connected Lie groups into two classes. The first part describes the use of tools from potential theory to establish the classification and to show that the analytic and algebraic approaches to the classification are equivalent. Part II covers geometric theory of the same classification and a proof that it is equivalent to the algebraic approach. Part III is a new approach to the geometric classification that requires more advanced geometric technology, namely homotopy, homology and the theory of currents. Using these methods, a more direct, but also more sophisticated, approach to the equivalence of the geometric and algebraic classification is made.

Background material is introduced gradually to familiarise readers with ideas from areas such as Lie groups, algebraic topology and probability, in particular, random walks on groups. Numerous open problems inspire students to explore further.

N. Th. Varopoulos was for many years a professor at Université de Paris VI. He is an honorary member of the Institut Universitaire de France.

NEW MATHEMATICAL MONOGRAPHS

All the titles listed below can be obtained from good booksellers or from Cambridge University Press. For a complete series listing visit www.cambridge.org/mathematics

1. M. Cabanes and M. Enguehard *Representation Theory of Finite Reductive Groups*
2. J. B. Garnett and D. E. Marshall *Harmonic Measure*
3. P. Cohn *Free Ideal Rings and Localization in General Rings*
4. E. Bombieri and W. Gubler *Heights in Diophantine Geometry*
5. Y. J. Ionin and M. S. Shrikhande *Combinatorics of Symmetric Designs*
6. S. Berhanu, P. D. Cordaro and J. Hounie *An Introduction to Involutive Structures*
7. A. Shlapentokh *Hilbert's Tenth Problem*
8. G. Michler *Theory of Finite Simple Groups I*
9. A. Baker and G. Wüstholz *Logarithmic Forms and Diophantine Geometry*
10. P. Kronheimer and T. Mrowka *Monopoles and Three-Manifolds*
11. B. Bekka, P. de la Harpe and A. Valette *Kazhdan's Property (T)*
12. J. Neisendorfer *Algebraic Methods in Unstable Homotopy Theory*
13. M. Grandis *Directed Algebraic Topology*
14. G. Michler *Theory of Finite Simple Groups II*
15. R. Schertz *Complex Multiplication*
16. S. Bloch *Lectures on Algebraic Cycles (2nd Edition)*
17. B. Conrad, O. Gabber and G. Prasad *Pseudo-reductive Groups*
18. T. Downarowicz *Entropy in Dynamical Systems*
19. C. Simpson *Homotopy Theory of Higher Categories*
20. E. Fricain and J. Mashreghi *The Theory of H(b) Spaces I*
21. E. Fricain and J. Mashreghi *The Theory of H(b) Spaces II*
22. J. Goubault-Larrecq *Non-Hausdorff Topology and Domain Theory*
23. J. Śniatycki *Differential Geometry of Singular Spaces and Reduction of Symmetry*
24. E. Riehl *Categorical Homotopy Theory*
25. B. A. Munson and I. Volić *Cubical Homotopy Theory*
26. B. Conrad, O. Gabber and G. Prasad *Pseudo-reductive Groups (2nd Edition)*
27. J. Heinonen, P. Koskela, N. Shanmugalingam and J. T. Tyson *Sobolev Spaces on Metric Measure Spaces*
28. Y.-G. Oh *Symplectic Topology and Floer Homology I*
29. Y.-G. Oh *Symplectic Topology and Floer Homology II*
30. A. Bobrowski *Convergence of One-Parameter Operator Semigroups*
31. K. Costello and O. Gwilliam *Factorization Algebras in Quantum Field Theory I*
32. J.-H. Evertse and K. Győry *Discriminant Equations in Diophantine Number Theory*
33. G. Friedman *Singular Intersection Homology*
34. S. Schwede *Global Homotopy Theory*
35. M. Dickmann, N. Schwartz and M. Tressl *Spectral Spaces*
36. A. Baernstein II *Symmetrization in Analysis*
37. A. Defant, D. Garcia, M. Maestre and P. Sevilla-Peris *Dirichlet Series and Holomorphic Functions in High Dimensions*
38. N. Th. Varopoulos *Potential Theory and Geometry on Lie Groups*
39. D. Arnal and B. Currey *Representations of Solvable Lie Groups*

Potential Theory and Geometry on Lie Groups

N. TH. VAROPOULOS
Université de Paris VI (Pierre et Marie Curie)

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi–110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107036499
DOI: 10.1017/9781139567718

First published 2021

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Names: Varopoulos, N., 1940– author.
Title: Potential theory and geometry on Lie groups / N. Th. Varopoulos.
Description: Cambridge ; New York, NY : Cambridge University Press, 2020. |
Series: New mathematical monographs ; 38 |
Includes bibliographical references and index.
Identifiers: LCCN 2019038658 | ISBN 9781107036499 (hardback)
Subjects: LCSH: Lie groups. | Potential theory (Mathematics) | Geometry.
Classification: LCC QA387 .V365 2020 | DDC 512/.482–dc23
LC record available at https://lccn.loc.gov/2019038658

ISBN 978-1-107-03649-9 Hardback

To some of the women I have loved

Contents

Preface

This book is both a synthesis and an extensive elaboration of a series of papers that I wrote on the subject in the years 1994–2000. These papers were highly interconnected and this was one of the reasons why they were very difficult to read. I must also admit that at the time I did not make the necessary effort to make these papers more reader friendly.

An effort to write down the early part of this work (essentially Chapters 2 and 3) in book form was undertaken by me and a co-author in the late 1990s. But this failed. To describe that project as a joint enterprise is not really accurate because I expected the co-author to do all the hard work and the writing. This of course was not a good idea in principle, and the fact that I had got away with exactly this strategy a decade earlier with a previous book was no excuse. At that time, my co-authors were young and motivated, they did all the hard work and the book that came out was by all accounts a success. Anyway, to try to make the same 'trick' work a second time reflected very poor judgement on my part.

Having said all this I should also add that since then the feedback I got from colleagues who were interested in the subject and who tried to read my 1994–2000 papers was that they found them impossible to read.

It was for these reasons and to expiate my past sins that in the spring of 2011 I embarked on a new project to write a book that would be useful and accessible to students in Lie theory and even to try to make it 'easy reading'.

The 'easy reading' part of the project, however, very quickly turned out to be wishful thinking.

The reason lies in the nature of the subject, which relies on a tremendous amount of background material from many distinct branches of mathematics. This very rapidly became my main preoccupation in the writing of the book. To recall what was needed separately in detail was not an option since that would have expanded the size of the book out of proportion.

To resolve this problem I have resorted to placing this background material into layers that intertwine gradually with the theory as it is developed in the book. These 'layers' are organised in each of the three parts and even in each chapter of the book, and they become progressively more sophisticated as we go along. So the reader could read the first layer of a certain topic, which should be relatively easy reading and which also explains the underlying ideas, and then move to another topic. Extensive indications and 'guides for the reader' are given in the text as to how one could navigate in that respect. The aim of it all is to try to minimise difficulties for the reader.

At any rate, as always happens with 'past sins', the writing of this book turned out to be much harder than I originally thought. In my effort, three people have helped me and it is my pleasure and duty to thank them here.

First and foremost I want to thank David Tranah, the mathematics editor at Cambridge University Press. It is not an exaggeration to say that without his assistance this book would not have been written. At a technical level, he went so far as to finish the TEX himself. But what I am most grateful to him for is the time and effort he put into teaching me how books should be written and where the difference lies between a book and a series of papers that nobody can read.

David made several suggestions to improve the presentation, he read the whole book and made many direct corrections on my manuscript. I also discovered that David knew a lot of mathematics as to my surprise (and also vexation), he even made mathematical corrections!

The world has changed and in the domain of scientific publications very many old and prestigious editorial institutions, having bought each other (!), have then turned themselves into money-making supermarkets. It is therefore refreshing and encouraging to find an exception like Cambridge University Press and people like David Tranah in its staff. I also want to thank Clare Dennison at Cambridge University Press and Alison Durham for the excellent job they did with the editing and preparation of the manuscript. I am very grateful to them for their kindness and for their patience with my incompetence in the final stage of the process.

Next, but also very importantly, I wish to thank my friend and colleague Leo di Michele. In the early 2000s, Leo was responsible for the setting up of the mathematics department of a new university that was then starting in Italy: the Bicocca Milano University. That department soon became a wonderful place to be.

My French mathematical career was then coming to a close and Leo invited me to join Bicocca. I found there a congenial and friendly atmosphere that contrasted very much with the cold isolation that I'd become used to.

My stay in Bicocca breathed new life into me and helped me embark on this project. For all this and for his generous friendship, I wish to express here my gratitude to Leo.

Last but not least it is my pleasure to thank Ivan Kupka who is a friend and a colleague from Université Paris VI. Kupka is a distinguished geometer and without his patient and constant help I would not have been able to understand the geometry that I needed to write Parts II and III of the book.

Learning that geometry was one bright spot in the long and austere task of writing this book. It was like an old man who at the end of his life falls in love with a young girl – if that happens to you my advice is not to ask the young lady how she feels about it. And for the same reason I feel very nervous about my geometric 'performance' in the book!

From my point of view however, whether I was any good in my geometry or not, I enjoyed the experience very much.

Paris

1
Introduction

In this chapter we shall describe in *informal terms* what the book is all about. Furthermore, we shall introduce a number of concepts, notation and conventions that will stay throughout the book.

1.1 Distance and Volume Growth

Distance Unless otherwise stated, all the locally compact groups that will be considered will be compactly generated. This means that there exists $e \in \Omega = \Omega^{-1} \subset G$, some compact symmetric neighbourhood of the identity $e \in G$ such that $\bigcup_{n \geqslant 1} \Omega^n = G$.

When $G = \Gamma$ is a discrete group this means that there exists a finite symmetric set of generators $(\gamma_1, \dots, \gamma_n) \subset \Gamma$. A natural distance is then associated to these generators. For $x \in \Gamma$ we write $x = \gamma_{i_1} \cdots \gamma_{i_k}$ where k is assumed as small as possible. We then set $|x| = k$ and $|e| = 0$ and define the distance $d(x,y) = |x^{-1}y|$. This distance is left invariant and $d(gx, gy) = d(x,y)$, for x, $y, g \in \Gamma$, and is referred to by the 'word distance'. The same definition extends to locally compact groups and $|x| = k$ with k the smallest integer for which $x \in \Omega^k$.

The Haar measure and the volume growth Haar measures on G are positive Radon measures that are invariant under group translation (see Bourbaki, 1963, Chapter 7). For a discrete group Γ the volume growth is defined to be

$$\gamma(n) = [\text{number of elements } x \in \Gamma \text{ with } |x| \leqslant n].$$

For a general locally compact group G,

$$\gamma(n) = \text{ Haar measure } [x \in G;\ |x| \leqslant n] = \text{ Haar measure } [B_n(e)].$$

Two Haar measures exist on G: the left-invariant $dg = d^{\mathrm{l}}g$, with $g \in G$, and right-invariant $d^{\mathrm{r}}g$ for which, respectively, $d^{\mathrm{l}}(xg) = d^{\mathrm{l}}g$ and $d^{\mathrm{r}}(gx) = d^{\mathrm{r}}g$, for $x, g \in G$. Both are unique up to a multiplicative constant and we shall normalise throughout by setting $d^{\mathrm{r}}(x^{-1}) = d^{\mathrm{l}}(x)$. The locally compact groups for which, up to a multiplicative constant, $d^{\mathrm{r}}g = d^{\mathrm{l}}g$ are called unimodular. The modular function is defined by $d^{\mathrm{r}}g = m(g)\,d^{\mathrm{l}}g$. Since $d^{\mathrm{r}}(x^{-1}) = d^{\mathrm{l}}x$ and since Ω is symmetric (i.e. stable under $x \mapsto x^{-1}$), it is irrelevant in the definition of $\gamma(n)$ whether we use the left or the right Haar measure. The definition of $\gamma(n)$ depends on the choice of Ω but it does so in an inessential way in the sense that if we change Ω to a new Ω there exist constants $C > 0$ such that

$$\gamma_{\mathrm{new}}(n) \leqslant C\gamma_{\mathrm{old}}(Cn) + C; \quad n \geqslant 0.$$

Convention on the letters C, c In this book we shall use systematically the letters C and c, possibly with indices $C_1, \ldots$, to indicate *positive* constants that are independent of the main parameters of the formula where they occur. These constants may differ from place to place even in the same formula. For $P, Q > 0$ we shall also write $P \lesssim Q$ for $P \leqslant CQ$, and write $P \sim Q$ when both $P \lesssim Q$ and $Q \lesssim P$ hold.

Notes and references A general reference on measure theory that will be used is Bourbaki (1963). The classical references for locally compact groups and the Haar measure are Pontrjagin (1939) and Weil (1953).

For the invariant distance on a locally compact group see Gromov (1981), Varopoulos et al. (1992).

1.2 A Classification of Unimodular Locally Compact Groups

The following theorem is contained in Gromov (1981). Let Γ be some discrete finitely generated group. Then Γ belongs exactly to one of two categories:

(α) Groups of polynomial volume growth. For these groups there exists C such that $\gamma(n) \leqslant Cn^C + 1$. The group Γ is of polynomial volume growth if and only if there exists $\Gamma_1 \subset \Gamma$, a subgroup that is nilpotent and such that the index $[\Gamma, \Gamma_1] < +\infty$ is finite.

(β) The groups that are not of polynomial volume growth. For all $A > 1$, for these groups there exists C such that $\gamma(n) \geqslant Cn^A$ for $n \geqslant 1$.

In these two statements the following standard practice is used; whenever we say 'there exists C or $c \ldots$', or 'for some constant C, $c, \ldots$', we always mean *positive* constants $C, c, \ldots$. This will be done throughout the book.

Another way of measuring the growth at infinity of Γ is to consider $\mu \in \mathbb{P}(\Gamma)$, some finitely supported probability measure that charges a set of generators $\gamma_1, \dots, \gamma_n$, and is symmetric, that is, $\mu(\gamma) = \mu(\gamma^{-1})$, $\gamma \in \Gamma$ and $\mu(\gamma_j) > 0$, $j = 1, \dots, n$ and $\mu(e) > 0$. We shall define $\mu^{*n} = \mu * \cdots * \mu$, the nth convolution power; what will be relevant will be $\mu^{*n}(e)$. Among the groups Γ we can again consider two categories:

(A) There exist $C, c > 0$ such that $\mu^{*n}(e) \geqslant cn^{-C}$, $n \geqslant 1$.
(B) For all $A > 0$ there exists $C > 0$ such that $\mu^{*n}(e) \leqslant Cn^{-A}$, $n \geqslant 1$.

Here it is less clear how the asymptotics of $\mu^{*n}(e)$ behave under a change of μ to a new $\mu_1 \in \mathbb{P}(G)$ and whether the above is a classification that is independent of a change of μ.

However, it is the case that the above two classifications are identical and for any discrete group Γ, $(\alpha) \Leftrightarrow (A)$, $(\beta) \Leftrightarrow (B)$.

The symmetry of μ is essential for the above. To see this consider $\Gamma = \mathbb{Z}$, $\mu(0) = \varepsilon$, $\mu(1) = 1 - \varepsilon$ for some small $\varepsilon > 0$.

The above generalise verbatim to unimodular locally compact groups, that is, we have the (α)–(β) classification determined by the volume growth and the (A)–(B) classification that depends on $\mu^{*n}(e)$, which is now interpreted to be $\phi_n(e)$ with $d\mu^{*n}(g) = \phi_n(g)\,dg$, where μ is assumed to be symmetric compactly supported with $\phi_1(g)$ continuous. Here again we have the equivalence $(\alpha) \Leftrightarrow (A)$, $(\beta) \Leftrightarrow (B)$.

Notes and references Hall (1959) is a good reference for the notion of nilpotency of discrete groups. Feller (1968) or Woess (2000) are good references for $\mu^{*n}(e)$ which is the return probability of the random walk $(z(n) \in \Gamma;\ n \geqslant 1)$ with $\mathbb{P}[z(n+1) = \gamma_1,\ z(n) = \gamma_2] = \mu(\gamma_1^{-1}\gamma_2)$.

The classifications (α), (β), (A), (B) and their equivalences can be found in Gromov (1981), Varopoulos et al. (1992).

1.3 Lie Groups

1.3.1 Convolution powers of measures

Let G be some locally compact group. We shall consider throughout $\mu \in \mathbb{P}(G)$ that are symmetric (i.e. stable by the involution $g \to g^{-1}\colon\ \mu(g) = \mu(g^{-1})$) and of the form $d\mu(g) = \phi(g)\,d^r g$ where ϕ is continuous and compactly supported. We shall denote in general $d\mu^{*n}(g) = \phi_n(g)\,d^r g$ and assume that $\phi(e) \neq 0$. Here and throughout we adopt the (fairly) standard notation $\mathbb{P}(X)$ to indicate the space of probability measures on X.

To fix ideas let us assume that G is a connected real Lie group, that is, at the other end of the 'spectrum' from the discrete groups that we considered above. For such a group we shall now state the following classification theorem.

Theorem *Let G be some real connected Lie group. Then G satisfies exactly one of the following alternatives:*

(B) *For every $\mu \in \mathbb{P}(G)$ as above there exist $\lambda \geqslant 0$ and $C_1, C_2, c_1, c_2 > 0$ such that*

$$C_2 \exp(-\lambda n - c_2 n^{1/3}) \leqslant \phi_n(e) \leqslant C_1 \exp(-\lambda n - c_1 n^{1/3}); \quad n \geqslant 1.$$

(NB) (Non-B). *For every $\mu \in \mathbb{P}(G)$ as above there exist $\lambda, \alpha \geqslant 0$ and $C_1, C_2 > 0$ such that*

$$C_2 n^{-\alpha} e^{-\lambda n} \leqslant \phi_n(e) \leqslant C_1 n^{-\alpha} e^{-\lambda n}; \quad n \geqslant 1.$$

The parameter λ depends on μ: but either $\lambda = 0$ for all μ as above and then we say that G is amenable; or $\lambda > 0$ for all μ and then we say that G is not amenable (see Reiter, 1968; Greenleaf, 1969; and §3.1 and (4.35)). Typical non-amenable groups are the semisimple groups of non-compact type. These groups are (NB) by Bougerol (1981). For the definitions on Lie groups see Varadarajan (1974). This classification for unimodular amenable groups is contained in Varopoulos et al. (1992, Chapter VII). If we drop the condition of unimodularity this classification becomes much more subtle and, surprisingly, the exponent α in (NB) varies continuously with μ in general. In 'general', α can then take any value $\alpha \geqslant \alpha_0(G)$.

The proof of the above analytic fact and the algebraic and geometric conditions on G that determine the B–NB classification are, to a large extent, the subject matter of this book.

More general locally compact groups can be classified as above as long as they are connected. Connectedness here is taken in the sense of the locally compact topology of G. But connectedness (or rather irreducibility) in the sense of algebraic geometry can also be used and the classification still persists for say algebraic subgroups of $\mathrm{GL}(V)$ over a field that is a finite extension of the field of p-adic numbers. As we shall see, the proofs for these algebraic groups are but easy modifications of the ones given for real Lie groups. These algebraic groups are a sideshow and the reader can ignore this aspect without missing much. Notice, however, that these algebraic groups give examples of totally disconnected locally compact groups that are not necessarily compactly generated and of where the theory applies.

1.3.2 The heat diffusion semigroup

The B–NB classification can be seen from another point of view that is closer to the one adopted in Varopoulos et al. (1992). We say that X, a vector field on G (i.e. a first-order differential operator), is left invariant if $Xf_g = (Xf)_g$, where $f_g(x) = f(gx)$, $f \in C_0^\infty(G)$. We consider then such invariant fields $X_1, X_2, \dots, X_p$, which together with their successive brackets $[X_i, X_j]$, $[[X_i, X_j], X_k]$ span $T_e(G)$, the tangent space at e. This condition on the brackets is called the Hörmander condition; see Hörmander (1967), Varopoulos et al. (1992). We recall the definition of the bracket which is the first-order differential operator $[X_i, X_j] = X_iX_j - X_jX_i$. We denote by $\triangle = -\sum X_j^2$ the corresponding sub-Laplacian which is a second-order subelliptic operator. If $X_1, \dots, X_p$ is a basis of $T_e(G)$ at e then $\triangle$ is in fact elliptic and the reader could think of it in these terms. This clearly generalises the standard Laplacian $-\sum \frac{\partial^2}{\partial x_i^2}$ in $\mathbb{R}^d$.

We can then appropriately close $\triangle$ and obtain a self-adjoint positive operator on $L^2(G, d^r g)$ and define the corresponding semigroup $T_t = e^{-t\triangle}$. The following facts are then well known: $T_t f = f * \mu_t$, $f \in C_0^\infty$ where $\mu_t \in \mathbb{P}(G)$, $t > 0$, which satisfy $\mu_t(g) = \mu_t(g^{-1})$, $\mu_t * \mu_s = \mu_{t+s}$, $d\mu_t(g) = \phi_t(g)\, d^r g$ for some $\phi_t \in C^\infty(G)$. We have thus the continuous-time analogue of the convolution powers of a measure of §1.3.1. The only difference is that the measures μ_t are not, in general, compactly supported. They satisfy instead a Gaussian decay at infinity. To wit, if we denote

$$E_t(r) = \mu_t(G \backslash B_r); \quad B_r = [g \in G;\ |g| \leqslant r],$$

then $E_t(r) \leqslant C \exp(-cr^2)$ where the constants depend on t; see §2.12 or Varopoulos et al. (1992) for more details.

Having defined $d\mu_t(g) = \phi_t\, d^r g$ as above, the B–NB classification of G can then be given by inserting $\phi_t(e)$ in (B), (NB) in §1.3.1 for $t = 1, 2, \dots$.

This variant of the B–NB classification is important because the semigroup T_t is in many ways more flexible than the convolution powers of a measure. This semigroup can be used, for instance, to define the negative powers of the Laplacian

$$\triangle^{-\alpha/2} = c_\alpha \int_0^\infty t^{-\alpha/2} T_t\, dt; \quad \alpha > 0.$$

Using this, Hardy–Littlewood–Sobolev (HLS) estimates and isoperimetric inequalities can be considered on G. These amount to analysing the boundedness properties of

$$\triangle^{-\alpha/2} \colon L^p(G) \to L^q(G); \quad \alpha > 0,\ 1 \leqslant p,\ q \leqslant \infty. \tag{1.1}$$

These HLS estimates and the Gaussian decay of μ_t played a central role in

Varopoulos et al. (1992). The classification in §I.2 of that book was in fact done via these HLS estimates. This is no longer the case in non-unimodular groups where only very few HLS results exist and we have many open problems but few theorems about (1.1) (cf. Varopoulos, 1996a).

Gaussian decay when the group is amenable can be illustrated by the following sharp estimate (see Varopoulos, 2000a):

$$E_t(r) \leqslant C \exp\left(-\frac{r^2}{ct}\right); \quad r, t \geqslant C,$$

where the constants depend only on G and $\triangle$. The same estimate holds for the convolution powers of a measure. This type of result is interesting but will not be considered in this book.

Notes and references The essential prerequisite for reading this book is a working knowledge of Lie group theory. There are several excellent books on the subject. The one that I shall follow very closely in results, spirit and notation is Varadarajan (1974). Another reference that I shall follow closely is Helgason (1978). A beginner in the subject will find Sagle and Walde (1973) helpful.

Amenability is a subject in itself and we shall refer to Reiter (1968) and Greenleaf (1969), but in fact everything there that will be needed will be recalled in §3.1. A standard general reference for connected locally compact groups is Montgomery and Zippin (1955).

For the little use that we shall make of algebraic groups we shall use the historical references Chevalley (1951, 1955). For our purposes these remain the best sources.

1.4 The Geometric B–NB Classification of Lie Groups. An Example

For the (A)–(B) classification of unimodular Lie groups we used $\gamma(n)$, the volume growth of G. For non-unimodular groups on the other hand it is well known (see Varopoulos et al., 1992, §IX.1), and very easy to see, that $\gamma(n) \geqslant ce^{cn}$, $n \geqslant 1$, and therefore something more subtle has to be done to express geometrically the B–NB classification of the previous section. We have to consider volumes of sets with respect to lower-dimensional Hausdorff measures. For these we shall need to introduce the following notions.

1.4.1 Invariant Riemannian structures on G and quasi-isometries

Let G be some connected real Lie group. On $T_e(G)$, the tangent space at the identity, we can fix some scalar product, that is, some Euclidean structure, and we can then use the left group translation $g \to xg$ to obtain the corresponding inner product structure on each $T_g(G)$, $g \in G$. These are the left-invariant Riemannian structures on G and two different inner products on $T_e(G)$ give rise to quasi-isometric Riemannian structures in the following sense.

Let $f\colon M_1 \to M_2$ be some diffeomorphism between two Riemannian manifolds. We then say that f is a quasi-isometry if there exists C such that $|df|$, $|df^{-1}| \leqslant C$.

For Riemannian manifolds M we shall need to recall the following two notions.

The Hausdorff measure of $E \subset M$ is denoted $\mathrm{Vol}_\alpha(E)$, $0 < \alpha \leqslant d = \dim M$. For $\alpha = d$ this is the Riemannian volume of E. For $\alpha = d-1$ this is the surface area of a hypersurface. For $\alpha = 1$ this is the length of a one-dimensional curve in E. Federer (1969) is a good reference.

Closely related to this is the notion of Lipschitz functions. Let $f\colon M_1 \to M_2$ be a continuous mapping between two Riemannian manifolds; we then say that $\mathrm{Lip}\, f \leqslant R$ if

$$d_2(f(x_1), f(x_2)) \leqslant R d_1(x_1, x_2); \quad x_1, x_2 \in M_1$$

for the two Riemannian distances d_1, d_2 on M_1, M_2.

1.4.2 An important example

The example that will guide us in the geometric considerations that follow is A, the group of affine motions $x \mapsto ax + b, x \in \mathbb{R}$. Here $a > 0, b \in \mathbb{R}$ and A can be identified with the upper half-plane $H = [(a,b);\ a > 0]$. The Poincaré metric $ds^2 = a^{-2}(da^2 + db^2)$ can then be assigned on H and this, identified with a metric on A, is a left-invariant metric on A. The space H with the above metric is a very important structure and it realises the non-Euclidean (Lobatchevski) geometry of the plane. The geodesics (i.e. the straight lines) for that geometry are the circles in H that are orthogonal to the boundary. For readers familiar with elementary differential geometry this is a simply connected negatively curved manifold. These manifolds are called Cartan–Hadamard manifolds (see Cheeger and Ebin, 1975; Helgason, 1978, Chapter 1). By conformal mapping, H is also the unit disc on which much complex analysis happens, and A is also the simplest non-Abelian Lie group (and the only one in dimension 2).

What counts on H is that two points $X_1, X_2 \in H$ can be joined by a unique geodesic and, exactly as in the Euclidean plane, we can use these geodesics to make a retract of H to some fixed point $O \in H$. This is done by the mapping $h(X,t) \in H, X \in H, 0 \leqslant t \leqslant 1$, which is defined by $h(X,1) = X, h(X,0) = O$ and $h(X,t)$ is the point on the geodesic that joins O to X at distance $d(O,h(X,t)) = td(O,X)$.

The following facts can easily be verified (either by elementary geometry – and then it is tedious – or by the general theory of Jacobi fields on the geodesics – and then it is automatic).

1.4.3 Isoperimetric inequalities

Let $\Gamma \subset H$ be some closed smooth curve and let $O \in \Gamma$ be some fixed point. Let us use the above retract, with centre O, and set $D = [h(X,t);\ X \in \Gamma,\ 0 \leqslant t \leqslant 1]$. Here D is the image in H: that is, $D = \alpha(e)$ where $e = e_2$ is the topological 2-cell and $\alpha\colon e \to H$ is a mapping with $\alpha(\dot{e}) = \Gamma$. Here $\dot{e} = S^1$ is the unit sphere which is the boundary of e. We also have

$$\mathrm{Vol}_2(D) \leqslant C(\mathrm{Vol}_1 \Gamma + 1) \tag{1.2}$$

for the corresponding Hausdorff measures. Incidentally, the same type of isometric inequality holds in the Euclidean plane $\mathbb{R}^2$ but there the right-hand side has to be replaced by $(\mathrm{Vol}_1 \Gamma + 1)^2$ instead.

The above motivates the following general definition.

Let M be some Riemannian manifold that is topologically $\cong \mathbb{R}^d$. We shall denote by e_n the topological n-cell (Euclidean n-ball) for $2 \leqslant n \leqslant d$; then $\dot{e}_n = S^{n-1}$, will denote its boundary.

We could say that M has the volume polynomial filling property if there exists a constant C such that for every smooth $\alpha\colon \dot{e}_n \to M$ we can extend $\hat{\alpha}\colon e_n \to M$ smoothly and in such a way that $\mathrm{Vol}_n\, \hat{\alpha}(e_n) \leqslant C[\mathrm{Vol}_{n-1}\, \alpha(\dot{e}_n) + 1]^C$. The inequality (1.2) then says that H admits this polynomial filling property.

One problem with using the above definition is that it is complicated to extend it to $n > d$ and also, clearly, the various volumes have to be counted with ‘multiplicity’ if the definition is to be coherent. It is preferable therefore to consider a closely related notion which in our context of Lie groups is essentially equivalent.

The above notions were introduced in Gromov (1991), where I learned about them.

1.4.4 The polynomial filling property

Rather than topological cells, it will be preferable here to consider $[0,1]^n \subset \mathbb{R}^n$, the unit cube; then $\partial^{n-1} = \partial[0,1]^n \subset \mathbb{R}^n$ denotes its topological boundary and they are both assigned their Euclidean distance. Going back to H and using the retract h we see that exactly the same considerations (either elementary or using the properties of negative curvature) show the following.

Theorem *Let $R > 0$ and let $f\colon \partial^1 \to H$ be such that $\operatorname{Lip} f \leqslant R$. Then we can extend f to $\hat{f}\colon [0,1]^2 \to H$ in such a way that $\hat{f}\big|_{\partial^1} = f$ and $\operatorname{Lip} \hat{f} \leqslant C(R+1)^2$.*

More generally, again for any Riemannian manifold M that is topologically $\cong \mathbb{R}^d$ and any $n \geqslant 2$, we say that M has the polynomial filling property in dimension n if there exists $C > 0$ such that, for all $R \geqslant 1$ and all $f\colon \partial^{n-1} \to M$ with $\operatorname{Lip} f \leqslant R$, we can find an extension $\hat{f}\colon [0,1]^n \to M$ such that

$$\hat{f}\big|_{\partial^{n-1}} = f, \quad \operatorname{Lip} \hat{f} \leqslant CR^C. \tag{1.3}$$

We consider only $R > 1$ because we are interested only in the behaviour of the manifold far out at infinity. For brevity we shall say that M satisfies the PFP if this holds in every dimension.

1.5 A Special Class of Groups and the Metric Classification

This special class, which will be called *models*, are the Lie groups U, which, like the group of affine motions A of the previous section, are diffeomorphic with some Euclidean space $E = \mathbb{R}^a$; that is, we have a diffeomorphism $U \simeq E$. It is clear that this topological property is essential if the notion of the polynomial filling property (PFP for short) of §1.4.4 is to be used in a systematic way.

We shall explain in this section how these groups can be used as the building blocks of the geometric theory. But before that, for those readers who know what the terminology means, we should point out that these models can be characterised algebraically by the fact that U is a model if and only if U is a simply connected soluble group.

1.5.1 The geometric classification for models

One of the basic results in the geometric theory is the following.

Theorem *Let U be a model. Then,*

(i) *if U is an* NB*-group in the sense of* §*1.3.1, U admits the polynomial filling property*;
(ii) *if U is a* B*-group, U does not admit the PFP.*

The second part of this theorem is difficult to prove and it takes essentially Chapters 8–10 to do so. On the other hand, if we restrict ourselves to models this is a very satisfactory way to express the B–NB classification theorem of §1.3.1 in geometric terms.

In the remainder of this section we shall explain how this theorem can be used to give the general geometric (or 'metric') classification of any connected Lie group.

1.5.2 The coarse quasi-isometries

This notion refers to general metric spaces (M,d) with a distance function: $d(m_1,m_2) \geqslant 0$; $m_1,m_2 \in M$.

A mapping $\alpha \colon M_1 \to M_2$ between two such metric spaces with distances d_1, d_2 will be called a *coarse quasi-contraction* if there exist constants C such that

$$d_2(\alpha(m),\alpha(m')) \leqslant C d_1(m,m') + C; \quad m,m' \in M_1.$$

We then say that the metric spaces M_1, M_2 are *coarse quasi-isometric* if there exist two coarse quasi-contractions $M_1 \underset{\beta}{\overset{\alpha}{\rightleftarrows}} M_2$ and $C \geqslant 0$ such that

$$d_1(m_1,\beta\circ\alpha(m_1)) \leqslant C, \quad d_2(m_2,\alpha\circ\beta(m_2)) \leqslant C; \quad m_1 \in M_1,\ m_2 \in M_2.$$

This means that α and β are almost inverses of each other.

The first example that comes to mind is $\alpha \colon \mathbb{Z} \to \mathbb{R}$, the natural integers and the real line, with their natural distance. This coarse quasi-isometry is clearly an equivalence relation and, here at least, we shall use the notation $M_1 \sim M_2$ (as good notation as any!).

Given what we have already done, a more significant example of coarse quasi-isometries occurs when we consider in §1.1 two different distances on the same compactly generated locally compact group that are given by two different neighbourhoods Ω_1, Ω_2 of the identity, as explained there.

1.5.3 Coarse quasi-isometric models

The next result has very little to do with Lie groups and the proof consists of easy manipulation of the kind one finds in elementary (and old-fashioned) homotopy theory. At any rate what this says is the following.

Theorem *Let U_1, U_2 be two models assigned with their group distance and let us assume that they are coarse quasi-isometric*: $U_1 \sim U_2$. *Then if one of two models admits the PFP, so does the other.*

1.5.4 A general connected Lie group

Now let G be some connected group. Then using classical structure theorems on Lie groups (some of which are quite deep), we can find some model U such that $G \sim U$. Furthermore, if G is a B-group (resp. NB-group), the model U can be chosen to be a B-group (resp. NB-group).

1.5.5 The general metric B–NB classification

If we put together §§1.5.1–1.5.4 we see that, within our framework at least, we have achieved our aim. Let us be more explicit but slightly informal.

We start from some connected Lie group G. To decide geometrically whether G is B or NB we do this:

First find some model U such that $G \sim U$. Then check if U admits the PFP. If it does then G is an NB-*group. If not then G is a* B-*group*.

Furthermore, the above criterion easily extends to general connected locally compact groups. It proves that if two such groups are coarse quasi-isometric, then one is B (resp. NB) if the other is. Such groups, for our purposes, can in fact be approximated by Lie groups (see Chapter 6).

1.5.6 The drawback of this metric classification

The main reason why we do not stop here and instead go on to give additional geometric classifications is the following. In the above we have mixed two notions that do not mix well. The first is PFP which is a continuous notion that involves homotopy theory. The second is the 'discontinuous' (almost discrete, one could say) notion of coarse quasi-isometries.

The metric theory that leads to the criterion of §1.5.5 is developed in Part II of the book (Chapters 7–11). In the final Part III we develop another geometric B–NB criterion that is interesting in its own right, but which also gets round the drawback that we have just pointed out. This new classification only uses the continuous notion of homotopy retract that will be explained in the next section.

1.6 Homotopy Retracts

1.6.1 The classical retract

It is a classical and deep (and difficult) theorem in the structure theory of Lie groups (see Hochschild, 1965) that the following happens:

Let G be some connected Lie group. Then there exists some smooth function $H(g,t) \in G$*, with* $g \in G$, $0 \leqslant t \leqslant 1$*, and some compact subgroup* $K \subset G$*, with the following properties*:

$$H(g,1) = g,\ H(g,0) \in K,\ H(k,t) = k; \quad g \in G,\ k \in K,\ 0 \leqslant t \leqslant 1.$$

The function $H(g,t)$ is called a *homotopy retract* of G onto the compact subgroup K that can be taken to be a maximal compact subgroup, but this additional information is not relevant here. In fact, more can be said and we actually have a diffeomorphism $G \simeq K \times \mathbb{R}^a$ for some $a = 0, 1, \ldots$.

1.6.2 The polynomial retract

We shall say that the connected Lie group G has the *polynomial retract property* if the homotopy $H(g,t)$ of the previous subsection can be chosen to satisfy the following polynomial bound for its gradient:

$$|dH(g,t)| \leqslant C(1+|g|)^C; \quad g \in G,\ 0 \leqslant t \leqslant 1,$$

for appropriate constants.

In informal terms, this says that the speed with which we collapse G to its compact subgroup is at most polynomial.

Example The identity is the only compact subgroup of the group A of affine motions that we considered in §1.4.2. (This is in fact the case *for every model*.) The above homotopy therefore retracts A to a point (i.e. the space is retractible). The considerations of §1.4.2 (whether Jacobi fields are used or elementary means) in effect amount to saying that A has the polynomial retract property.

1.6.3 The polynomial retract property used in the B–NB classification

The notion is natural and from our point of view it can be used as follows.

Theorem (B–NB; Ht) *Let G be some connected Lie group. Then,*

(i) *if G is an* NB*-group, G admits the polynomial retract property*;

(ii) *if G is a* B*-group, G does not admit the polynomial retract property.*

Another notion of polynomial retract (less standard in topology; cf. §12.2.2) can be given by requiring that, together with the polynomial estimate of §1.6.2, the homotopy is such that $H(g,1) = g$ and $H(g,0) \in P$ lies in *some* compact set $P \subset G$. *This definition of polynomial retract can be substituted in the above theorem and the theorem still holds good*. As pointed out in §1.6.1, if we do not impose the polynomial estimate, all Lie groups can be retracted to a compact set in this sense. Notice, however, that only a very restricted class of manifolds admit this general property.

1.6.4 The investment/return ratio; or what it takes to prove the (B–NB; Ht) theorem

The proof of part (i) of the theorem is given in Appendix F but it would be fairer to say that it uses bits and pieces from all over Parts I and II of the book.

The proof of part (ii) is an entirely different story and for that one has to use the homology theory that will be developed in Part III.

This brings us to the dilemma that the author has had to face, namely, a decision about the 'return on investment' that, I might add, we often have to make in real life as well. Part III of the book is quite long and in some sense also quite difficult because it involves sophisticated ideas from algebraic topology and manifold theory. One of the reasons that it is so long is that Chapter 12 consists almost entirely of the background material I felt we had to recall in order to give the average reader of this book a chance.

So, is it worth the investment?

The (B–NB; Ht) theorem is probably the most important single result in the geometric theory and to amputate it from its 'only if' part would certainly be a pity. Even so, if it were just for this theorem, we might have been tempted to skip the 100-odd pages of the proof.

On the other hand, this homology theory that we have to develop is interesting in its own right. I also feel, and wish to convince you, the reader, that it is here that most of the future prospects of the theory lie. One cannot after all forget that homology is the most important invariant of any geometric object (and not just of them).

As a consequence I decided to embark on Part III.

In the next two sections we shall describe in fairly complete terms the homology that is used. This will give readers a chance to decide for themselves whether it is worth the 'investment' for the 'return' obtained.

1.7 Homology on Lie Groups

Let us start by recalling how two cohomology theories can be defined on a manifold M: the de Rham cohomology in this section and the Čech cohomology in the next.

1.7.1 The de Rham complex

Let M be some C^∞ manifold and let $\Lambda_p(M)$ be the space of smooth differential forms of degree p (set to zero if $p < 0$). The exterior differential

$$\mathrm{d}\colon \Lambda_p \to \Lambda_{p+1}$$

is a mapping such that $\mathrm{d}^2 = 0$ and therefore $\operatorname{Image}\mathrm{d} \subset \operatorname{Ker}\mathrm{d}$ for any $\Lambda_{p-1} \xrightarrow{\mathrm{d}} \Lambda_p \xrightarrow{\mathrm{d}} \Lambda_{p+1}$; that is, the image is contained in the kernel. The pth Betti number is $\dim[\operatorname{Ker}\mathrm{d} \subset \Lambda_p]/[\operatorname{Image}\mathrm{d} \subset \Lambda_p]$. For general manifolds these Betti numbers are usually ∞. However, they are finite for compact manifolds and, as we shall see, for Lie groups too.

To understand the above one has to know what differential forms are. If you do not, don't despair. The other (Čech) cohomology that we shall examine in the next section is equivalent (i.e. gives the same Betti numbers) and can be defined by entirely elementary means.

1.7.2 The case of a Lie group

Here the manifold is a connected Lie group G and is also assigned a left-invariant distance (which will not be used until later). In this case, the space $\Lambda_p(G)$ can be given a particularly simple description. We start with $\omega_1, \dots, \omega_n$, a basis of $T_e^*(G)$, the dual of the tangent space; then $\omega_I = \omega_{i_1} \wedge \omega_{i_2} \wedge \cdots \wedge \omega_{i_p}$ gives a basis of $\bigwedge^p T_e^*(G)$ as I runs through the various multi-indices $I = (i_1 < i_2 < \cdots < i_p)$. The group left translation $(x \mapsto gx)$ can then be used to obtain the corresponding basis $\omega_I(g)$ of $\bigwedge^p T_g^*(G)$, with $g \in G$, and the differential forms on G can then be written

$$\omega(g) = \sum a_I(g)\omega_I(g).$$

Readers who only 'half-know' what differential forms are should note that unlike the usual definition which only uses local coordinates (i.e. $\sum a_I \, \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_p}$), here, when we take the exterior differential, it is *not* in general true that $\mathrm{d}\omega_i = 0$. The advantage, however, of the above representation of differential forms is that it is global and we can define intrinsically

$$|\omega(g)| = \sum_I |a_I(g)|; \quad g \in G,$$

where the summation runs through the distinct multi-indices that give a basis (to $\bigwedge T^*G$). It is clear that, up to multiplicative constant, this definition is independent of the choice of the basis $\omega_1, \dots, \omega_n$.

Be that as it may, in the case of Lie groups, two important things have to be highlighted. The first is a deep theorem, but also an automatic consequence of the retract property that we described in §1.6.1. This theorem says that a Lie group has finite homology and has in fact exactly the same Betti numbers as does the maximal compact subgroup (K in the notation of §1.6.1).

The second thing that we shall highlight is that a different complex can now be defined. This will be called the *polynomial complex*. For this we can consider $\Lambda_{\mathrm{Pol}}(G)$ which is the subspace of differential forms ω on G that are of polynomial growth, that is, the forms for which constants C exist such that

$$|\omega(g)| \leqslant C(1+|g|)^C; \quad g \in G.$$

Obviously it is not in general true that $\mathrm{d}^2\Lambda_{\mathrm{Pol}} \subset \Lambda_{\mathrm{Pol}}$, so if we want to have a complex (i.e. $\mathrm{d}^2\colon \Lambda \to \Lambda$, $\mathrm{d}^2 = 0$) we must defined this complex by $\Lambda_{\mathrm{Pol}} \cap \mathrm{d}^{-1}\Lambda_{\mathrm{Pol}}$ (if ω lies in there then $\mathrm{d}\omega \in \Lambda_{\mathrm{Pol}}$ but also, by $\mathrm{d}^2 = 0$, automatically $\mathrm{d}\omega \in \mathrm{d}^{-1}\Lambda_{\mathrm{Pol}}$).

Using this complex, new Betti numbers can be defined as before: the polynomial Betti numbers. If these are finite we say that the Lie group G has *finite polynomial homology*.

To end the suspense, here is the key theorem.

Theorem (B–NB; Hl) *Let G be some connected Lie group. Then,*

(i) *if G is an* NB*-group it has finite polynomial homology*;
(ii) *if G is a* B*-group it does not have finite polynomial homology.*

Before we comment further let us illustrate this theorem and see what it means in a simple concrete case.

Example Let U be a model as in §1.5 and let ω be some closed differential form on U without constant term (i.e. $\mathrm{d}\omega = 0$). We shall assume that ω is of polynomial growth. Then since U as a differential manifold is a Euclidean space $\mathbb{R}^a$, we can surely solve the Poincaré equation $\mathrm{d}\theta = \omega$ for some other smooth differential form θ on U. The question that arises is whether that solution θ can also be chosen to be of polynomial growth. Our theorem provides the answer: yes if U is NB; no if U is B.

1.7.3 The homological investment/return ratio [sic]

It is very easy to see that (i) in the (B–NB; Hl) theorem is an easy consequence of the corresponding part (i) of the (B–NB; Ht) theorem. Therefore, as explained in §1.6.3, the proof of this is given in Appendix F, though it is essentially contained in Parts I and II of the book.

This is not so for part (ii) of (B–NB; Hl). The proof of this is much more difficult and this interacts with the corresponding homotopy result in §1.6. Let us explain. To get this part (ii) we use Part III of the book: pretty much the whole of it. This is in fact what Part III is all about. But now, and this is again automatic, part (ii) of (B–NB; Ht) follows at once. (*Grosso modo*: (i) Ht $\Longrightarrow$ (i) Hl; (ii) Hl $\Longrightarrow$ (ii) Ht.) Readers who have had some exposure to algebraic topology can no doubt see why all this happens. We shall leave it at that, except to say that this fact puts Part III in a broader perspective. The moral is that although in Part III the investment is quite high, the return is high also.

1.8 Čech Cohomology on a Lie Group

As promised, we now describe another complex on a Lie group that gives the same Betti number as the de Rham complex. This construction is more general and can be adapted to any topological space that is not a manifold. We give this here also because the construction is more elementary and some readers may prefer it. It should be noted too that in Part III of the book, systematic use will be made of both these complexes.

1.8.1 A good cover of a Lie group

The terminology 'good' for a cover, say for a manifold, in the context of Čech cohomology indicates a number of specific properties (see Bott and Tu, 1982) which here, in the case of a Lie group, can be described in the following concrete way. Let $\mathscr{U} = (U_\alpha; \alpha \in A)$ be a cover by open sets of the connected Lie group G where A is a countable *ordered* index set. The sets U_α are left translates of a small neighbourhood of the identity $e \in U \subset G$, that is, $U_\alpha = g_\alpha U$. To be specific, $U = B_a$ is some open ball of radius $a > 0$ centred at e, for some left-invariant Riemannian metric that has been fixed once and for all. As for the points $[g_\alpha \in G; \alpha \in A]$, they give rise to a subset of G that is chosen to be maximal under the condition that the distance $d(g_\alpha, g_\beta) \geqslant b$, with $\alpha, \beta \in A, \alpha \neq \beta$; here $b > 0$ is fixed. (Zorn's lemma, among other things, can be used. Maximal means that if we add an extra point, the condition breaks down.)

It is then very easy to verify that if $a \gg b$ then the open sets $U_a = g_a U$ give a locally finite cover of G (the 'locally finite' comes from the condition $d(g_\alpha, g_\beta) \geqslant b$, which implies that $g_\alpha B_{b/10}$ are disjoint balls of the same Haar measure).

The pivotal property that has to be imposed on the size of the small ball U is that it is geodesically convex for the left-invariant Riemannian metric on G. Geodesically convex open sets on a Riemannian manifold are diffeomorphic with $\mathbb{R}^n$, and intersections of geodesically convex sets are geodesically convex. The definition is what it says: any two points x, y on the set can be joined by a unique minimising geodesic. That geodesic lies entirely in the set and no other geodesic that joins them lies entirely in the set. The well-known Whitehead lemma says that if the radius a of the ball U is small enough, then U and therefore all the sets U_α and all their intersections $U_{\alpha_1,\dots,\alpha_p} = U_{\alpha_1} \cap \dots \cap U_{\alpha_p}$, with $\alpha_1 < \dots < \alpha_p$, are also geodesically convex.

1.8.2 The Čech complex

Now let $\mathscr{U} = (U_\alpha;\ \alpha \in A)$ be some cover as in §1.8.1. Here this could be any cover of any set; the only thing that counts is that we can define the sets $U_{\alpha_1,\dots,\alpha_p}$ as above. On each of these sets we specify the space of constant functions with value $c_{\alpha_1,\dots,\alpha_p} \in \mathbb{R}$. This gives a one-dimensional vector space. The direct product of all these vector spaces for which the length of the multi-index $\alpha_1 < \dots < \alpha_p$ is p is denoted by $C^{p-1} = C^{p-1}(\mathscr{U})$ and its elements are called *chains*. The operator $\delta\colon C^{p-1} \to C^p$ is then defined by

$$(\delta c)_{\alpha_1,\alpha_2,\dots,\alpha_{p+1}} = \sum_{j=1}^{p+1} (-1)^j c_{\alpha_1,\dots,\hat{\alpha}_j,\dots,\alpha_{p+1}},$$

with the following explanations for this formula: the 'hat' means that α_j has been suppressed and for any $c \in C^{p-1}$ the $c_{\beta_1,\dots,\beta_p}$ indicates the coordinates of c in the direct product vector space. Furthermore, $c_{\alpha_1,\dots,\hat{\alpha}_j,\dots,\alpha_{p+1}}$ is a constant function of $U_{\alpha_1,\dots,\hat{\alpha}_j,\dots,\alpha_{p+1}}$; therefore it stays a constant function on $U_{\alpha_1,\dots,\alpha_{p+1}}$.

Here are the facts that can then be verified. Together with the above spaces C^{p-1}, $p \geqslant 1$, we let $C^p = 0$ for $p < 0$. We then have $\delta^2 = 0$ when $C^{p-1} \underset{\delta}{\to} C^p \underset{\delta}{\to} C^{p+1}$ and $\{C^p;\ p \in \mathbb{Z}\}$ is a complex. The proof is easy.

What also holds, but is less easy to prove, is that the Betti numbers obtained from this complex $\dim([\operatorname{Ker}\delta \subset C^p]/[\operatorname{Image}\delta \subset C^p])$ are exactly the same as the Betti numbers obtained from the de Rham complex of §1.7.1.

The above holds, as long as the cover is a 'good' cover, on any manifold and is not restricted to groups.

1.8.3 The polynomial complex

We shall denote by C^{p-1}_{Pol} the subspace of chains $\sigma \in C^{p-1}$ for which there exist constants $C > 0$ such that

$$|\sigma_{\alpha_1,\dots,\alpha_p}| \leqslant C(1+|g_{\alpha_1}|)^C; \quad \alpha_1 < \alpha_2 < \cdots < \alpha_p.$$

More symmetric but equivalent notation would have replaced $|g_{\alpha_1}|$ in the above by distance$(e, U_{\alpha_1,\dots,\alpha_p})$.

The situation is more satisfactory here than for the de Rham complex because $\delta C^p_{\mathrm{Pol}} \subset C^{p+1}_{\mathrm{Pol}}$ and therefore we can use this to define a new complex and the corresponding polynomial Betti numbers $\dim([\mathrm{Ker}\,\delta \subset C^p_{\mathrm{Pol}}]/[\mathrm{Image}\,\delta \subset C^p_{\mathrm{Pol}}])$.

At this point we go back to $U = B_a$, which was used to define the Čech complex on the Lie group G; we can prove the following.

Theorem *The Betti numbers of the polynomial de Rham complex and of the polynomial Čech complex are identical as long as the radius a of U is small enough.*

This Čech complex can therefore be used to give the geometric characterisation of the B- or NB-groups.

The de Rham complex is the one that is more often used on manifolds. The notion of the Čech complex is closer to combinatorial ideas and in particular to that of a graph. We felt therefore that in the case of a Lie group (which is both a manifold and a group) both of these complexes deserved to be mentioned in this introductory chapter.

1.9 The Role of the Algebra in the B–NB Classification

What we have described in this chapter can be rephrased by saying that we have classified connected Lie groups into two classes and this we do in two different ways.

Analytic This is done in §1.3.

Geometric This is done in §§1.4–1.8.

Seen like this, the issue is to prove that the two classifications are identical.

As it happens, this equivalence is not proved directly; rather we have to make the classification a third way. This is achieved by conditions on $\mathfrak{g}$, the Lie algebra of the group.

Algebraic We consider $\mathfrak{g}$, a real Lie algebra. We introduce algebraic conditions that classify $\mathfrak{g}$ into two classes: the B-algebras and the NB-algebras.

To close the circle we show first that the connected Lie group G is a B-group (resp. NB-group) if its algebra is a B-algebra (resp. NB-algebra). This is done in the analytic theory in Part I of the book.

Similarly in Parts II and III, we prove the same equivalence between the geometric and algebraic classifications. In the equivalence of §1.10 below we show schematically how we pass from one classification to the other.

We see from the above that the game is played on three different levels. This explains, to some extent, why the story is so long in the telling.

An important prospect Here in fact lies the single most important prospect for the further development of the theory. The issue is to shortcut the Lie algebra altogether, as was done in Varopoulos et al. (1992) for unimodular groups (see §1.1 above), and obtain directly the proof of the equivalence between the analytic and geometric classifications. One hopes that this will open the way to other classes of groups. Why not be optimistic: all discrete finitely generated groups as in §1.1, or at least lattices in Lie groups? And so on.

1.10 A Broader Overview and Suggestions for the Reader

There are two aspects to the material of this book. The first is qualitative and is the classification of connected Lie groups and is given at three different levels:

$$\text{analytic} \Leftrightarrow \text{algebraic} \Leftrightarrow \text{geometric.}$$

The methods of the first equivalence are functional analytic and we do not use any 'hard' estimates.

To illustrate the issue let $\mu \in \mathbb{P}(G)$ and ϕ_n be as in §1.3.1 and let $e^{-\lambda}$ be the operator norm of $f \mapsto f * \mu$ in $L^2(G;d^r g)$, $\lambda \geqslant 0$. It is then an easy matter to show (see §3.1.3) that for all $\varepsilon > 0$ there exist C, C_ε (here C_ε depends on ε) such that

$$C_\varepsilon e^{-(\lambda+\varepsilon)n} \leqslant \phi_n(e) \leqslant C e^{-\lambda n}.$$

The analytic $\Leftrightarrow$ algebraic classification then consists in finding algebraic conditions on G under which one or the other of the conditions below hold:

(B) $\phi_n(e) \leqslant C \exp(-\lambda n - c n^{1/3})$.
(NB) $C e^{-\lambda n} n^{-c} \leqslant \phi_n(e)$.

The second aspect of the theory builds on this classification and gives precise results. This aspect is well illustrated by (NB): with $\lambda = \lambda(\mu)$ as in (NB) of §1.3.1 the issue is to determine explicitly the $\alpha = \alpha(\mu)$ for which (NB) holds. This is done in terms of G and the appropriate geometric invariants of μ. The problem is quite difficult and relies on delicate probabilistic estimates. This result is the generalisation of the local central limit theorem in $\mathbb{R}^d$ where $\alpha = d/2$. Results of this kind, however, are definitely less basic than the general classification of the groups.

For these reasons I have emphasised the first aspect, and the exact computation of α and such like can 'wait for another day'! Furthermore, apart from being more fundamental, once the methods of the classification are well understood, the sharp results are then obtained by refining these methods. Finally, we should point out that this first aspect of the theory has reached its final form.

Guide for the reader for Part I The analytic $\Leftrightarrow$ algebraic classification for amenable groups is contained in Chapters 2 and 3. There the reader will find the basic logic of the proofs and should study these chapters carefully. To pass to the general (non-amenable) case, new ideas are needed and these are developed in Chapters 4 and 5.

The proofs in these chapters are not intrinsically difficult; they are, however, long and they necessitate reasonable familiarity with Lie algebra theory, the structure of Lie groups, probability theory. All of these come from different mathematical cultures and herein lies the main difficulty in reading this book.

To deal with this difficulty and so as not to break the logic of the arguments, a deal of additional material related to these topics is deferred to the appendices.

The algebraic $\Leftrightarrow$ geometric equivalence is an entirely different subject that can be read independently. But here again the methods differ wildly from place to place; for example, combinatorics are used in Chapter 9 and de Rham cohomology in Chapters 12–14.

From the above it should be clear that the proofs in the three parts should not be read 'linearly'. An alternative suggestion would be to pick up one set of techniques, for example random walks in Part I, or de Rham and Čech cohomology in Part III, and see how these are used to provide the required component of the B–NB classification.

PART I

ANALYTIC AND ALGEBRAIC CLASSIFICATION

2

The Classification and the First Main Theorem

An Overview of Chapters 2 and 3

In these two chapters we shall give the classification theorem of §1.3.1 for Lie groups that are amenable. These we recall are the groups for which $\lambda = 0$ in that theorem. In that case the groups are called C and NC groups and they are treated in Chapters 2 and 3 respectively (C stands for 'condition' and NC for non-C). The amenable groups are the main building blocks for the more general groups of Chapters 4 and 5.

The proofs in these two chapters are long and technical; therefore, in a first reading, only the Parts 2.1 and 3.1 of these chapters should be studied. In these first parts of the chapters we examine the convolution powers of compactly supported measures. In fact, in Chapter 1 a slight generalisation is considered where we prove the results for the convolution of several distinct measures. This is not because we aim for maximal generality but because this generalisation is essential in Chapter 4 and at any rate the proofs are not more difficult.

After Parts 2.1 and 3.1, it is clear that we have here a good classification as in §1.3 and it is only a matter of finding a way of introducing the spectral gap λ and the exponential $e^{-\lambda t}$ in the estimates. This is done in Chapters 4 and 5. The alternative version of the classification that uses the diffusion kernel of some invariant operator (see §1.3.2) is done in the second parts of Chapters 2 and 3.

The algebraic conditions on the Lie algebra that give the C–NC classification are given in §§2.1–2.3. Basic linear algebra is used there and in particular the 'root space decompositions'. This can be found in standard textbooks, for example Varadarajan (1974) or Jacobson (1962) among others.

In §§2.4–2.5 we give some easy elementary properties on convolution of measures that will be used throughout.

Once this is done, the C-part of the theorem is given in §§2.7–2.11. But before that we shall give a detailed 'plan of the proof' that we hope will help the reader unravel and understand the interconnections between these sections of Chapter 2.

In these two chapters a scheme emerges which we follow consistently in the first part of the book. To wit,

(i) probabilistic- and potential-theoretic or other accessory ad hoc tools are generally given in the appendices;
(ii) basic special cases of groups are identified and the theorems are first proved in these particular cases;
(iii) the general results follow by reductions to these special cases.

The reductions needed in (iii) make extensive use of the structure theory of Lie groups (see Pontrjagin, 1939; Weil, 1953; Hochschild, 1965; Helgason, 1978; Varadarajan, 1974), a subject that is not trivial. The reader not familiar with this theory could skip these reductions altogether.

In Chapter 3 the special cases of (ii) are the simply connected groups, or better still, groups that are appropriate semidirect products $N \leftthreetimes V$ (see §3.4). In Chapter 2 it is more difficult to pinpoint these appropriate special cases on which the theorem is to be proved (this is done in §2.6). Identifying these special cases is an essential ingredient of the proofs.

The pivotal part of the proofs is the proof for the special cases (ii). Here the key parts of Chapters 2 and 3 are §§2.7–2.11 and §§3.4–3.5 respectively.

Standing convention Unless otherwise stated, all the Lie algebras will be finite-dimensional over the field $\mathbb{R}$ of real numbers. The definitions, the notation and the basic facts come from Varadarajan (1974), Jacobson (1962).

Part 2.1: Algebraic Definitions and Convolutions of Measures

2.1 Soluble Algebras and Their Roots. The Levi Decomposition

2.1.1 The nilradical

Let $\mathfrak{q}$ be some soluble Lie algebra and let $\mathfrak{n} \lhd \mathfrak{q}$ be its nilradical, that is, the largest nilpotent ideal (see Jacobson, 1962, §§3.5, 3.7, 3.8). We denote by $[\cdot,\cdot]$ the multiplication in the algebra and for $A, B \subset \mathfrak{q}$ we denote by $[A,B]$ the linear combination of $[a,b]$, $a \in A$, $b \in B$, and we then have $[\mathfrak{q},\mathfrak{q}] \subset \mathfrak{n}$ (Varadarajan, 1974, §3.8.3). We also denote $[\mathfrak{n},[\mathfrak{n},[\dots,\mathfrak{n}]]\dots] = \mathfrak{n}^p$ (with p factors). With this notation we recall that '$\mathfrak{q}$ is a Lie algebra' means $[x,y]+[y,x]=0$, $[x,[y,z]]+$

$[z,[x,y]]+[y,[z,x]]=0$. An ideal $\mathfrak{h}$ is called nilpotent if $\mathfrak{h}^p=[\mathfrak{h},[\mathfrak{h},\ldots]]=0$ for some p. The definition of solubility can be taken to mean $[\mathfrak{q},\mathfrak{q}]\subset\mathfrak{n}$.

For any real vector space V we denote its complexification by $V^{\mathrm{c}}=V\otimes_{\mathbb{R}}\mathbb{C}$ and also write $\mathrm{ad}\,x\colon \mathfrak{q}^{(\mathrm{c})}\to\mathfrak{q}^{(\mathrm{c})}$ for the mapping induced by $\mathrm{ad}\,x\colon y\to[x,y]$, $x,y\in\mathfrak{q}$. Sometimes, for additional clarity, we shall write $\mathrm{ad}\,x=\mathrm{ad}(x)=(\mathrm{ad}\,x)$. When confusion could arise we shall denote the complexification by $V^{(\mathrm{c})}$.

Now let $V=\mathfrak{q}/\mathfrak{n}$, $W=\mathfrak{n}/[\mathfrak{n},\mathfrak{n}]$; the action ad induces mappings

$$\mathrm{ad}\,v\colon W\to W,\quad \mathrm{ad}\,v\colon W^{\mathrm{c}}\to W^{\mathrm{c}};\quad v\in V. \tag{2.1}$$

We can identify V with a commutative algebra of linear transformations and define the roots and the root space decomposition of the space W^{c} under the action $\mathrm{ad}\,v$; see Jacobson (1962, §II.4).

To be more specific, the roots $\lambda_j\in\mathrm{Hom}_{\mathbb{R}}[V,\mathbb{C}]$ are complex-valued linear functionals of V and to each such root there corresponds a complex subspace $0\neq W_j\subset W^{\mathrm{c}}$, called the root space, and the defining relation is

$$(\mathrm{ad}\,v-\lambda_j(v))^n\,w=0;\quad v\in V,\ w\in W_j,\ n=\dim\mathfrak{q}+10. \tag{2.2}$$

These subspaces are invariant under the V action and give the direct space decomposition

$$W^{\mathrm{c}}=W_1\oplus\cdots\oplus W_t. \tag{2.3}$$

We also recall that the action of V on W_j is given by a scalar matrix $\lambda_j(v)I$ plus a nilpotent transformation, that is, strictly upper triangular for an appropriate basis. We can decompose

$$\lambda_j=\mathrm{Re}\,\lambda_j+i\,\mathrm{Im}\,\lambda_j,\quad L_j=\mathrm{Re}\,\lambda_j \tag{2.4}$$

into their real and imaginary parts. If we collect together $\lambda_{\alpha_1},\ldots,\lambda_{\alpha_s}$, all the roots with the same real part $L=L_{\alpha_1}=\cdots=L_{\alpha_s}$, we obtain

$$W_{\alpha_1}\oplus\cdots\oplus W_{\alpha_s}=W_L=\widetilde{W}_L\otimes\mathbb{C}, \tag{2.5}$$

where $\widetilde{W}_L\subset W$. To see this we shall consider on W^{c} the complex conjugation that sends $u+iv$ to $u-iv$ for $u,v\in W$. We denote this as usual by $x\to\bar{x}$, with $x\in W^{\mathrm{c}}$, and from the above it follows that $\overline{W_L}=W_L$ because the complex conjugate takes the root space with root λ_j to the root space with root $\bar{\lambda}_j$. In other words, the W_L are real spaces and we can use the exercise below. We obtain thus

$$W=\widetilde{W}_1\oplus\cdots\oplus\widetilde{W}_p. \tag{2.6}$$

This will be called the real root space decomposition that corresponds to the distinct real parts $\mathscr{L}=(L_1,\ldots,L_p)\subset V^*$ the real dual of V, and in an abuse of terminology that we shall adopt throughout, the L_j will be called the *real roots*. To stress the point and avoid confusion, this set is *not* the set of the roots that happen to be real. This set reduces to the empty set if and only if the algebra reduces to zero.

Exercise A subspace $E \subset W \otimes \mathbb{C}$ over $\mathbb{C}$ is called real if $\overline{E} = E$ for the above complex conjugation. Such a space is of the form $E = \widetilde{E} \otimes \mathbb{C}$ for some $\widetilde{E} \subset W$. It suffices to observe that when $E \neq \{0\}$ it contains some $0 \neq v \in E \cap W$ (here W is considered as a real subspace of W^{c}). To see this let $0 \neq x \in E$; then $2x = (x + \bar{x}) + i(x - \bar{x})/i$. So $v\mathbb{C} \subset E$ and therefore $E = v\mathbb{C} \oplus E_1$ for another real subspace E_1 of lower dimension (to see this use a basis $v, v', \dots$ for W). The same thing for more general field extensions is used in §6.5.4 below.

2.1.2 The radical and the Levi decomposition

Let $\mathfrak{g}$ be a general Lie algebra and let $\mathfrak{q} \lhd \mathfrak{g}$ be its radical, that is, its largest soluble ideal. We can then write $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ as a semidirect product (see Varadarajan, 1974, §3.14), where $\mathfrak{s}$ is a Levi subalgebra which is a semisimple algebra, that is, its radical is 0 (see Varadarajan, 1974, §3.8). This means, we recall, that $\mathfrak{g} = \mathfrak{q} + \mathfrak{s}$ is a direct vector space sum, that is, that $\mathfrak{q} \cap \mathfrak{s} = \{0\}$; it is called the Levi decomposition of the algebra.

2.2 The Classification

2.2.1 Soluble algebras

Let $\mathfrak{q}$ be a soluble algebra and let $\mathscr{L} = (L_1, \dots, L_p)$ be the distinct real roots as in §2.1.1. We then say that

- $\mathfrak{q}$ *is a* C-*algebra if there exist* $\alpha_1, \dots, \alpha_p \geqslant 0$ *for* $1 \leqslant j \leqslant p$ *such that* $\sum_{i=1}^{p} \alpha_i L_i = 0$ *and such that for at least one real root* L_j *we have* $\alpha_j L_j \neq 0$;
- $\mathfrak{q}$ *is an* NC-*algebra* (*non*-C-*algebra*) *if for all* $\alpha_1, \dots, \alpha_p \geqslant 0$ *the condition* $\sum_{i=1}^{p} \alpha_i L_i = 0$ *implies* $\alpha_i L_i = 0$, $1 \leqslant i \leqslant p$.

Clearly every soluble algebra is either C or NC. This says, in particular, that nilpotent algebras are NC, for then $\mathscr{L} = (0)$ or $\mathscr{L} = \emptyset$.

For a general algebra $\mathfrak{g}$, we say that it is a C-algebra (resp. NC-algebra) if its radical is a C-algebra (resp. NC-algebra).

Exercise 2.1 If $\mathfrak{n}$ is the nilradical, show that $\mathfrak{g}/\mathfrak{n}^2$ is a C-algebra if and only if $\mathfrak{g}$ is. If $\mathfrak{z}$ is the centre of $\mathfrak{g}$ then $\mathfrak{g}/\mathfrak{z}$ is a C-algebra if and only if $\mathfrak{g}$ is. The product algebra $\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{g}_2$ is NC if and only if $\mathfrak{g}_1$ and $\mathfrak{g}_2$ both are. *Hint*. The nilradical of $\mathfrak{g}/\mathfrak{n}^2$ is $\mathfrak{n}/\mathfrak{n}^2$. This holds because if $\xi \in \mathfrak{g}$ is such that its image in $\mathfrak{g}/\mathfrak{n}^2$ is in the nilradical of that algebra, then $(\mathrm{ad}\,\xi)^k\mathfrak{g} \subset \mathfrak{n}^2$ for some k and by iterating (i.e. $(\mathrm{ad}\,\xi)^s\mathfrak{n} \subset \mathfrak{n}^3$, for some s, and so on ...), this implies that $(\mathrm{ad}\,\xi)^p = 0$ for some p. This shows that $\xi \in \mathfrak{n}$ because it is in the radical and we can use Varadarajan (1974, §3.8.3). By an analogous argument we see that the nilradical of $\mathfrak{g}/\mathfrak{z}$ is $\mathfrak{n}/\mathfrak{z}$. Note, however, that for the proof of the exercise, the characterisations in §§2.3.3–2.3.4 below are much better adapted.

2.2.2 Amenability and the R-condition

We also recall that $\mathfrak{s}$ a semisimple algebra is said to be of compact type if the Lie groups that correspond to $\mathfrak{s}$ are compact (the point here is that if one such group is compact then they all are; Varadarajan, 1974, §4.11.6; Helgason, 1978, II, §6.9). The Lie groups G for which the Levi subgroups $S \subset G$ are compact are called *amenable*; S is the analytic subgroup that corresponds to the Levi subalgebra of the Lie algebra. In Chapter 3 we shall give more details on amenability. But provisionally this will be taken to be the definition.

The definition of R-algebras is widely used in the literature; see Guivarc'h (1973), Jenkins (1973). We say that $\mathfrak{g}$ is an R-algebra if and only if the real roots of §2.1.1 for the radical $\mathfrak{q} \lhd \mathfrak{g}$ are all zero (i.e. $\mathscr{L} = (0)$), and if in addition $\mathfrak{s}$, the Levi subalgebra, is of compact type. According to our definition such algebras are NC. In this chapter we shall use this notion only in §§2.5.4 and 2.3.5 and this just to illustrate the main theorem. The reader can therefore ignore this notion for the time being if they so wishes.

The above classification extends to connected real Lie groups. If G is such a group we say that it is a C-group (resp. NC-group, R-group) if its Lie algebra $\mathfrak{g}$ is.

2.3 Equivalent Formulations of the Classification and Examples

2.3.1 Affine geometry

Let $\mathfrak{q}$ be soluble and $V = \mathfrak{q}/\mathfrak{n}$ as in §2.1.1. Then $\mathfrak{q}$ is an NC-algebra as long as there exists $x \in V$ such that $Lx > 0$ for $L \in \mathscr{L}\backslash\{0\} = \mathscr{L}_1$. Conversely, if the soluble algebra $\mathfrak{q}$ is an NC-algebra we can use Hahn–Banach in the finite-dimensional space V^* to separate 0 from the convex hull $(\mathscr{L}_1)$ and this gives $x \in V$ such that $Lx > 0, L \in \mathscr{L}_1$.

Similarly, $\mathfrak{q}$ is a C-algebra if and only if there exists $V_1 \subset V, V_1 \neq V$ (i.e. a strict subspace), and a constant $c > 0$ such that

$$L_1^+(x) + L_2^+(x) + \cdots + L_p^+(x) \geqslant c|\hat{x}|_{V/V_1}; \quad x \in V. \tag{2.7}$$

Here $L^+ = \sup(L, 0) = L \vee 0$ and $\hat{x}$ is the image of x in V/V_1. Here we have fixed some norm on V and in (2.7) we use the quotient norm. *Grosso modo* this says that $\inf_{x \in \hat{x}} \sum L_j^+(x)$ can be used as an equivalent norm on V/V_1.

If we use the previous characterisation of NC-algebras we see that (2.7) implies that $\mathfrak{q}$ is not NC. Indeed, if $x \in V$ is such that $Lx > 0, L \in \mathscr{L}_1$, we may suppose that $x \notin V_1$. But then substituting $-x$ in (2.7) we get a contradiction with (2.7).

Conversely, assume that $\mathfrak{q}$ is C and let $L_{i_1}, \ldots, L_{i_k}$ be a minimal set of non-zero real roots that satisfy $\sum_{j=1}^k \gamma_j L_{i_j} = 0$ and $\gamma_j > 0$. The L_{i_j} thus form a non-trivial simplex in some subspace $E \subset V^*$ that contains 0 in its interior. But then for all $0 \neq X \in E^*$ in the dual we have $\max_j \langle L_{i_j}, X \rangle > 0$; otherwise all the L_{i_j} would be on one side of a hyperplane in E. By elementary (finite-dimensional) duality we can identify E^* with $V/E^\perp$ and (2.7) follows.

2.3.2 Examples

(i) For a nilpotent group in §2.1.1 all the roots λ_j equal 0 because by the definition, $V = \{0\}$ in (2.1). These groups are therefore R-groups.

Less trivial but typical illustrations of the definitions that we have given are supplied by *semidirect products*. We recall that $G = H \leftthreetimes K$ if H is a closed normal subgroup, K is a closed subgroup and $K \cap H = \{0\}$. See Varadarajan (1974, §3.14), Hochschild (1965, IX) and §8.3 in Part II of the book.

(ii) The group of affine motions on $\mathbb{R}$, $x \mapsto ax + b$, $a = e^\alpha > 0$, $b \in \mathbb{R}$ is soluble and a semidirect product of the translations $x \mapsto x + b$ which is the nilradical, and the dilations. This group is the semidirect product $\mathbb{R} \leftthreetimes \mathbb{R}$. There is only one root $\lambda(\alpha) = \alpha$. This group is therefore a soluble NC-group.

(iii) We can generalise the previous example to the group of affine motions on $\mathbb{R}^d$ which is $\mathbb{R}^d \leftthreetimes \mathrm{GL}_d(\mathbb{R})$ with the natural action of GL_d on $\mathbb{R}^d$. To obtain a soluble group we can consider $\pi\colon \mathbb{R}^m \to \mathrm{GL}_d(\mathbb{R})$ a homomorphism for some $m \geqslant 1$ and the induced semidirect product $\mathbb{R}^d \leftthreetimes \mathbb{R}^m = W \leftthreetimes V = G$. A concrete example is obtained by

$$\begin{gathered}\pi(v) = \text{diagonal}(\exp(L_1(v)), \ldots, \exp(L_d(v))); \\ L_1, \ldots, L_d \in V^* = \text{the real dual.}\end{gathered} \tag{2.8}$$

The group G is then soluble and its roots $L_1, \dots, L_d$ are real. We have a C-group when $d = 2, m = 1, L_1 v = v, L_2 v = -v$.

2.3.3 The use of Lie's theorem

Let $\mathfrak{q}$ be some soluble Lie algebra. By Lie's theorem, once we complexify, we can find a basis in $\mathfrak{q}^c = \mathfrak{q} \otimes \mathbb{C}$ for which all the linear mappings $\operatorname{ad} x$ become simultaneously triangular (see Varadarajan, 1974, §3.7.3):

$$\operatorname{ad} x = \begin{pmatrix} \lambda_1(x) & & \bigstar \\ & \ddots & \\ O & & \lambda_n(x) \end{pmatrix}. \tag{2.9}$$

These new $\lambda_j \in \operatorname{Hom}_{\mathbb{R}}[\mathfrak{q}, \mathbb{C}]$ are also called roots and as before we can define a new set $\mathscr{L}' = (L_1, \dots, L_n)$ with $L_j = \operatorname{Re} \lambda_j \in \mathfrak{q}^*$. This new set can be used as before to classify C- and NC-algebras. This classification thus obtained is identical to the one that we have already given.

Exercise 2.2 Prove this: the proof is a consequence of the Jordan–Hölder composition series (see Jacobson, 1989, Volume 2, §3.3) and of the fact that $[\mathfrak{q}, \mathfrak{q}] \subset \mathfrak{n}$, the nilradical (cf. Varadarajan, 1974, §3.8.3). Use §2.3.4 below and the series

$$\mathfrak{q} \rhd \mathfrak{n} \rhd \mathfrak{n}^2 \rhd \cdots \rhd \mathfrak{n}^r = \{0\}.$$

Use this characterisation to prove that if $\mathfrak{r} \lhd \mathfrak{q}$ and if $\mathfrak{q}$ is NC then $\mathfrak{q}/\mathfrak{r}$ is also (this is not necessarily the case for $\mathfrak{r}$: for example, consider $Q_1 \times Q_2$, with $Q_i = (b_1, \alpha_i)$, the direct product of two copies of the group in §2.3.2(ii) and R, the subgroup for which $\alpha_1 = -\alpha_2$).

Note also that in (2.9) at least one λ_i is 0 because we always have $[\mathfrak{q}, \mathfrak{q}] \subset \mathfrak{n}$.

No use of this formulation of the classification will be made in this chapter. More details will, however, be given in the next two subsections.

2.3.4 The composite roots

In the proof of the NC-theorem in the next chapter, the ad-action of $\mathfrak{q}$ on $\mathfrak{n}$ will have to be examined in more detail. We shall now elucidate this but the reader can skip this subsection until it is actually needed in Chapter 3.

Observe first that $V = \mathfrak{q}/\mathfrak{n}$ acts by ad- also on each space $W_p = \mathfrak{n}^p/\mathfrak{n}^{p+1}$ (the notation is as in §2.1.1) and these spaces, once complexified, admit root space decompositions. The point to note here is that if $\lambda_1, \lambda_2, \dots$ are the roots of the action on $W = \mathfrak{n}/\mathfrak{n}^2$ then the roots of the action on W_p are the *composite* roots of the form $\lambda_{\alpha_1} + \cdots + \lambda_{\alpha_p}$. To see this we observe that $[\mathfrak{n}^p, \mathfrak{n}^q] \subset \mathfrak{n}^{p+q}$ and therefore $W_\infty = \sum W_p$ admits a structure of a graded Lie algebra for which $W_p = [W_1, [W_1, [\ldots] \ldots]]$. Both $\mathfrak{q}$ and V act naturally on W_∞ by derivations and the two actions can be identified in a natural way. We then use the fact that the bracket of two root spaces of W_∞ by the action of V lies in the root space with root the sum of the two roots, provided that this sum is a root and is 0 otherwise (see Jacobson, 1962, §III.2 and Chapter II, Exercise 8). As before we can take their real parts; these will be called composite real roots and are of the form $L_{\alpha_1} + \cdots + L_{\alpha_p} \in V^*$. By the Jordan–Hölder theorem we see that this new set of composite real roots counted with multiplicity is exactly the set $\mathscr{L}'$ defined in §2.3.3 provided that in that counting we ignore the roots that are 0. The characterisation of §2.3.3 follows because the NC-condition for $\mathscr{L}'$ clearly holds if and only if it holds for the set of their simple components in $\mathscr{L}$.

These composite real roots will be explicitly used in Chapter 3 in an important special case that we shall describe in the next few lines.

We shall consider $\mathfrak{h} \subset \mathfrak{q}$ some nilpotent subalgebra such that $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$. We shall see in §3.4.2 that such an algebra always exists. The composite real roots $\mathscr{L}'$ can then be identified with real linear functionals in $(\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n})^* \subset \mathfrak{h}^*$. These roots can, however, also be defined directly as follows.

The algebra $\mathfrak{h}$ induces by the ad-action some nilpotent algebra of linear transformations on $\mathfrak{n}$ and it induces therefore a root space decomposition of the complexified $\mathfrak{n}^{\mathrm{c}} = \mathfrak{n}_0^{\mathrm{c}} \oplus \cdots \oplus \mathfrak{n}_p^{\mathrm{c}}$ (see Jacobson, 1962, §II.4, and §3.8.2 below). This decomposition will be examined in some detail in §3.8. Here we shall simply point out that we can consider again $\mathscr{L}'' = (L_1, \ldots, L_p)$ the real parts of the roots of this decomposition. The same Jordan–Hölder argument shows that $\mathscr{L}' = \mathscr{L}''$ (provided again that we ignore the zero roots, i.e. $\mathscr{L}' \backslash \{0\} = \mathscr{L}'' \backslash \{0\}$).

2.3.5 An illustration: the modular function

This subsection will only be needed in the illustration in §2.5.4 of the C-theorem. The reader can either skip the details or treat them as exercises in Lie group theory.

Let G be connected Lie group and $\mathfrak{g}$ its Lie algebra. Then $\mathrm{Ad}\, g\colon\ \mathrm{GL}(\mathfrak{g}) \to \mathrm{GL}(\mathfrak{g})$ is the differential at e of $x \to gxg^{-1}$. For this and related notions and facts, see Varadarajan (1974, §2.13). By that definition we have, in §1.1, $m(g) = \frac{d^r g}{d^l g} = \det(\mathrm{Ad}\, g)$.

Exercise To see this use the differential forms of §12.5.1 from the geometric theory. If you get stuck, check it out in Helgason (1984, §2.5). But also by the definition, $\mathrm{Ad}(\exp\xi) = e^{\mathrm{ad}\,\xi}$, with $\xi \in \mathfrak{g}$, and therefore $m(\exp\xi) = e^{\mathrm{trace\,ad}\,\xi}$ (see Varadarajan, 1974, §2.13.8).

One immediate consequence of this is that G is unimodular if and only if $\mathrm{trace\,ad}\,\xi = 0$, with $\xi \in \mathfrak{g}$.

Also if G is soluble, then $\mathrm{trace\,ad}\,\xi = \sum L_j(\xi)$, $\xi \in \mathfrak{g}$, for the real roots as defined in (2.9): observe that the trace is always real. The conclusion one draws from this is that if G is soluble and unimodular then it has to be a C-group unless it is an R-group (verify this also for amenable groups). A much more complete discussion on the interaction between unimodularity and the C–NC classification can be found in Appendix B at the end of Part I of the book.

2.4 Measures on Locally Compact Groups and the C-Theorem

2.4.1 A class of measures

Let G be some connected locally compact group. We shall fix $K \subset G$, some compact subset, and $e \in \Omega = \Omega^{-1} \subset G$, some relatively compact neighbourhood of the identity element e of G, and $C, c > 0$, both positive constants. We shall then consider probability measures $\mu \in \mathbb{P}(G)$ that satisfy the following three conditions:

(i) $\mathrm{supp}\,\mu \subset K$;
(ii) $d\mu(g) = f(g)\,d^r g$ for some $f \in L^\infty$ with $\|f\|_\infty < C$: here we use the right Haar measure, but nothing changes if we use the left measure instead;
(iii) $f(g) \geqslant c$, $g \in \Omega$.

For a sequence of measures $\mu_1, \mu_2, \ldots,$ as above, we shall *abuse notation* slightly and write $\mu^n = \mu_1 * \cdots * \mu_n = f^{(n)}(g)\,dg$ where here $f^{(n)}$ is continuous for $n \geqslant 2$; see Weil (1953, §13). For this slightly annoying reason, many of our estimates below can only be formulated for $n \geqslant 2$. It is, however, technically more flexible to assume in (iii) that f in L^∞ rather than continuous. We can now state one of the main theorems in the subject.

Theorem 2.3 (The C-theorem) *Let G be a* C-*connected real Lie group* (*i.e. the radical of the Lie algebra is a* C-*algebra*) *and let K, Ω, C, c and measures $\mu_1, \mu_2, \ldots$; $\mu^n = f^{(n)}(g)\,dg$ satisfy conditions* (i), (ii), (iii) *above. Then there exist positive constants C_1, c_1, that depend only on G, K, Ω, C, c, such that*

$$f^{(n)}(e) \leqslant C_1 \exp(-c_1 n^{1/3}); \quad n \geqslant 2. \tag{2.10}$$

This is the upper estimate in §1.3.1 in the amenable case $\lambda = 0$. The converse of this will be proved in the next chapter. We shall show there that if G is NC instead, and is amenable, and if μ is symmetric, that is, if $\mu(x) = \mu(x^{-1})$ and satisfies the above conditions (i), (ii), (iii), then we have

$$f^{(n)}(e) \geqslant C_2 n^{-c_2}; \quad n \geqslant 2, \tag{2.11}$$

for appropriate constants C_2, $c_2 > 0$ and $f^{(n)}(g)\,dg = \mu^{*n} = \mu * \cdots * \mu$. The necessary facts on amenability will be given in §3.1 and we shall see that amenability is essential for (2.11) to hold. The symmetry of μ is also essential for obvious reasons. To see this, think of a measure in $\mathbb{R}$ that has most of its mass $\subset [100, \infty[$.

The above facts put together give the analytic side of the C–NC classification of §1.3 (see also §1.10) provided that we restrict ourselves to amenable groups because (C) $\Rightarrow$ (2.10), (NC) $\Rightarrow$ (2.11).

The proof of the C-theorem that we give in this chapter in fact only works for amenable groups. But, as we shall explain in §3.1, if G is not amenable the much stronger estimate $f^{(n)}(e) = O(e^{-cn})$ holds anyway.

2.5 Preliminary Facts

2.5.1 The Harnack principle for convolution

There is nothing special about the identity $e \in G$ in the formulation of the C-theorem. To see this we maintain the notation and hypothesis of the theorem and write $f(n,g) = f^{(n)}(g)$, $g \in G$, $n \geqslant 2$, and fix $P \subset G$ some compact subset. We shall then use the fact that there exists $k \geqslant 1$ and $C_1 > 0$ such that

$$f(n,gx) + f(n,xg) \leqslant C_1 \tilde{f}(n+k,g); \quad n = 2,\ldots, g \in G,\ x \in P. \tag{2.12}$$

The unfortunate thing here is that $\tilde{f}$ is defined by a different sequence of measures $\tilde{\mu}_1, \ldots$ that in general also depends on n, k. These new measures, however, satisfy conditions (i), (ii) and (iii) of §2.4.1 with the same K, Ω, c, C. Despite this unpleasant complication, (2.12) will be seen to serve its purposes. Observe also that in the special case when the original measures are identical, that is, $\mu = \mu_1 = \mu_2 = \cdots$, this complication disappears and $\tilde{f} = f$. In that case we have a genuine Harnack estimate

$$\mu^{*n}(xg) + \mu^{*n}(gx) \leqslant C_1 \mu^{*(n+k)}(g), \tag{2.13}$$

where here and throughout when confusion does not arise, abusing notation we write $\mu(g)$ for $f(g)$ when $\mu = f(g)\,dg$.

The proof of (2.12) is immediate because with measures $\mu_1, \mu_2, \dots$ as, in the theorem, conditions (i), (ii), (iii) imply that, with $x \in P$, δ_x the Dirac point mass at x and k large enough, there exists C_1 such that

$$\delta_x * \mu_1 \leqslant C_1 \mu_2 * \cdots * \mu_k, \quad \mu_1 * \delta_x \leqslant C_1 \mu_2 * \cdots * \mu_k. \tag{2.14}$$

All this holds for general locally compact groups that are *connected*. To see this observe that, by §2.4.1(iii), for every $L \subset G$ relatively compact, there exists k_0 such that for $k \geqslant k_0$ we have $\mu_2 * \cdots * \mu_k = f\,dg$ where the function f is strictly positive on L. More explicitly, f is bounded below on L by some $c_1 > 0$ that depends only on Ω, c of (iii) and k and L (to verify this, use a smaller neighbourhood of e such that $\Omega_1^5 \in \Omega$). For (2.14) it suffices that $P(\operatorname{supp} \mu_1) \cup (\operatorname{supp} \mu_1) P \subset L$.

From the above proof we see that a more precise version of (2.12) actually holds. Namely, when P is fixed there exists some k_0 such that, for *all* $k \geqslant k_0$, there exists C_1 such that (2.12) or (2.14) holds. Furthermore, k_0 depends only on K, Ω, c, C, P, and C_1 depends on K, Ω, c, C, P, k.

The important point in all this is that the choice of k or k_0 and C_1 is *independent of n and even of the particular measures μ_i that one uses, as long as they verify (i), (ii), (iii) with fixed K, Ω, c, C.*

In §5.4 we shall come back to this Harnack principle in a context that makes the construction more transparent.

2.5.2 Applications of Harnack

Because of (2.12) we can replace e in (2.10) by any $g \in G$. Using (2.12) we also see that (2.10) in the theorem can be replaced by the following equivalent condition: for every compact set $Q \subset G$ there exist C_1, c_1 that also depend on Q such that

$$\mu_1 * \cdots * \mu_n(Q) \leqslant C_1 \exp(-c_1 n^{1/3}); \quad n \geqslant 1. \tag{2.15}$$

Conditions (i) and (ii) in §2.4.1 already show that (2.15) implies (2.10) (we use (2.15) for n and deduce (2.10) for $n+1$). We can also use (2.12) to see how $f^{(n)}(e)$ in the theorem behaves under a projection $\pi\colon G \to G/H$, where G is an arbitrary locally compact group and H is a closed normal subgroup.

Let us denote by $\check{\mu}_j = \check{\pi}(\mu_j)$ the direct image measures (Bourbaki, 1963) in G/H for measures μ_j in G that satisfy conditions (i), (ii), (iii) above. It then follows that the $\check{\mu}_j$ satisfy these conditions on G/H. To see this we use the fact that when $d\mu = f\,d^r g$ then $d\check{\mu} = \check{f}(\dot{g})\,d^r\dot{g}$ with

$$\check{f}(\dot{g}) = \int_H f(hg)\,d^r h \quad \text{for any } g \in \dot{g} \in G/H;$$

see (2.25) and (2.62) below for more details.

Let us write $\check{\mu}^n = \check{\mu}_1 * \cdots * \check{\mu}_n = \check{f}^{(n)}(\dot{g})\,d\dot{g}$, $\dot{g} \in G/H$. For any compact subset $Q \subset G/H$ we have $\mu^n(\pi^{-1}(Q)) = \check{\mu}^n(Q)$ because the image of the convolution is the convolution of the images. Therefore if we use formulation (2.15) we see that the C-theorem holds for G whenever it holds for G/H. To fix ideas assume that G is connected and that $\mu_1 = \mu_2 = \cdots$; in view of (2.13), we can then write, more precisely,

$$f^{(n)}(e) \leqslant C_1 \check{f}^{(n+k)}(\dot{e}), \tag{2.16}$$

for the identity elements e and $\dot{e}$ of G and G/H, and when H is compact we have

$$c_1 f^{(n)}(e) \leqslant \check{f}^{(n+k)}(\dot{e}) \leqslant C_1 f^{(n+2k)}(e). \tag{2.17}$$

Here the constants k, C_1, c_1 depend only on K, Ω, C, c in conditions (i), (ii) and (iii). This also holds for different measures $\mu_1, \mu_2, \ldots$ provided that (2.16) is interpreted as is (2.12) and we pass from f to $(\tilde{f})\check{}$, and similarly for (2.17).

2.5.3 A technical reduction

In proving Theorem 2.3 we may restrict ourselves to the special case where the first measure in the product has been fixed to be $\mu_1 = \nu$, some preassigned $\nu \in \mathbb{P}(G)$ that will be chosen in advance so as to satisfy appropriate conditions. This is clear from the same argument as in (2.14) that shows that $\mu_1 \leqslant \nu * \mu^{*k}$ for some k large enough and μ satisfying (i), (ii) and (iii) in §2.4.1. This particular technical point will be used in §§2.8–2.9.

2.5.4 Unimodular groups

The connection of the present theory with the more general theory of unimodular locally compact groups, which was briefly described in §1.2, lies in the fact that G is an R-group if and only if $\gamma(n)$ grows polynomially (see Guivarc'h, 1973; Jenkins, 1973), and as already pointed out in §2.3.5, if G is unimodular, amenable and not an R-group then G has to be a C-group.

For unimodular groups the C-theorem extends to general locally compact groups and we can assert that $f^{(n)}(e) = O(\exp(-cn^{\alpha/(\alpha+2)}))$ as long as $\gamma(n) \geqslant C\exp(cn^{\alpha})$, for some $0 < \alpha \leqslant 1$. For these results see Varopoulos et al. (1992, §7.5).

I came across this 'strange' bound $\exp(-cn^{1/3})$ in my early work on discrete groups (: 1983). Realising that this was in the nature of things played an important role in the development of the theory.

Guide to the Proof

Here we shall explain the interconnection of the subsections that follow.

In §2.6 we construct a special case on which the proof will be given. From these groups, in §2.11.3 we deduce, almost, the general case. This is quite convincing because in that way we obtain all simply connected groups and all soluble groups. The full generality is not done, however, until §2.11.3 and in a first reading the reader could skip this technical subsection. The proof for the special case of §2.6 is broken up in §§2.7–2.11 as follows.

In §2.7.1 we give a probabilistic interpretation of the problem, which is essential if one wants to comprehend what the proof is all about. Although nothing concrete is proved there the reader is strongly urged to understand that probabilistic interpretation because here lies the germ of the whole proof. In §2.7.2 again nothing concrete is done, other than the introduction of convenient notation. The interaction of these two subsections reflects the classical interplay between the notion of conditional expectation in probability theory and the disintegration of measures in measure theory. For more on this see Bourbaki (1963, Chapter 6, Historical Notes).

Although pivotal in the proof, §2.8 is just an ad hoc combinatorial lemma on convolution of measures on $\mathbb{R}^n$ and is an essential accessory to the proof.

Once we have the notation of §2.7.2 and the combinatorics of §2.8 we put these together in §2.9 to reduce the theorem to an explicit statement on classical random walks in $\mathbb{R}^n$.

For this finally we use the probabilistic estimates from the appendix, and the random walk statement of §2.9 is proved in §2.10.

2.6 Structure Theorems for Lie Groups and the Exact Sequence

Some familiarity with the basic facts about the structure of Lie groups will be needed in this section, but nothing very elaborate (e.g. Varadarajan, 1974). Alternatively, I suggest you believe what is in this section and move on.

2.6.1 The use of structure theory

Let G be some connected real Lie group and let $N \lhd G$ be its nilradical, that is, the subgroup that corresponds to $\mathfrak{n} \lhd \mathfrak{g}$, the nilradical of the Lie algebra. The nilradical N is closed but not necessarily simply connected (see Varadarajan, 1974, §3.18.13). We can, however, find $T \lhd N$, a compact torus, that is $\cong \mathbf{T}^p =$

$(\mathbb{R}(\mathrm{mod}\,1))^p$ such that N/T is simply connected. As a matter of fact T is a central subgroup both in N and G. To see this we use the simply connected cover $\pi\colon \widetilde{N} \to N$ (see Varadarajan, 1974, §2.6) and then $\ker\pi$ lies in the analytic central subgroup $\widetilde{Z} \subset \widetilde{N}$ (see Varadarajan, 1974, §3.6.4). Then T is $\widetilde{Z}/\ker\pi$. Alternatively, we can take for T a maximal normal torus in G; such a group has to be central in G because the automorphism group of T is discrete and therefore $T \subset N$. More details on T will be given later, in §5.8* and §11.3.1.

We shall consider then G/T and its nilradical N/T. Now, it is clear from the definition that G/T is a C-group if and only if G is a C-group (see Exercise 2.1 in §2.2.1). If we use the reduction of §2.5.2 we finally conclude that, in proving the C-theorem for the group G, we may assume that the nilradical $N \lhd G$ is simply connected.

Now let $N_2 \lhd N$ be the analytic subgroup that corresponds to $\mathfrak{n}^2 = [\mathfrak{n},\mathfrak{n}]$. The simple-connectedness of N implies that N_2 is then a closed subgroup normal in G (see Varadarajan, 1974, §3.18.1). The group $H = N/N_2$ is then simply connected Abelian (see Varadarajan, 1974, §3.18.2) and therefore is a Euclidean space, that is, $\cong \mathbb{R}^d$ and we have an exact sequence

$$0 \to H \to G/N_2 \to G/N = K \to 0. \tag{2.18}$$

By Exercise 2.1 in §2.2.1, G is a C-group if and only if G/N_2 is, and here H is the nilradical of G/N_2. For this last point see the hint in Exercise 2.1.

Here we shall use the reductions of §2.5 and for our C-theorem we end up with having to prove a special case of the theorem that will be analysed in detail in the next subsection.

2.6.2 Special case of the C-theorem

Let G be some C-group. Let us assume that H, the nilradical of G, is a Euclidean space so that we have an exact sequence

$$0 \to H \to G \to G/H = K \to 0. \tag{2.19}$$

Our aim is to show that estimate (2.10) of the C-theorem holds for G. What we have seen is that once this has been done, the C-theorem follows in full generality. The proof of this is done in several steps. In this subsection we shall examine more closely what the C-condition means for the exact sequence (2.19).

From (2.19) we shall consider the exact sequence of the corresponding Lie algebras

$$0 \to \mathfrak{h} \to \mathfrak{g} \to \mathfrak{g}/\mathfrak{h} = \mathfrak{k} \to 0. \tag{2.20}$$

Furthermore, let $\mathfrak{q} \lhd \mathfrak{g}$ be the radical of $\mathfrak{g}$ and let us denote by $\mathfrak{h} = \mathfrak{h}_1 \oplus \cdots \oplus \mathfrak{h}_p$ the real root space decomposition of $\mathfrak{h}$ with respect to the action of $\mathfrak{q}/\mathfrak{h}$ on $\mathfrak{h}$ as in (2.6). On the other hand, for any Lie algebra $\mathfrak{g}$, if $\mathfrak{q}$ is the radical and $\mathfrak{n}$ is the nilradical, we have $[\mathfrak{g}, \mathfrak{q}] \subset \mathfrak{n}$ (see Varadarajan, 1974, §3.8.3). From this and the Levi decomposition we conclude that we can identify $\mathfrak{k} = \mathfrak{q}/\mathfrak{h} \oplus \mathfrak{s}$, where $\mathfrak{s}$ is the Levi subalgebra of $\mathfrak{g}$. Since $\mathfrak{h}$ is Abelian, the algebra $\mathfrak{k}$ acts naturally by derivations on $\mathfrak{h}$. This will be made explicit in the next sentence at the level of the corresponding Lie groups. Furthermore, the $\mathfrak{h}_i$, $i = 1, \ldots, p$, are ideals in $\mathfrak{g}$ because, by the commutation of the above direct sum for $\mathfrak{k}$, the action of $\mathfrak{s}$ on $\mathfrak{h}$ stabilises the root spaces (2.2).

In terms of the Lie groups in the sequence (2.19), this means that the Euclidean space H can be decomposed as $H = H_1 \oplus \cdots \oplus H_p$, and that K acts by inner automorphisms on each subspace H_i. To fix notation, for all $k \in K$ and $\hat{k} \in G$, a pre-image of k in G, the action is $\tau_k(h) = \hat{k}h\hat{k}^{-1}$, $h \in H$. The definition of τ_k is independent of the particular choice of $\hat{k}$ because H is commutative, $k \mapsto \tau_k$ is a homomorphism $K \to \mathrm{GL}(H_j)$ and we shall write $\exp(\Lambda_j(k)) = |\det \tau_k|$. This can be expressed in an alternative way by saying that $\check{\tau}_k(\mathrm{Haar}\, H_j) = \exp(\Lambda_j(k))\, \mathrm{Haar}\, H_j$. The Haar measure is of course the Euclidean measure of the corresponding vector space and $k \to \exp(\Lambda_j(k))$ are homomorphisms $K \to \mathbb{R}_*^+$. The fact that, in (2.20), $\mathfrak{k} \cong \mathfrak{q}/\mathfrak{h} \oplus \mathfrak{s}$ splits means that we can find a covering map $\widetilde{K} = V \oplus S \to K$, where V is the Euclidean space that corresponds to $\mathfrak{q}/\mathfrak{h}$, and S is semisimple and simply connected (see Varadarajan, 1974, §2.6). The Λ_j can be identified with $\Lambda_j \colon \widetilde{K} \to \mathbb{R}$ and by the semisimplicity $\Lambda_j \equiv 0$ on S, since the only homomorphic images of semisimple groups are semisimple (see Jacobson, 1962, §III.5). From this it follows that we can identify $\Lambda_j = d_j L_j$ for the real roots of (2.4) and the dimension d_j of the corresponding root space of (2.6).

Exercise 2.4 Prove that $\Lambda_j = d_j L_j$. To see this, use the exponential mapping (see Varadarajan, 1974, §2.10) to identify $\mathfrak{h}$ with H and $\mathfrak{h}_j$ with H_j, and this identifies the corresponding Lebesgue measures. Then we have $\exp(\mathrm{Ad}(k)\xi) = k\exp(\xi)k^{-1}$, $\xi \in \mathfrak{g}$, $k \in V$ and also $\mathrm{Ad}(\exp(\zeta)) = e^{\mathrm{ad}\,\zeta}$ for $\zeta \in \mathfrak{q}/\mathfrak{n}$ which is the Lie algebra of V (see Varadarajan, 1974, §2.13, for all these facts). From this we see that $\det \tau_k$ is the product of characteristic roots of the action $\mathrm{Ad}\, k$ on H_j and this is equal to $\exp(\lambda_{\alpha_1}(\zeta) + \cdots + \lambda_{\alpha_s}(\zeta))$ when $k = \exp(\zeta)$ and $\lambda_{\alpha_1}, \ldots$ are the roots (2.2) that have the same real part L_j as in (2.5). Since the number of these roots is $s = d_j$ the assertion follows.

2.6.3 The C-condition on the exact sequence

Let us recapitulate what we have done in the previous subsection and consider some locally compact group G and $H \lhd G$, some closed normal subgroup that is a Euclidean space and that can be decomposed into subspaces $H = H_1 \oplus \cdots \oplus H_p$ where each H_j is a normal subgroup. The group $G/H = K$ acts then by inner automorphisms on each H_j and we obtain homomorphisms $\Lambda_j \colon K \to \mathbb{R}$ as above by the condition $\check{\tau}_k(\operatorname{Haar} H_j) = \exp(\Lambda_j(k)) \operatorname{Haar} H_j$, where $k \to \tau_k, K \to \operatorname{Aut} H$ denotes the action of K on H. Furthermore, we shall restrict ourselves to the case where $K = V \times S$ as in §2.6.2, and then the Λ_j can be identified elements of V^*.

Definition 2.5 We say that the above exact sequence (2.19), $0 \to H \to G \to K \to 0$, *satisfies the* C*-condition* if the homomorphisms Λ_j satisfy the C-condition, that is, $\sum \alpha_j \Lambda_j = 0$ for some $\alpha_j \geqslant 0$ with at least one $\alpha_{j_0} \Lambda_{j_0} \neq 0$.

What we saw in our previous considerations is that a C-Lie group, as in the special case of the C-theorem of §2.6.2, gives rise to the C-exact sequence (2.19).

What has to be done now is to show that when $H \lhd G$ give rise to a C-exact sequence then the estimate (2.10) of the special case of the C-theorem of §2.6.2 holds for the group G. The remainder of this chapter will be devoted to this task.

A generalisation It is of interest to extend this definition without imposing any special structure on K. We then have $\check{\tau}_k(\operatorname{Haar} H_j) = \sigma_j(k) \operatorname{Haar} H_j$ for homomorphisms $\sigma_j \colon K \to \mathbb{R}_*^+$, the multiplicative groups of positive reals. The relation in the C-condition of §2.2.1 is then replaced by $\sigma_1^{\alpha_1} \sigma_2^{\alpha_2} \cdots \sigma_p^{\alpha_p} = 1$ for $\alpha_j \geqslant 0$ and this, under the C-condition, should imply that $\sigma_i^{\alpha_i} = 1$ for each i. We shall see in Chapter 6 that this definition applies to other natural classes of groups.

2.7 Notation, Heuristics and Disintegration of Measures

2.7.1 Probabilistic language

For any group G we use the standard notation $x^y = yxy^{-1}$, $x, y \in G$ for any $H \lhd G$ closed normal subgroup, and for the projection $\pi \colon G \to G/H = K$ we shall also use the notation $\dot{x} = \pi(x), x \in G$. For simplicity we shall assume first that H is a semidirect factor in $G = H \leftthreetimes K$ (see §2.3.2) where K is identified with a subgroup such that $H \cap K = \{e\}$ as in §2.3.2. With $x_j = h_j k_j \in G$, $j = 1, \ldots, h_j \in H$, $k_j \in K$, we can then write

$$s_n = x_1 \cdots x_n = (h_1 h_2^{\dot{s}_1} \cdots h_n^{\dot{s}_{n-1}}) \dot{s}_n = \tilde{h}_n \dot{s}_n. \tag{2.21}$$

We shall think of s_n as the partial sums of a random walk valued in G and then $k_1 k_2 \cdots k_n = \dot{s}_n \in K \subset G$ are also the partial sums of a random walk. From this point of view the first factor $\tilde{h}_n \in H$ has a 'long memory' and is not a Markovian process.

To elaborate further in this spirit, the above (time-inhomogeneous) random walk is determined by $\mathbb{P}[x_j \in E] = \mu_j(E)$ and G is locally compact and H is closed. Then the distribution of $s_n \in G$ is given by the convolution product $\mu^n = \mu_1 * \cdots * \mu_n$. To analyse that measure we are going to think of s_n as the product of $\dot{s}_n$ with the process $\tilde{h}_n \in H$, which although not Markovian does become Markovian if we *condition* on $\dot{s}_1, \dot{s}_2, \ldots$. This is the fundamental idea that permeates the whole theory.

Observe also that in the above expression (2.21), the semidirect product structure of G is not essential as long as H is Abelian because we can then define unambiguously $h^{\dot{k}} = khk^{-1}$ for any pre-image $\pi(k) = \dot{k}, h \in H$ (see §2.6.2). Rather than pursuing the probabilistic interpretation of μ^n, we shall give below the equivalent analytic notion of the *disintegration of measures* (see Bourbaki, 1963, §6.3) in the fibration $\pi\colon G \to G/H = K$ where all the groups considered are locally compact. At the end we shall return to probabilistic language again.

Suggestion Those readers who do not feel comfortable with sophisticated measure theory might like to think of discrete groups in the next section. Or alternatively, to restrict themselves to smooth measures as in §2.15 below.

2.7.2 The disintegration of measures, and notation

In this section we shall freely use standard notions and notation from measure theory, and follow Bourbaki (1963). To present the disintegration we shall consider a general exact sequence

$$0 \to H \to G \to G/H = K \to 0 \tag{2.22}$$

of locally compact groups, where H is Abelian. As before we shall denote $\dot{g} = Hg = gH \in K$, with $g \in G$, and if we fix g we can use the right (resp. left) identification of H with the coset $\dot{g}$ by $h \to hg$ (resp. $h \to gh$). If ν is some measure on H we shall denote by ν_g (resp. ${}_g\nu$) the image measure that we obtain on the coset by that mapping. For any measure ν on H we can use this identification to identify ν with a measure $\nu_{\dot{g}}$ (resp. ${}_{\dot{g}}\nu$) on $\dot{g}$. This identification depends on the representative $g \in \dot{g}$ but it is unique up to translations by elements of H. For notational convenience and when no confusion arises we shall drop the dots and denote these measures by ν_g and ${}_g\nu$. With this notation we have

$$(\nu^x)_x = {}_x\nu; \quad \lambda_x * {}_y\mu = (\lambda * \mu^{xy})_{xy}, \tag{2.23}$$

for measures ν, λ, μ on H, $x,y \in G$, and where ν^x is the image of ν by the inner automorphism $h \to xhx^{-1}$.

The measures μ on G can be written as vector integrals of measures on the fibres gH, that is, measures of the form ${}_{\dot{g}}\nu(\dot{g})$ against some measure λ on K. Here, to clarify the notation we may fix $p\colon K \to G$ some Borel section, that is, $\pi \circ p$ is the identity on K. Let $\nu(\dot{g})$ be some measure on H that depends on $\dot{g} \in K$ (i.e. $\nu(\dot{g})$ is a function of $\dot{g}\colon K \ni \dot{g} \to$ measures on H). We shall then translate ν, as explained before, to the measure ${}_{\dot{g}}\nu(\dot{g})$ on $p(\dot{g})H = \dot{g}$ (i.e. we use the translation from H to $\dot{g}$ given by $h \to p(\dot{g})h$). We can then write down the integral $\int_K {}_{\dot{g}}\nu(\dot{g})\, d\lambda(\dot{g})$ for measures λ on G/H and the basic fact is that every bounded measure μ on G can be represented this way. The representation is clearly not unique, but for positive measures uniqueness can be achieved if we specify $\lambda = \check{\mu}$ the direct image of μ by the mapping $\pi\colon G \to K$ of (2.22). The $\nu(\dot{g})$ are then essentially uniquely determined but they depend on the section p. We end up therefore with a representation as a vector integral

$$\mu = \int_K {}_{\dot{g}}\nu(\dot{g})\, d\check{\mu}(\dot{g}). \tag{2.24}$$

If $\mu \in \mathbb{P}(G)$ is a probability measure then $\check{\mu} \in \mathbb{P}(K)$ is also a probability measure and so are $\nu(\dot{g}) \in \mathbb{P}(H)$, $\dot{g} \in K$, $\check{\mu}$ almost everywhere. If we do not specify the Borel section p, these ν are uniquely determined only up to translation.

When $d\mu(g) = f(g)\, dg$, $f \in L^1$, then (see (2.62) below and Bourbaki, 1963)

$$d\check{\mu}(\dot{g}) = \check{f}(\dot{g})\, d\dot{g}, \quad \check{f}(\dot{g}) = \int_H f(gh)\, dh, \tag{2.25}$$

where $g \in G$ is some representative of $\dot{g}$. The measure $\nu(\dot{g})$ on H is then given up to translations in H by

$$\nu(\dot{g}) = {}_g f(h)\, dh; \quad {}_g f(h) = \frac{f(gh)}{\check{f}(\dot{g})}, \quad \text{taken to be 0 when } \check{f}(\dot{g}) = 0. \tag{2.26}$$

In (2.25) and throughout we maintain the notation of §1.1, and denote left Haar measure by $d\dot{g}$, dh for the groups K and H respectively. Formulas such as (2.25) and (2.26) can of course also be written with right Haar measure instead. Furthermore, in all the above, we have for simplicity dropped the qualifications that the formulas in question only hold 'almost everywhere'; for example, (2.25) holds for almost all $\dot{g} \in K$ for the Haar measure. This practice will be adopted throughout.

We shall convolve these measures and the translations by elements of H will commute with the convolutions because H is Abelian. More explicitly, let

$$\mu_i = \int_K {}_g\nu^i(g)\, d\check{\mu}_i(g); \quad i = 1, 2, \ldots \tag{2.27}$$

be a sequence of probability measures as above, where for notational convenience we have suppressed the dots and have written g rather than $\dot{g}$. When confusion does not arise this notational simplification will be tacitly used throughout.

The convolution product in G can then be written

$$\mu_1 * \cdots * \mu_n = \int_K \int_K \cdots \int_K {}_{g_1}\nu^{(1)}(g_1) * \cdots * {}_{g_n}\nu^{(n)}(g_n)\, d\check{\mu}_1(g_1) \ldots d\check{\mu}_n(g_n). \tag{2.28}$$

The integrand in the above multiple integral is uniquely determined up to translation by elements of H and the factors in the convolution product are identified to measures of G which are then convolved in the group G.

To close the circle of ideas we shall go back to the random walk notation $s_j = g_1 \cdots g_j \in K$ where we suppress the dots above the elements of K (thus write s and g rather than $\dot{s}$ and $\dot{g}$), and transform the integrand of (2.28) as follows:

$${}_{g_1}\nu^{(1)} * {}_{g_2}\nu^{(2)} * \cdots * {}_{g_n}\nu^{(n)} = \Big((\nu^{(1)}(g_1))^{s_1} * (\nu^{(2)}(g_2))^{s_2} * \cdots * (\nu^{(n)}(g_n))^{s_n} \Big)_{s_n}. \tag{2.29}$$

Here the convolution product on the right has been taken in the group H and then, using right translation this time, that convolution product is identified with a measure on the coset Hs_n. The proof is but a repeated use of (2.23). This means that for fixed $s_1, s_2, \ldots, s_n$ (keeping these fixed is the conditioning that we alluded to in the previous subsection) we obtain a convolution in H. Up to translation, this is the distribution of $\widetilde{h}_n$ in (2.21). This conditioned process becomes therefore a random walk in H.

2.8 Special Properties of the Convolutions in H

Here $H = H_1 \oplus \cdots \oplus H_p$ is a Euclidean space. We shall decompose it in some fixed way into a direct sum of Euclidean subspaces. This was the set-up in

§2.6 in the definition of the C-condition for the exact sequence. In this section we shall describe a general method of estimating convolution products $\nu_1 * \nu_2 * \cdots * \nu_n$, $\nu_j \in \mathbb{P}(H)$, that take into account the product structure of H. Later, we shall apply this method to the specific convolution products that occur in (2.29). Let us assume that $\nu_j \leqslant (f_1^{(j)} \otimes \cdots \otimes f_p^{(j)})\, dh$ where the $f_k^{(j)}$ are L^1 functions. Then $\nu_1 * \cdots * \nu_n = F(h)\, dh$ and we can assert that $F \in L^\infty$ as long as for every $1 \leqslant k \leqslant p$ we can find $f_k^{(j_k)} \in L^\infty(H_k)$ (i.e. that the density of the kth factor is L^∞). We can put this in a quantitative estimate and assume that $\|f_i^{(j)}\|_1 \leqslant C$ for some $C > 0$. We then have

$$\|F\|_\infty \leqslant C_1 \|f_1^{(j_1)}\|_\infty \cdots \|f_p^{(j_p)}\|_\infty, \tag{2.30}$$

for some C_1 that depends on C and H. The important thing here and in what follows is that the constant C_1 is independent of n. To see this we use the fact that the ν_j are probability measures and we can ignore, in the product $\nu_1 * \cdots * \nu_n$, all the factors except those ν_j for which j is one of $j_1, \ldots, j_p$. A more precise form of (2.30) can be stated as follows. Let $\nu_1, \ldots$ be arbitrary, let the $f_k^{(j)}$ be as above and set

$$m_k = \inf_j \|f_k^{(j)}\|_\infty, \quad k = 1, \ldots, p. \tag{2.31}$$

Then (2.30) can be restated as follows:

$$\|F\|_\infty \leqslant C_1 m_1 \cdots m_p. \tag{2.32}$$

If $A \subset [1, 2, \ldots, n]$ is a subset, then by ignoring in $\nu_1 * \cdots * \nu_n$ all the factors with $j \notin A$ and by denoting instead

$$m_k = \inf_{j \in A} \|f_k^{(j)}\|_\infty, \quad k = 1, \ldots, p, \tag{2.33}$$

we obtain the same estimate (2.32) where here we make the assumption $\nu_j \leqslant f_1^{(j)} \otimes \cdots \otimes f_p^{(j)}$ only for $j \in A$.

We shall recapitulate and refine the above estimates slightly. Let $\nu_1, \ldots, \nu_n \in \mathbb{P}(H)$ and let $A \subset [1, \ldots, n]$ be some subset such that for every $j \in A$,

$$\nu_j \leqslant \left(f_1^{(j)} \otimes \cdots \otimes f_p^{(j)} \right) dh, \tag{2.34}$$

where we assume that $\|f_k^{(j)}\|_1 \leqslant C$ for some positive constant and $1 \leqslant k \leqslant p$, $j \in A$. Let us define m_k as in (2.33) (possibly $m_k = +\infty$) for $k = 1, \ldots, p$. There exist then $C_1 > 0$ such that we can estimate $\nu_1 * \cdots * \nu_n = F\, dh$ by

$$\|F\|_\infty \leqslant C_1 m_1 \cdots m_p. \tag{2.35}$$

In applying this formula it will be convenient to set $m_j^- = m_j \wedge 1 = \min[m_j, 1]$

and replace the right-hand side of (2.35) by $C_1 m_1^- \cdots m_p^-$. This can be done as long as we already have the additional information that $m_j \leqslant C$ for some appropriate constant and $1 \leqslant j \leqslant p$; as we shall see, in our applications this information comes for free.

2.9 The Reduction to the Random Walk Estimate

Here G, H satisfy the conditions of the C-exact sequence of (2.19), as in Definition 2.5, and $\mu_1, \ldots$ is a sequence of probability measures on G that satisfies conditions (i), (ii) and (iii) of §2.4.1. These measures will then be disintegrated $\mu_j = \int_{K\ g} \nu^{(j)}(g)\, d\check{\mu}_j(g)$ as in (2.27) and, with the notation of (2.28), we shall denote

$$\nu(g_1, \ldots, g_n) = (\nu^{(1)}(g_1))^{s_1} * \cdots * (\nu^{(n)}(g_n))^{s_n}; \quad g_1, \ldots, g_n \in K, \tag{2.36}$$

where, as before, the dots have been suppressed for the elements of K and these measures are defined only up to translation by elements of H. This is a convolution product on the Euclidean space $H = H_1 \oplus \cdots \oplus H_p$ and to be able to use estimates of the type (2.35) we must find a way to dominate appropriately $\nu^{(j)}(g) \leqslant C\big(f_1^{(j)} \otimes \cdots \otimes f_p^{(j)}\big)\, dh$ as in (2.34). It is clear that this is not possible for all $g \in K$. But by conditions (i), (ii) and (iii) that hold for the measures μ_j and their images $\check{\mu}_j = \check{f}^{(j)} d\dot{g}$ (see §2.5.2), it will follow from (2.26) that there exist constants $C, c_1, \ldots$ such that for the subsets $\Omega_j = [g;\ \check{f}^{(j)}(g) > c_1] \subset K$ (dots have been suppressed) we have $\check{\mu}_j(\Omega_j) > c_2$ and

$$\begin{gathered} \nu^{(j)}(g) \leqslant C\left(f_1^{(j)} \otimes \cdots \otimes f_p^{(j)}\right) dh, \\ \|f_k^{(j)}\|_1 + \|f_k^{(j)}\|_\infty \leqslant C; \quad g \in \Omega_j,\ j \geqslant 1,\ 1 \leqslant k \leqslant p. \end{gathered} \tag{2.37}$$

Furthermore, to be able to use the sharper estimate with the m_j^- instead, we shall use the reduction of §2.5.3 and assume that μ_1 has been fixed and is such that $\check{\mu}_1(\Omega_1) = 1$.

In the notation that follows, we shall again drop the ‘dot’ above the elements of K. For $g \in \Omega_j$ if we use the transformation $\nu^{(j)}(g) \to (\nu^{(j)}(g))^s$ for each factor of (2.36) and $s \in K$, we can dominate this new transformed measure by a similar product measure $\big(\tilde{f}_1^{(j)} \otimes \cdots \otimes \tilde{f}_p^{(j)}\big)\, dh$ where now $\|\tilde{f}_k^{(j)}\|_1 \leqslant C$ and by the definition of Λ_k in §2.6.2 we have

$$\|\tilde{f}_k^{(j)}\|_\infty \leqslant C\ \exp(\Lambda_k(s)); \quad 1 \leqslant k \leqslant p,\ s \in K,\ j \geqslant 1, \tag{2.38}$$

where C are appropriate constants. To see this we use the Jacobian of the transformation on H_k induced by $h \to h^s = shs^{-1}$, with $h \in H, s \in K$. In fact it is $-\Lambda_k$

that should be used but we do not need to start 'chasing' signs here because $(-\Lambda_1,\dots)$ in Definition 2.5 also satisfy the C-condition.

Let us go back to $\nu(g_1,\dots,g_n)$ of (2.36) for some fixed $g_1,\dots,g_n$ and let A be the subset of the integers $j=1,\dots,n$ for which $g_j\in\Omega_j$. As before, we denote $s_j=g_1\cdots g_j\in K$. Because of (2.38) we shall abusively use the same notation and denote $m_k=\inf_{j\in A}\exp(\Lambda_k(s_j))$ and $m_k^-=m_k\wedge 1$ as before. If we combine (2.35) and (2.38) we obtain

$$\nu(g_1,\dots,g_n)=F\,dh,\quad \|F\|_\infty\leqslant Cm_1^-\cdots m_p^-, \tag{2.39}$$

for an appropriate constant.

In view of the representation (2.28) what we need to show to finish the proof of our theorem in the special case of §2.6.2 is

$$\int_K\cdots\int_K m_1^-m_2^-\cdots m_p^-\,d\check{\mu}_1(g_1)\dots d\check{\mu}_n(g_n)\leqslant C\exp(-cn^{1/3});\quad n\geqslant 2, \tag{2.40}$$

for some $C,c>0$. And, as explained in the next few lines, the additional translation s_n in the right-hand side of (2.29) makes no difference.

The best way to verify that this implies (2.10) is to use the Harnack estimate of §§2.5.1–2.5.2 and estimate the scalar product of the measure in (2.28) with some smooth non-negative function with small support near $e\in G$ that looks like $\phi(h)\otimes\psi(k)$, $h\in H$, $k\in K$. In this notation we use a local section of $G\to K$ and we identify G (as a manifold) in a small neighbourhood of $e\in G$ with $H\times K$. In formal terms (using conditional expectations $\mathbb{E}(\cdot\|\cdot)$ and (2.39)) this scalar product can be estimated by

$$\begin{aligned}\langle\mu_1*\cdots*\mu_n,\phi\otimes\psi\rangle&=\int\langle(\nu(g_1,\dots,g_n))_{s_n},\phi\otimes\psi\rangle\,d\check{\mu}_1(g_1)\cdots d\check{\mu}_n(g_n)\\&\leqslant C\|\phi\|_1\int\psi(s_n)m_1^-\cdots m_p^-\;d\check{\mu}_1(g_1)\cdots d\check{\mu}_n(g_n)\\&\leqslant C\|\phi\|_1\int\mathbb{E}(m_1^-\cdots m_p^-\|s_n\in dk)\psi(k)\\&\leqslant C\|\phi\|_1\|\psi\|_\infty\mathbb{E}(m_1^-\cdots m_p^-).\end{aligned} \tag{2.41}$$

Presented like this the C-theorem becomes a problem on random walks on the group K. To wit, $g_1,\ldots\in K$ are independent random variables with distributions $\check{\mu}_j$ and $s_j=g_1\cdots g_j\in K$ is the corresponding random walk. The m_k are then random variables, and what we need is the estimate

$$\mathbb{E}(m_1^-\cdots m_p^-)=O(\exp(-cn^{1/3})). \tag{2.42}$$

This is, of course, estimate (2.40) written out in probabilistic terms. In (2.42) $\mathbb{E}$ is the expectation in the probability space of the paths of the random walk $s_j=g_1\cdots g_j\in K$ of §2.7.1.

2.10 The Random Walk and the Proof of the Theorem in the Special Case $G/H \cong \mathbb{R}^d$

Here we shall give a proof of (2.42) in the special case when $G/H = K$ is a Euclidean space. In the next section we shall see that the general case easily follows by slightly refining the method. We shall recapitulate and reformulate the problem in probabilistic terms as follows.

Let $U_1, U_2, \ldots \in V$ be a sequence of independent random variables that take their values in the Euclidean space V. We shall assume that $|U_j| \leqslant C$ and we shall also fix $\Omega_j \subset V$, a sequence of subsets such that $\mathbb{P}[U_j \in \Omega_j] > c$, where the C, c are appropriate constants. In what follows it will be convenient to rename $K = V$ and $K \ni \dot{g}_j = U_j$.

An additional condition has to be imposed on these variables, which at first sight seems artificial from the point of view of probability theory: we demand that the U_j have densities $\psi_j(x)\,dx$ and that $\psi_j(x) > \varepsilon_0$ for $|x| \leqslant \delta_0$ for some ε_0, $\delta_0 > 0$. This condition in our context is perfectly natural by (iii) of §2.4.1. This condition is exploited in the proofs in the appendix to this chapter by the fact that it implies that the characteristic functions, that is, the Fourier transforms $\phi_j = \hat{\psi}_j$ satisfy $\phi_j(0) = 1$, $|\phi_j(\xi)| \leqslant \exp(-c|\xi|^2)$ for $|\xi| \leqslant 1$ and $|\phi_j(\xi)| \leqslant 1 - \eta$ for $|\xi| \geqslant 1$ for some $c, \eta > 0$. Seen like this, the reader who has had some exposure to elementary probability theory recognises the basic prerogative of the proof of the central limit theorem.

Be this as it may, we shall further assume that $\Lambda_1, \ldots, \Lambda_p \in V^*$ are linear functionals that satisfy the C-condition of Definition 2.5 and therefore the $-\Lambda_p$ also satisfy the same condition. Here we shall adopt the equivalent formulation (2.7) and for some subspace $V_1 \subset V$, $V_1 \neq V$, we have

$$\begin{aligned} \Lambda(x) = \sum -\Lambda_k^-(x) \geqslant c|\hat{x}|_{V/V_1}; \quad & \Lambda_k^- = \Lambda_k \wedge 0,\ (-\Lambda_k)^+ = -\Lambda_k^-, \\ & x \in V,\ x \in \hat{x} \in V/V_1; \end{aligned} \tag{2.43}$$

here ˆ indicates the image in V/V_1. We shall denote

$$M_n = \sup_{\substack{1 \leqslant j \leqslant n \\ U_j \in \Omega_j}} \Lambda(S_j),\ S_j = U_1 + \cdots + U_j. \tag{2.44}$$

By (2.43), $M_n \geqslant \sup[|\hat{S}_j|_{V/V_1};\ 1 \leqslant j \leqslant n,\ U_j \in \Omega_j]$; and this by (2A.8), (2A.14) (a result that will be proved in the appendix) implies

$$\mathbb{E}\exp(-M_n) \leqslant C_1 \exp(-c_1 n^{1/3}), \tag{2.45}$$

where C_1, c_1 depend only on the variables U_j and the C-condition (2.7).

Now, in the expressions below, we shall take sup and inf to mean sup and inf as $1 \leqslant j \leqslant n$ with $U_j \in \Omega_j$. We have

$$-\sum_k \sup\big(-\Lambda_k^-(S_j)\big) \leqslant -\sup \Lambda(S_j) \tag{2.46}$$

and since $\inf \Lambda_k^-(S_j) = -\sup(-\Lambda_k^-(S_j))$ we conclude that

$$\sum_k \inf \Lambda_k^-(S_j) \leqslant -M_n. \tag{2.47}$$

If we switch back to the notation of §2.9 and pass to the exponentials, the left-hand side of (2.47) becomes the $m_1^- \cdots m_p^-$ of (2.39). And it follows that (2.42) is a consequence of (2.45). This completes the derivation of (2.10) for the special case §2.6.2 when $K = V$. From this, the C-theorem in §2.4.1 follows when G is soluble and simply connected for, in that case, $G/N = V$ is a Euclidean space.

2.11 The Random Walk and the Proof of the C-Theorem in the General Case

2.11.1 The idea of the proof

In the general case, as we already pointed out in §2.6.2, the Lie algebra of $K = G/H$ in (2.19), (2.20) is the product of an Abelian algebra with a semisimple algebra and locally $K \cong V \times S$, where V is a vector space and S is semisimple. Assume for simplicity that we have $K = V \times S$ globally. The argument that we gave in the previous section, when $K \cong V$, can be pushed through to this product case. The idea is simple enough: if $g_1, \ldots \in K$ are K-valued random variables and $s_j = g_1 \ldots g_j$ is the random walk, we can project $K \to V$, $g_j \to U_j$ and since the roots $\Lambda_1, \ldots, \Lambda_p$ for the exact sequence $0 \to H \to G \to K \to 0$ vanish on the semisimple group S (see §2.6.2) we are back to the probabilistic argument of the previous subsection. In this argument it is the condition $g_i \in \Omega_i \subset K$ in §2.9 that complicates matters because it is tricky in general to relate this to a corresponding condition $U_j \in \Omega'_j \subset V$. This can of course be done but it becomes essentially trivial when S is compact. It's here that amenability as defined in §2.2.2 simplifies the proof considerably. We shall therefore assume, as we may, that G is amenable (see the final remark of §2.4).

2.11.2 The proof

We shall assume first that in (2.19), $K = V \times S$ where $V \cong \mathbb{R}^d$ is a vector group and S is compact (we do not assume here that S is semisimple). This happens when G is soluble because then $G/N = K$ (where N is the nilradical) is Abelian and is therefore the product of a Euclidean space with a torus. This also happens is when G is simply connected, for then K is simply connected and the fact that it is amenable implies that S is semisimple of compact type (see §§2.2.2 and 2.4). In fact, this happens for all amenable groups, but a special ad hoc argument is needed to see this (cf. §F.3 in Part II on geometric theory).

Formula (2.28) is, as before,

$$\mu_1 * \cdots * \mu_n = \int_K \cdots \int_K {}_{g_1}\nu^{(1)}(g_1) \cdots {}_{g_n}\nu^{(n)}(g_n)\, d\check{\mu}_1(g_1) \ldots d\check{\mu}_n(g_n). \quad (2.48)$$

Now we shall exploit the fact that the cofactor S in K is compact to make a special choice of the sets Ω_j in §2.9.

For every $j = 1, 2, \ldots$ *we shall assume that there exists* $\Omega'_j \subset V$ *an open subset and* $c > 0$ *such that* $\check{\mu}_j(\Omega'_j \times S) > c$ *and such that for all* $g_j \in \Omega'_j \times S$ *the corresponding measure* $\nu^{(j)}(g)$ *in* (2.48) *satisfies conditions* (2.37) *for appropriate constants*.

What this assertion says is that the sets $\Omega_1, \ldots, \Omega_j, \ldots \subset K$ that are used in the proof are not arbitrary but can be modified and be assumed to be cylinders over sets $\Omega'_j \subset V$. By the compactness of S, this is certainly going to be the case if Ω_j is large enough, in the sense that it contains the ball $[|g| \leqslant C] \subset K$ for some appropriate large C. Conditions (i), (ii) and (iii) of §2.4.1 do not guarantee this of course, but if we block together $\mu_1 * \cdots * \mu_k$, $\mu_{k+1} * \cdots * \mu_{2k}$, $\mu_p^* = \mu_{pk+1} * \cdots * \mu_{(p+1)k}, \ldots$ products of k such measures, then the new measures μ_j^* will also satisfy the conditions, and the new Ω's in condition (iii) will be as large as we like provided that k is large enough. So this can be done and the additional assumption be made. Now the product of n factors μ_j^* is the product of the first nk factors μ_j. For the products of length $nk + r$, $r = 1, \ldots, k-1$, in between, we use the Harnack estimate of §2.5.2.

With this special choice of Ω's in the integral (2.48) we write $g_j = (U_j, \sigma_j) \in V \times S = K$ and the $U_1, U_2, \ldots \in V$ are independent random variables that satisfy the conditions of §2.10 and we write $S_j = U_1 + \cdots + U_j$.

Now, to estimate the integral (2.48), we write the integrand as the product (2.36) translated on the coset Hs_n as in (2.29). For each factor of (2.36) we still use (2.38) with Λ_k defined in §2.6.2 for the group $K = V \times S$, and Λ_k vanishes on S because S is compact. The σ_j coordinate can therefore be ignored and the estimate (2.39) holds with $m_k = \inf[\Lambda_k(S_j);\ U_j \in \Omega'_j,\ 1 \leqslant j \leqslant n]$ with these m_k. We must therefore prove (2.40) because m_k can be replaced by m_k^- by the same argument as before (see §2.5.3). The probabilistic interpretation is again (2.42) with $\mathbb{E}$ referring to the random walk on K. But since the m_k depend only on the U_j, this is the same as the expectation on the path space of $S_j \in V$ and then the estimate (2.42) is a consequence of (2.43)–(2.47) together with the results of the appendix.

2.11.3 An exercise in Lie groups

This subsection is redundant if you use §F.3 from Part II, but it is 'fun' if you enjoy manipulating Lie groups.

To reduce the proof for the general amenable group G to the special case of the simply connected groups of §2.11.2, we consider the simply connected covering group $\pi\colon \widetilde{G} \to G$ and the corresponding exact sequence as in §2.6,

$$0 \longrightarrow \tilde{H} \longrightarrow \widetilde{G} \longrightarrow \tilde{K} = \frac{\widetilde{G}}{\tilde{H}} = V \times S,$$

where $\tilde{H}$ is the nilradical of $\widetilde{G}$ and is therefore a closed simply connected subgroup (see Varadarajan, 1974, §3.18.2). Exactly as in §2.11.2, $V \cong \mathbb{R}^d$ is a Euclidean space and S is some compact semisimple group. Furthermore, since π is a covering map from $\tilde{H}$ onto H, and H is a Euclidean space, π reduces to an isomorphism $\pi\colon \tilde{H} \to H$. We shall denote by $\alpha\colon \widetilde{G} \to \widetilde{G}/\tilde{H} = \tilde{K}$ the canonical projection and let $Z = \ker \pi$. We shall need to show that

$$\alpha(Z) = \tilde{Z} \subset \tilde{K} \ \text{ is a closed subgroup.} \tag{2.49}$$

We shall assume this for the moment and complete the required reduction as follows.

First of all $\tilde{Z}$ lies in the centre of $\tilde{K}$. This centre is $V \times Z(S)$ where $Z(S)$ is the centre of S and thus, by the semisimplicity of S, is a finite subgroup. This means that $\tilde{Z}_1 = \tilde{Z} \cap V$ is a subgroup of $\tilde{Z}$ of finite index and therefore $\alpha^{-1}(\tilde{Z}_1) \cap Z = Z_1$ is a subgroup of Z of finite index.

We can then consider the group $G_1 = \widetilde{G}/Z_1$ which is intermediate between $\widetilde{G}$ and G in the sense that π factorises as

$$\widetilde{G} \underset{\pi_1}{\longrightarrow} G_1 = \frac{\widetilde{G}}{Z} \underset{\pi_2}{\longrightarrow} G. \tag{2.50}$$

We can also use π_1 to construct the corresponding exact sequence as before:

$$0 \longrightarrow H_1 \longrightarrow G_1 \longrightarrow K_1 = \frac{G_1}{H_1} \longrightarrow 0. \tag{2.51}$$

Here H_1 is the nilradical of G_1 and is therefore closed (see Varadarajan, 1974, §3.18.13). Since the isomorphism $\pi\colon \tilde{H} \to H$ also factorises into $\tilde{H} \underset{\pi_1}{\to} H_1 \underset{\pi_2}{\to} H$ and π_1, π_2 are covering maps, it follows that H_1 is also a Euclidean space. From the definitions, on the other hand, it follows that $K_1 \approx \big(V/\tilde{Z}_1\big) \times S$. Since on the other hand $V/\tilde{Z}_1$ is a product of a Euclidean space with a torus, the group K_1 is the product of a Euclidean space with a compact group as in §2.11.2.

What we have therefore in (2.51) is an exact sequence that satisfies all the conditions of the exact sequences used in §2.11.2. The C-theorem (Theorem 2.3) is therefore valid for the group G_1.

Finally, in the covering map $\pi_2\colon G_1 \to G$ of (2.50) the kernel $\ker \pi_2 = Z/Z_1$ is a finite group. The reduction (2.17) therefore applies, and therefore the C-estimate (2.10) also holds for the original group G. And we are done.

It remains to give the proof of (2.49). This is a consequence of the following.

Lemma *Let G be some locally compact group and let $Z, H \subset G$ be two closed normal subgroups. Let $\zeta : G \to G/Z$, $\chi : G \to G/H$ denote the corresponding canonical projections. The subgroup $\chi(Z) \subset G/H$ is closed if and only if the subgroup $\zeta(H) \subset G/Z$ is closed.*

The lemma says that the property is symmetric with respect to Z and H. To see this we just have to notice that the subgroup $\chi(Z)$ is *not* closed if and only if we can find two subsequences $z_1, z_2, \ldots \in Z$ and $h_1, h_2, \ldots \in H$ that tend to infinity such that $z_n h_n^{-1} \to 0$ in G. This situation is symmetric with respect to the two subgroups Z and H.

Part 2.2: The Heat Diffusion Kernel and Gaussian Measures

2.12 The Heat Diffusion Semigroup

In the rest of this chapter we are concerned with the heat diffusion kernel and diffusion measures. Let $X_1, \ldots, X_m$ be left-invariant vector fields on the connected Lie group G as in §1.3.2 and suppose that $\triangle = -\sum X_j^2$ is subelliptic, that is, we shall assume that the fields satisfy the Hörmander condition which means that these fields, together with their successive brackets, span the tangent space (see §1.3.2). For simplicity the reader may assume that the X_j already span the tangent space and that $\triangle$ is in fact elliptic because for the global behaviour nothing much changes. The operator $\triangle$ is formally self-adjoint on $L^2(G; d^r g)$ because, for the L^2 scalar product, we have $\langle X_j f_1, f_2 \rangle = -\langle f_1, X_j f_2 \rangle$. Note that with left-invariant fields we have to use right-invariant measure. We then close $\triangle$ in L^2 and obtain a positive operator in L^2 that generates the corresponding semigroup $T_t = e^{-t\triangle}$. An adequate reference for this and for the basic facts that we shall describe below is Varopoulos et al. (1992).

The aim of this second part of the chapter is to give the classification of §1.3

in terms of this semigroup. This is not in the mainstream of the subject and the reader who so wishes can skip the rest of this chapter and also anything else in the book that has to do with the heat diffusion semigroup T_t.

In the next two subsections we shall state without proofs the main properties of the semigroup that will be needed. With the use of these properties we shall see that the rest of this chapter is but a natural and interesting variant of what we have done in the first part for compactly supported measures. What makes it a 'variant' is that the general strategy of the proofs is the same. What makes it 'interesting' is that we have to make several modifications in the proofs that make them simpler in some ways but harder in others.

2.12.1 Heat diffusion kernel and the Harnack principle

There exists $\phi_t(g)$, a smooth positive function in $t > 0$ and $g \in G$, and probability measures $d\mu_t = \phi_t(g)\, d^r g = \phi_t(g)m(g)\, dg$; because of the self-adjointness of the operators, these satisfy

$$T_t f(x) = f * \mu_t = \int_G f(xy^{-1})\, d\mu_t(y) = \int_G \phi_t(y^{-1}x) f(y)\, dy; \qquad \mu_t(y^{-1}) = \mu_t(y). \tag{2.52}$$

It follows that $u(x,t) = \phi_t(x) = T_t \delta_e(x)$ for the Dirac mass at $e \in G$ satisfies the heat equation

$$\frac{\partial}{\partial t} u = \sum X_j^2 u; \quad t > 0,\ x \in G. \tag{2.53}$$

This is the direct analogue of the classical heat equation $\left(\frac{\partial}{\partial t} - \sum \frac{\partial^2}{\partial x_i^2}\right) u = 0$ in the Euclidean space $(x_1, \ldots, x_d) \in \mathbb{R}^d$.

Just as in the classical situation, any positive solution of (2.53), $u(t,g)$, then satisfies the parabolic Harnack estimate which here is, for $C > 0$ and $0 < a_i$, $1 \leqslant i \leqslant 3$, there exists C_0 such that

$$u(xy,t) < C_0\, u(x,t+\tau); \quad x \in G,\ |y|_G \leqslant C,\ t \geqslant a_1,\ \tau \in [a_2, a_3]. \tag{2.54}$$

This is a local result and it is first proved for $x = e$ and then the invariance of the operator by left translation gives (2.54) for all x. This Harnack estimate is standard when $\triangle$ is elliptic but perhaps less so in the subelliptic case (see Varopoulos et al., 1992, §III.2). When $u(x,t) = \phi_t(x)$ this gives the direct analogue of (2.12). To see this use the fact that $\phi_t(x^{-1}) = \phi_t(x)m(x)$ and this allows us to pass from xy to yx in (2.54) – here we use $m(xy) = m(x)m(y)$.

2.12.2 Gaussian measures

For a locally compact group G we say that the function $\phi(g)$ is Gaussian if there exist positive constants such that

$$C_1 \exp(-c_1 |g|_G^2) \leqslant \phi(g) \leqslant C_2 \exp(-c_2 |g|_G^2); \quad g \in G, \tag{2.55}$$

and we say that a measure $d\mu = \phi \, d^r g = \phi(g) m(g) \, dg$ is Gaussian, or $\mathrm{Gs}(G)$ for short, if ϕ is Gaussian. The C_i, c_i are then called the Gaussian constants. Since m is a multiplicative character it satisfies $m(g)^{\pm 1} = O(e^{c|g|})$ and therefore in the above we can use indiscriminately the right or left Haar measure. The fundamental fact here is that the measures μ_t and the kernels ϕ_t of the semigroup T_t in (2.52) are Gaussian with constants that stay uniform when $t \in [a, b]$ for fixed $0 < a < b$ (see Varopoulos et al., 1992).

2.13 The C-Theorem

Here we shall assume that G is some connected C-Lie group that is amenable. The theorem below says that in the C-theorem of §2.4 we can replace conditions (i)–(iii) of §2.4.1 for the measures, by the Gaussian condition.

2.13.1 The Gaussian C-theorem

Theorem *Let us assume that $\mu_1, \ldots \in \mathbb{P}(G)$ are Gaussian with uniform constants and let us denote $\mu^n = \mu_1 * \cdots * \mu_n$, $n \geqslant 1$ as in §2.4.1. Then for every $P \subset G$, a compact subset, there exist $C, c > 0$, which depend also on the Gaussian constants, such that*

$$\mu^n(P) \leqslant C \exp(-cn^{1/3}); \quad n \geqslant 1. \tag{2.56}$$

This is the analogue of (2.15) of §2.4. No pointwise estimate like (2.10) can a priori be deduced for the density of μ^n because we do not have at our disposal the analogue of the Harnack estimate (2.12) in this generality. For the heat diffusion kernel on the other hand, (2.54) can be used and from (2.56) we deduce the estimate

$$\phi_t(e) \leqslant C \exp(-ct^{1/3}); \quad t > 1. \tag{2.57}$$

Nonetheless, in the general case, the c in the exponent of (2.56) can be chosen to be independent of P.

2.14 Distances on a Group and the Geometry of Gaussian Measures

We shall need to prove some properties of Gaussian measures on the group G and to do that we shall need to recall first some general properties of the distance $|\ |$ of §1.1. We shall state these properties in the first part of this section and the proofs will be given in the second part.

2.14.1 Statements of the facts

We shall first recall some notation from §1.1. Let G be some locally compact, compactly generated group and let $H \subset G$ be some closed subgroup that is also assumed to be compactly generated. For every $h \in H$ we can then define $|h|_H$ as in §1.1 with h considered as an element of H but also $|h|_G$ of h considered as an element of G. It is obvious that there exist constants for which

$$|h|_G \leqslant C|h|_H + C; \quad h \in H. \tag{2.58}$$

It is not at all trivial to prove that for connected Lie groups G, H there exist constants for which we have the following converse inequality:

$$|h|_H \leqslant \exp(C|h|_G + C); \quad h \in H. \tag{2.59}$$

Fortunately, only easy cases of this will be needed and these will be proved below. One such is when G is nilpotent and simply connected. In that case, the exponential mapping from the Lie algebra to the group is a global diffeomorphism (see Varadarajan, 1974, §3.6) and furthermore if $|\xi|$ denotes some Euclidean norm on the algebra there exist constants such that

$$C|\xi| \geqslant |\exp(\xi)|_{\text{Group}} \geqslant c|\xi|^{c}; \quad \xi \in \text{Lie algebra}, \ |\xi| \geqslant C. \tag{2.60}$$

From this we see that there exist constants for which we have

$$|h|_H \leqslant C(|h|_G + 1)^C; \quad h \in H. \tag{2.61}$$

When G is nilpotent and simply connected and H some closed subgroup then H is compactly generated and (2.61) holds. In Ragunathan (1972, §§2.5–2.10) one finds proofs of this and a connected closed subgroup $\hat{H} \supset H$ is constructed with $\hat{H}/H$ compact. This and many other results that one can find in Varopoulos (1999a) (a very technical and reader-unfriendly reference!) are outside the scope of this book because, among other things, essential use is made of the theory of algebraic groups.

Exercise 2.6 If $H \subset G$ is normal then show that the distance on G/H is the quotient distance. This means that $|\dot{g}|_{G/H} = \inf |g|_G$ as $g \in \dot{g} \subset G$.

We recall from §1.1 the notation $\gamma(r) = \gamma_G(r)$, which is the Haar measure of G of B_r where $B_r = [g;\ |g|_G \leqslant r]$. It is well known, and we shall give the proof below, that for any locally compact group, $\gamma_G(r) \leqslant Ce^{cr}$ for appropriate constants. Similarly we can define $\tilde{\gamma}_H(r)$ as the Haar measure of H of $B_r \cap H$, the relative volume growth. Once more, by the involution $x \mapsto x^{-1}$, we see that in the definition we can take either the right or the left Haar measure. For appropriate constants we have again $\tilde{\gamma}_H(r) \leqslant Ce^{cr}$. We shall only need this in the case of a normal subgroup and in that case shall use the standard formula (see Bourbaki, 1963, Chapter 7, §2.7; Helgason, 1984, §1.1.2; Ragunathan, 1972, Chapter 1)

$$\int_G f(g)\, d_G g = \int_{G/H} d_{G/H}\dot{g} \int_H f(gh)\, d_H h; \quad f \geqslant 0, \tag{2.62}$$

for the left Haar measures of G, H and G/H. In the inner integral in (2.62), for any $\dot{g} \in G/H$, $g \in G$ is an arbitrary element for which $\pi(g) = \dot{g}$ for the canonical projection $\pi\colon\ G \to G/H$.

Now let $\phi(g)$ be some Gaussian function on G. The restriction ϕ to H is not in general Gaussian but when (2.59) holds, it clearly follows that there exist positive constants C_i for which

$$\phi(h) \leqslant \exp(-C_1 \log^2(|h|_H + 1) + C_2). \tag{2.63}$$

More generally, for H normal and $\dot{g} \in G/H$ we can use Exercise 2.6 to find $g \in G$ with $\pi(g) = \dot{g}$ as (2.62) and $|g|_G \leqslant 2|\dot{g}|_{G/H}$; we can then use the triangle inequality and for appropriate constants we have

$$\begin{aligned} |gh|_G^2 &\geqslant d_G^2(g, gh) - |g|_G^2 = |h|_G^2 - |g|_G^2 \\ &\geqslant c\log^2(|h|_H + 1) - |g|_G^2 - C; \quad g \in G,\ h \in H. \end{aligned} \tag{2.64}$$

We conclude therefore that

$$\phi(gh) \leqslant C\exp(c|\dot{g}|^2)\exp(-c\ \log^2(|h|_H + 1)); \quad g \in G,\ h \in H. \tag{2.65}$$

In §2.14.4 we shall use the exponential estimate on $\tilde{\gamma}_H$ and show that if $\mu \in \mathrm{Gs}(G)$ and H is a normal closed subgroup then $\check{\mu} = \check{\pi}(\mu) \in \mathrm{Gs}(G/H)$. The converse, when H is compact, is automatic and, for any $\check{\mu} = \check{f}\, d\dot{g} \in \mathrm{Gs}(G/H)$, the measure $\mu = \check{f}(\pi(g))\, dg$ is Gaussian on G.

None of the above facts, with the exception of (2.59) in full generality, are difficult but they do need proving. The proofs will be given below.

2.14.2 The distance distortion

For a locally compact group G we say that $H \subset G$ a closed subgroup is cocompact if there exists $P \subset G$, a compact subset such that $G = HP$. By considering $x \mapsto x^{-1}$ we also have then $G = P^{-1}H$ and P^{-1} is also compact.

Let $H \subset G$ be cocompact and let $h \in H$ and $e = g_0, g_1, \ldots, g_m = h$ a 'geodesic' in G in the same sense that $g_j \in G$, $m \leqslant |h|_G$ and $|g_{j+1}^{-1} g_j| \leqslant C$. We can then write $g_j = h_j p_j$, $h_j \in H$, $p_j \in P$, and then $|h_{j+1}^{-1} h_j|_H \leqslant C$. But then $e = h_0, h_1, \ldots, h_m, h$ is a 'geodesic' in H and this shows that $|h|_H \leqslant C|h|_G + C$. This means that we can reverse the inequality (2.58) and we then say that H is not distorted in G.

The only instance where we shall need the exponential distortion inequality (2.59) is the case where G is a soluble simply connected group and $H = N$ is its nilradical. Then what makes the proof of (2.59) easy is the fact that we can find $M \subset G$, some analytic subgroup that is nilpotent, closed and simply connected (a Cartan subgroup) such that $G = NM$. Given the one-to-one correspondence between closed connected subgroups of G with the subalgebras of the Lie algebra $\mathfrak{g}$ (see Varadarajan, 1974, 3.18.12), it follows that it suffices to find $\mathfrak{m} \subset \mathfrak{g}$ a nilpotent subalgebra such that $\mathfrak{g} = \mathfrak{n} + \mathfrak{m}$ for the nilradical $\mathfrak{n}$. This is a consequence of the general theory of Lie algebras and it will be explained in some detail in §3.4.2. This fact easily implies (2.59). To see how this is done we proceed as follows. By the definition of $|\ |_G$,

$$\begin{aligned} n = n_1 m_1 n_2 m_2 \cdots n_p m_p; \quad & p \leqslant |n|_G, \\ n_j \in N,\ m_j \in M,\ |n_j|,\ |m_j| \leqslant C. & \end{aligned} \tag{2.66}$$

Notice that for small distances it does not matter whether we consider distances in G or in a subgroup. We can then use (2.21) to write

$$\begin{aligned} & n = n_1 n_2^{M_1} n_3^{M_2} \cdots n_p^{M_{p-1}} M_p = N_p M_p; \\ & M_j = m_1 \cdots m_j \in M. \end{aligned} \tag{2.67}$$

It is clear that $|M_j|_M \leqslant cj$ and from this we can deduce

$$|n_j^{M_{j-1}}|_N \leqslant C \exp(cp). \tag{2.68}$$

To see this we use the general formula $\exp(\mathrm{Ad}(g)\xi) = g(\exp\xi)g^{-1}$, $g \in G$, $\xi \in \mathfrak{g}$ (see Varadarajan, 1974, 2.13), and combine this with the operator norm $\|\mathrm{Ad}(M_j)\| \leqslant C\exp(cj)$ for appropriate constants. We then use (2.60). On the other hand, $M_p \in N \cap M$ and since M is nilpotent it follows from (2.58), (2.61) and (2.67) that

$$|M_p|_N \leqslant C|M_p|_{N\cap M} \leqslant C(|M_p|_M + 1)^C \leqslant C(p+1)^C. \tag{2.69}$$

Putting together (2.67), (2.68) and (2.69) we obtain the required result. See also (8.11) for a variant of the same proof.

Since the radical of an amenable group is cocompact, we also see that with the above, we have a proof of (2.59) when G is amenable and H is the nilradical. Observe finally that because of (2.61), the same exponential distortion (2.59) also holds when H is any closed connected subgroup $H \subset N$.

Exercise To avoid having to use (2.61) for non-connected H, prove that in the above argument $N \cap M$ is connected. This is not a general fact but it holds here because N is normal and $G = NM$. We can argue as in §8.4.3 (in the geometric theory in Part II) to construct exponential coordinates $(\zeta_1, \dots, \zeta_r, \eta_1, \dots, \eta_p)$ on M (notation of §8.4.3) such that an element in M lies in N if and only if $\eta_1, \dots, \eta_p = 0$. The projection $G \to G/N$ could be used for this. The reader is invited to anticipate some of the geometric ideas from Part II and work out the details (cf. Exercise 8.9).

2.14.3 The volume growth

In the group G we denote balls by $B_r(g_0) = [g \in G;\ d_G(g, g_0) \leqslant r]$. In the ball $B_r = [|g| \leqslant r]$ centred at e we can choose $g_1, \dots, g_n \in B_r$ a finite subset that is maximal under the condition $d_G(g_i, g_j) \geqslant 10$, $i \neq j$. The balls $B_1(g_j)$ are disjoint and therefore $n \leqslant C\gamma(r+1)$. On the other hand, since any $g \in B_{r+10}$ can be written $g = g'g''$ with $g' \in B_r$ and $|g''| \leqslant 100$, and since by the definition of the g_j, $\bigcup_j B_{100}(g_j) \supset B_r$, it follows that g can in fact be written $g = g_k g'''$ for some $k = 1, \dots, n$ and $|g'''| \leqslant 10000$. The conclusion is that there exists $c > 0$ such that $\gamma(r+10) \leqslant c\gamma(r+1)$. The exponential bound $\gamma_G(r) \leqslant C\exp(cr)$ follows.

Now let $H \subset G$ be some normal closed subgroup and let us use formula (2.62) with $f = \chi_{r+2}$ the characteristic function of the ball B_{r+2} in G. For any $|\dot{g}| \leqslant 1$, by Exercise 2.6, we can then find g in (2.62) such that $|gh|_G \leqslant r+2$ for all $|h|_G \leqslant r$. For that choice of g, the inner integral on the right-hand side of (2.62) is therefore $\gtrsim \tilde{\gamma}_H(r)$. Since the left-hand side of (2.62) is $\gamma_G(r+2)$, the exponential estimate $\tilde{\gamma}_H(r) \leqslant Ce^{cr}$ follows.

2.14.4 The Gaussian estimates for the projected measure $\check{\mu}$

Let $d\mu = \phi(g)\,dg$, $d\check{\mu}(\dot{g}) = \check{\phi}(\dot{g})\,d\dot{g}$. Let us fix $\dot{g} \in G/H$ and $g \in G$ such that $\pi(g) = \dot{g}$ and $|g|_G \leqslant 2|\dot{g}|_{G/H}$. By formulas (2.25) and (2.62) we have $\check{\phi}(\dot{g}) =$

$\int_H \phi(gh)\,dh \leqslant C\int_H \exp(-c|gh|_G^2)\,dh$ and the integral can be estimated by

$$\sum_{N=0}^{\infty} \exp(-cN^2)[H- \text{ left Haar measure of } h\in H \\ \text{such that } N\leqslant |gh|_G \leqslant N+1]. \tag{2.70}$$

But, for large $|\dot{g}|$, by the choice of g such that $\pi(g)=\dot{g}$ as in Exercise 2.6, the second factor in (2.70) is zero unless $N\geqslant c|\dot{g}|$. If we use the exponential estimate on $\tilde{\gamma}_H$ and sum in (2.70) we obtain the upper Gaussian estimate for $\check{\mu}$.

The lower Gaussian estimate for $\check{\mu}$ is immediate from (2.55) because for appropriate constants and $|h|\leqslant C$ we have $\phi(gh)\geqslant c\exp(-c|\dot{g}|^2)$.

2.15 The Disintegration of Gaussian Measures

We shall consider again the exact sequence of (2.19),

$$0\to H\to G\to G/H=K\to 0. \tag{2.71}$$

Recall $H=H_1\oplus\cdots\oplus H_p$ and the Λ_j, $1\leqslant j\leqslant p$ from Definition 2.5 and assume that the C-condition is satisfied. All the notation in §§2.7–2.8 will be preserved. We shall consider $d\mu=\phi(g)\,dg\in \mathrm{Gs}(G)$ and disintegrate it as in (2.24):

$$\mu=\int {}_{\dot{g}}\nu(\dot{g})\,d\check{\mu}(\dot{g}),\quad \mathbb{P}(H)\ni\nu(\dot{g})= {}_g\phi(h)\,dh, \\ {}_g\phi(h)=\phi(gh)/\check{\phi}(\dot{g}),\quad \check{\phi}(\dot{g})=\int_H\phi(gh)\,dh, \tag{2.72}$$

where in the choice of the pre-image $g\in G$ such that $\pi(g)=\dot{g}$ for $\pi\colon\, G\to G/H$ we use Exercise 2.6 to guarantee that $|g|_G\lesssim 2|\dot{g}|_K$ and that we saw in §2.14.2, (2.65), (2.59) hold because H is the nilradical. This implies that there exist constants C,c and functions f_j on H_j such that

$$ {}_g\phi(h)\leqslant C\exp(c|\dot{g}|^2)f_1(h_1)\cdots f_p(h_p);\quad (h_1,\dots,h_p)\in H,\ \dot{g}\in K, \\ \|f_j\|_1+\|f_j\|_\infty\leqslant C;\quad j=1,2,\dots,p. \tag{2.73}$$

Notice here that, unlike what happens in (2.37), estimate (2.73) holds for all $\dot{g}\in K$ and not only for $\dot{g}\in\Omega\subset K$ in some subset with $\check{\mu}(\Omega)>c$. Apart from the upper Gaussian in (2.55) that is obviously used for (2.73), the lower Gaussian estimate in (2.55) for $\check{\mu}$ is also used here. Observe also that when $K=\mathbb{R}^d$, or at worst $\mathbb{R}^d\times$ (a compact group) (cf. §2.11.2), these Gaussian estimates for $\check{\mu}$ in the case of the heat diffusion kernels as in §2.12.2 come for free from the Euclidean case where we have explicit formulas, and we do not need the general theory of Varopoulos et al. (1992). See Varopoulos (1994a) for more details on this.

Now, we shall consider $\nu_1, \ldots, \nu_n \in \mathbb{P}(H)$ and $\dot{g}_1, \ldots, \dot{g}_n \in K$ such that for fixed constants we have

$$\begin{gathered} \nu_j \leqslant C \exp(c|\dot{g}_j|^2) f_1^{(j)} \otimes \cdots \otimes f_p^{(j)}\, dh; \quad 1 \leqslant j \leqslant n, \\ f_k^{(j)} \in L^\infty(H_k), \quad \|f_k^{(j)}\|_1 \leqslant C; \quad k = 1, 2, \ldots, p. \end{gathered} \tag{2.74}$$

As in (2.35) it then follows that $\nu_1 * \cdots * \nu_p = F(h)\, dh$,

$$\|F\|_\infty \leqslant C m_1 \cdots m_p; \quad m_k = \inf_{1 \leqslant j \leqslant n} \left[\exp\left(c|\dot{g}_j|^2\right) \|f_k^{(j)}\|_\infty \right]. \tag{2.75}$$

This will be applied to the convolution product of Gaussian probability measures $\mu_1 * \cdots * \mu_n$ which is represented as in (2.28). And with the same notation as in §§2.7, 2.9, we see that there exist constants for which

$$\begin{gathered} \nu(g_1, \ldots, g_n) = F\, dh, \quad \|F\|_\infty \leqslant C m_1, \ldots, m_p, \\ m_k = \inf_j \exp\left(c|\dot{g}_j|^2 + c\Lambda_k(\dot{g}_j)\right). \end{gathered} \tag{2.76}$$

This is the direct analogue of (2.39) and $\Lambda_k(\dot{g}_j)$ is the Jacobian as in (2.38). Here we have gained the fact that no sampling is required as in (2.39). The price that we had to pay is the additional factors $\exp(c|\dot{g}_j|^2)$ and the fact that we cannot replace Λ_k by $\Lambda_k^- = \Lambda_k \wedge 0$.

With this new definition of the m_k, what remains to be done is to prove the analogue of (2.40):

$$\int_K \int_K \cdots \int_K m_1 \cdots m_p\, d\check{\mu}_1(g_1) \ldots d\check{\mu}_n(g_n) \leqslant C \exp(-cn^{1/3}); \quad n \geqslant 1. \tag{2.77}$$

Once this estimate has been proved, the C-theorem follows for the group G in the exact sequence (2.71). To see this it suffices to use the analogues of (2.28) and (2.29) and take the scalar product of $\mu_1 * \cdots * \mu_n$ with a function of the form $\phi(h) \otimes \psi(k)$ as in (2.41).

Remark Since the compact set P in (2.56) is arbitrary, in the construction of the two functions ϕ and ψ of (2.41) we cannot use a local section of $G \to K$. A global section that is only Borel and not necessarily continuous, and that may also depend on P, has to be used. Nonetheless, G and $H \times (G/H)$ can be identified in a Borel fashion that way so that $\phi \otimes \psi$ is identically 1 on P and (2.41) goes through.

2.16 The Gaussian Random Walk on G/H and the Proof of the C-Theorem

We shall proceed as in the case of compactly supported measures. We shall first prove (2.56) in the special case when $G/H = V = \mathbb{R}^d$ is a Euclidean space and then $\Lambda_1, \dots, \Lambda_p \in V^*$ are linear functionals that satisfy the C-condition in §2.2.1. Let $U_1, \dots \in V$ be a sequence of independent random variables with distributions $\psi_k(x)\,dx, x \in V$, for which there exist uniform constants such that

$$C_1 \exp(-c_1|x|^2) \leqslant \psi_k(x) \leqslant C_2 \exp(-c_2|x|^2); \quad x \in V. \tag{2.78}$$

These are the Gaussian estimates (2.55) in V. We then define the random variables

$$A_n(\Lambda_k) = \inf_{1 \leqslant j \leqslant n} \exp(c|U_j|^2 + c\Lambda_k(U_j)) \tag{2.79}$$

for fixed constants. For the proof of (2.77) we need the analogue of (2.42) and what we need to prove now is the estimate

$$\mathbb{E}(A_n(\Lambda_1)A_n(\Lambda_2)\cdots A_n(\Lambda_p)) \leqslant C\exp(-cn^{1/3}); \quad n \geqslant 1, \tag{2.80}$$

for appropriate constants. This purely probabilistic estimate can be obtained from (2A.23) in the appendix to this chapter. The proof of our theorem in §2.14.1 with G as in (2.71) with $G/H = V$ is therefore complete.

Now, the procedure to prove the Gaussian C-theorem for a general amenable group is strictly identical to the case of compactly supported measures. To wit, the first step is to consider the exact sequence (2.71) when K is no longer Euclidean but is isomorphic to $V \times S$ where $V \cong \mathbb{R}^d$ and S is compact. The proof in that case is identical to the one given in §2.11.2; the argument applies verbatim in this Gaussian situation and is simpler because no sampling is involved. The additional facts used are that Gaussian measures projected on K are Gaussian and that for $k = (u,s) \in K$, we have far out $|k|_K \sim |u|_V$. The final reduction for the exact sequence (2.71) can be done as in §2.11.3 but at the very end of the argument we need to use the final remark of §2.14.1 on the lifting of Gaussian measures from Q/H to G when H is finite.

The last point that must be verified is that the reduction of a general amenable group to the exact sequence (2.19) that was done for compactly supported measures in §2.6.1 applies here also for Gaussian measures (see §2.13.1 on the Harnack estimate).

For this we make essential use of the fact that $\check{\mu}$ the image of a Gaussian measure μ by a projection $\pi\colon G \to G/H$ (with $\check{\mu} = \check{\pi}(\mu)$) is Gaussian. It follows from this that if the Gaussian C-theorem holds for the group G/H then it also holds for G. The reason is that for $\mu_1, \ldots, \mu_n \in \mathbb{P}(G)$, as in (2.56), we have

$$\mu_1 * \cdots * \mu_n(\pi^{-1}(P)) = \check{\mu}_1 * \cdots * \check{\mu}_n(P); \quad P \subset G/H. \tag{2.81}$$

We use the above and consider again $N \lhd G$, $T \lhd N$ and N_2 as §2.6.1. Proceeding as in §2.6 we can therefore reduce the problem to the exact sequence (2.19) and (2.71). This completes the proof.

2A Appendix: Probabilistic Estimates

The results in this appendix are exercises in 'elementary probability theory' and what I give in the next few pages are indications of how to achieve them. The reader who is not a 'fan' of that subject may simply register the results and forget about the proofs.

Let $U_1, U_2, \ldots \in V = \mathbb{R}^d$ be independent random variables and let $\mu_k \in \mathbb{P}(V)$ be their distributions and $\varphi_k = \hat{\mu}_k$ be their characteristic functions, that is, their Fourier transforms. We shall make the assumption that

$$|\varphi_k(\xi)| \leqslant \exp(-c|\xi|^2),\ |\xi| < 1; \quad |\varphi_k(\xi)| \leqslant 1 - \eta,\ |\xi| > 1, \tag{2A.1}$$

uniformly in k for some $c, \eta > 0$. This is certainly true if $d\mu_k = \psi_k\, dv$ with $\psi_k(x) > c$ for $|x - x_k| < c$, for some $x_k \in V$ and constants c, uniformly in k, because then $\psi_k(x) = \alpha\psi(x - x_k) + (1-\alpha)\widetilde{\psi}_k(x)$ for some $0 \leqslant \psi \in C_0^\infty$, $\widetilde{\psi}_k \geqslant 0$ both of integral $= 1$ and $0 < \alpha < 1$. In the case that we shall use (2A.1) we even have $x_k = 0$ and then the verification is even more obvious.

Conditions (2A.1) on the characteristic functions are standard in the proof of the central limit theorem. Here we shall define

$$S_n^* = \sup_{1 \leqslant k \leqslant n} |U_1 + \cdots + U_k|, \quad S_k = U_1 + \cdots + U_k, \tag{2A.2}$$

and use conditions (2A.1) to prove that there exist $C, c > 0$ such that

$$\mathbb{P}[S_N^* \leqslant M] \leqslant C \exp\left(-\frac{N}{cM^2}\right); \quad N, M \geqslant 1. \tag{2A.3}$$

Let $0 \leqslant \chi \in C_0^\infty(V)$ be such that $\chi(x) \geqslant 1$ for $|x| \leqslant 1$. For the Fourier transform

$\widehat{\chi}$ and $m = 1, 2, \dots, r > 0$ we then have, by Parseval's theorem,

$$\begin{aligned}\mathbb{P}[|U_1 + \cdots + U_{m^2}| < rm] &\leqslant \int \chi\left(\frac{x}{rm}\right) d\mu_1 * \cdots * \mu_{m^2}(x)\\ &\leqslant Crm \left| \int \varphi_1 \cdots \varphi_{m^2}(\xi)\widehat{\chi}(rm\xi)\, d\xi \right|\\ &\leqslant Crm \int \exp(-cm^2\xi^2)|\widehat{\chi}(rm\xi)|d\xi + (1-\eta)^{m^2} \|\widehat{\chi}\|_1\\ &\leqslant Cr + c(1-\eta)^{m^2} \leqslant \varepsilon_0; \quad m \geqslant C, \qquad (2A.4)\end{aligned}$$

for appropriate constants and some $\varepsilon_0 < 1$ provided that r is small enough because $\|\widehat{\chi}\|_\infty, \|\widehat{\chi}\|_1 \leqslant C$. By considering successive independent blocks $U_{jm^2+1} + \cdots + U_{(j+1)m^2}$, with $j = 0, 1, \dots, p-1$, we obtain

$$\mathbb{P}\left[S^*_{pm^2} \leqslant \frac{rm}{2}\right] \leqslant \varepsilon_0^{p-1}; \quad m, p = 1, 2, \dots, \qquad (2A.5)$$

where in (2A.4), (2A.5) the small values $m = 1, 2, \dots, m_0$ have to be examined separately, but we shall not actually need to use these small values in the argument that follows. Observe now that (2A.3) holds, for $M \leqslant c$ small and $N > C$ large, by the condition imposed on the U_j. Inequality (2A.3) also holds trivially if $\frac{N}{M^2} \leqslant C$. By choosing, in (2A.5), $m \sim M$ and $p \sim \frac{N}{M^2}$, we obtain then (2A.3) for the remaining values of M, N because r is fixed.

The estimate (2A.3) implies that for all $c > 0$ there exist $C, c_1, c_2 > 0$ such that

$$\begin{aligned}\mathbb{E}(\exp(-cS_n^*)) &\leqslant C \sum_{M=1}^{\infty} \exp\left(-c_2 M - c_2 \frac{n}{M^2}\right) \sim \sum_{M<n^{1/3}} + \sum_{M \geqslant n^{1/3}}\\ &\leqslant Ce^{-c_1 n^{1/3}}. \qquad (2A.6)\end{aligned}$$

In the above estimate n can clearly be assumed large.

2A.1 The sampling for bounded variables

The notation is as before and we assume that $A_k \subset V$, $k = 1, 2, \dots$ are such that $\mathbf{P}(U_k \in A_k) > c$. We shall then sample the supremum that was used to define the variable S_n^* and define the modified

$$\widetilde{S}_n = \sup_{\substack{1 \leqslant k \leqslant n\\ U_k \in A_k}} \left|U_1 + \cdots + U_k\right| \leqslant S_n^*. \qquad (2A.7)$$

We shall prove that we still have the estimate (2A.6) provided that we restrict ourselves to bounded variables $\left|U_j\right| \leqslant C$. More explicitly, for all $c > 0$, there exist C_1, c_1 such that

$$\mathbb{E}(\exp(-c\widetilde{S}_n)) \leqslant C_1 \exp(-c_1 n^{1/3}); \quad n \geqslant 1. \tag{2A.8}$$

This was essentially the estimate needed in (2.45). There is, however, a slight problem with the sampling in (2.45) that will be examined in the next subsection.

The argument that we shall use is combinatorial and elementary. Let $J = [1 \leqslant j_1 < j_2 < \cdots[$ be the set of integers defined by $j \in J$ if and only if $U_j \in A_j$. Let $J_n = J \cap [1, 2, \ldots, n]$. When $|J_n| \geqslant 2$ we shall define $G_n = \sup_{j_p \in J_n} (j_p - j_{p-1})$. It follows that there exist constants such that

$$\mathbb{P}[|J_n| \leqslant 10] \leqslant C \exp(-cn), \quad \mathbb{P}[G_n = k] \leqslant Cn \exp(-ck). \tag{2A.9}$$

To see this observe that for the event $[G_n = k]$ to hold there must exist some string of events $U_i \notin A_i, U_{i+1} \notin A_{i+1}, \ldots, U_{i+p} \notin A_{i+p}$ for $p = k-2$ and some $i = 1, 2, \ldots, n$. This proves the second estimate (2A.9). The proof of the first is similar and simpler.

On the set $[|J_n| \geqslant 10]$ we can write

$$[\widetilde{S}_n \leqslant m] = \bigcup_{k \leqslant n} \Big([\widetilde{S}_n \leqslant m] \cap [G_n = k] \Big) = \bigcup E_{n,k}. \tag{2A.10}$$

By the boundedness of the variables U_j, for fixed k on the event $E_{n,k}$ under the union sign, we have $S_n^* \leqslant \widetilde{S}_n + ck \leqslant m + ck$. If we use (2A.3) and (2A.9) we obtain therefore

$$\mathbb{P}(E_{n,k}) \lesssim \exp\left(-\frac{n}{c(m+k)^2} \right) \wedge n \exp(-ck), \tag{2A.11}$$

and if we sum in the two ranges $k \leqslant n^{1/3}$ and $k \geqslant n^{1/3}$ we see that this sum can be estimated by

$$n \exp\left(-\frac{n}{c(m+n^{1/3})^2} \right) + \exp(-cn^{1/3}). \tag{2A.12}$$

To obtain $\mathbb{E}\big(e^{-c\widetilde{S}_n};\ |J_n| \geqslant 10\big)$ we therefore multiply (2A.12) by e^{-cm} and sum in m. This gives the required estimate (2A.8) because for $[|J_n| \leqslant 10]$ we can use (2A.9).

2A.2 A variant in the sampling

Here we place ourselves in the set-up of §2.10 and we consider a projection $p\colon V \to V/V_1 = \widetilde{V} \neq \{0\}$ of the Euclidean space as in (2.44). Then the variables

$\hat{U}_k = p \circ U_k \in \widetilde{V}$ satisfy the same condition as the U_k and therefore, if we set $\hat{S}_n^* = \sup_{1 \leqslant k \leqslant n} |\hat{U}_1 + \cdots + \hat{U}_k|$, we have

$$\mathbb{E}(\exp(-c\hat{S}_n^*)) \leqslant C \exp(-cn^{1/3}). \tag{2A.13}$$

But we can also consider the following variant $\widetilde{S}_n = \sup_{1 \leqslant k \leqslant n, U_k \in A_k} |\hat{U}_1 + \cdots + \hat{U}_k|$ for subsets $A_j \subset V$ as before. The point here is that the sampling for the supremum is already done on the variables U_k before we project. The U_k are of course assumed again to be bounded. By exactly the same argument we obtain then the refinement that for all $c > 0$ there exist c_1, C such that

$$\mathbb{E}(\exp(-c\widetilde{S}_n)) \leqslant C \exp(-c_1 n^{1/3}); \quad n \geqslant 1. \tag{2A.14}$$

2A.3 Gaussian variables and the C-condition

We shall consider independent random variables $U_1, \ldots \in V$ valued in a Euclidean space that satisfy conditions (2A.1). But these variables will not be assumed to be bounded; we shall assume instead that there exist positive constants c_0, C for which the Gaussian estimate holds:

$$\mathbb{E}(\exp(c_0 |U_j|^2)) \leqslant C; \quad j = 1, 2, \ldots. \tag{2A.15}$$

We shall also consider $\Lambda_1, \Lambda_2, \ldots, \Lambda_p \in V^*$, linear functionals on V, and assume that they satisfy the C-condition as in the definition from §2.2.1 or (2.7). With $c > 0$ arbitrary but fixed and S_k as in (2A.2) we can then define

$$A_n(\Lambda_j^-) = \inf_{1 \leqslant k \leqslant n} \exp[c\Lambda_j^-(S_k)]; \quad \Lambda_j^- = \Lambda_j \wedge 0,\ 1 \leqslant j \leqslant p, \tag{2A.16}$$

and using (2A.6) by the same argument as in (2.43)–(2.47) we deduce that there exist C, c_1 such that

$$\begin{aligned} \mathbb{E}(A_n(\Lambda_1^-) \cdots A_n(\Lambda_p^-)) &\leqslant \mathbb{E} \inf_{1 \leqslant k \leqslant n} \exp(\Lambda_1^-(S_k) + \cdots + \Lambda_p^-(S_k)) \\ &\leqslant C \exp(-c_1 n^{1/3}); \quad n \geqslant 1. \end{aligned} \tag{2A.17}$$

No sampling is done in the definition of A_n. In the rest of this section we shall give an improvement of (2A.17).

Let Φ be some function on V; in fact, it will be $\Phi = \Lambda$ or $\Lambda^- = \Lambda \wedge 0$ for some linear functional $\Lambda \in V^*$. We shall set then

$$A[\Phi; p, q] = \inf_{p \leqslant k \leqslant q} \exp(\Phi(S_k)). \tag{2A.18}$$

Since for any $\Lambda \in V^*$ there exists $C > 0$ such that

$$\begin{aligned} \Lambda(S_1) \wedge \Lambda(S_2) \wedge \cdots \wedge \Lambda(S_n) &\leqslant 0 \wedge \Lambda(S_2) \wedge \cdots \wedge \Lambda(S_n) + C|U_1| \\ &\leqslant \Lambda^-(S_2) \wedge \cdots \wedge \Lambda^-(S_n) + C|U_1|, \end{aligned} \tag{2A.19}$$

it follows that

$$A(\Lambda_1;1,n)\cdots A(\Lambda_p;1,n) \leqslant \exp(C|U_1|)A(\Lambda_1^-;2,n)\cdots A(\Lambda_p^-;2,n). \quad (2A.20)$$

Estimate (2A.17) now applies in the product of the A's on the right-hand side of (2A.20) because the variable $S_2 = U_1 + U_2$ satisfies conditions (2A.1) and (2A.15). If we use these observations and Hölder we deduce from (2A.17) that there exist c, C such that

$$\mathbb{E}(A(\Lambda_1;1,n)\cdots A(\Lambda_p;1,n)) \leqslant C\exp(-cn^{1/3}); \quad n \geqslant 1. \quad (2A.21)$$

We shall finally consider the functional used in (2.79) and prove an improvement of (2A.21). For $c_1 > 0$ we set

$$A_n(\Lambda) = \inf_{1\leqslant k\leqslant n} \exp[c_1|U_k|^2 + \Lambda(S_k)]; \quad \Lambda \in V^*, \quad (2A.22)$$

and we need to show that there exist C, c such that

$$\mathbb{E}(A_n(\Lambda_1)\cdots A_n(\Lambda_p)) \leqslant C\exp(-cn^{1/3}); \quad n \geqslant 1. \quad (2A.23)$$

Hölder can be used to see this as long as c_1 is small enough – that is, $pc_1 \ll c_0$ of (2A.15) – because we can then deal with the additional cofactor $\exp(pc_1 \sup_{1\leqslant k\leqslant n} |U_k|^2)$ by (2A.15) and for the other factor in (2A.22) we use (2A.21) and then Hölder for the expectation of the product. The proof in the general case is more involved. (Observe, however, that for most of our applications in (2.80) this c_1 may be chosen as small as we like. The reason is that sharp Gaussian estimates for the heat diffusion kernel of §2.12.1 can be used. As explained in the few lines that follow (2.73), these are easy to obtain for the group K of (2.71) (see §E.1 for more on that).

We choose $N \geqslant 1$ to be specified later, and for $\alpha = 1, 2, \ldots$ we denote

$$I_\alpha = [\alpha N + 1, \ldots, (\alpha+1)N], \quad Y_\alpha = \inf_{j\in I_\alpha} |U_j|, \quad (2A.24)$$

and $j_\alpha \in I_\alpha$ the first integer for which $|U_{j_\alpha}| = Y_\alpha$. The proof of (2A.23) hinges on the following modifications of (2A.22):

$$\begin{aligned} B_n(\Lambda) &= \inf_\alpha [\exp(c_1 Y_\alpha^2 + \Lambda(S_{j_\alpha}));\ \alpha N \leqslant n], \\ D_n(\Lambda) &= \inf_\alpha [\exp(\Lambda(S_{j_\alpha}));\ \alpha N \leqslant n]. \end{aligned} \quad (2A.25)$$

We shall also need to use $\widetilde{A}_n(\Lambda) = A(\Lambda;1,n)$, which is the same as (2A.22) if we set $c_1 = 0$, and then we have the following comparisons:

$$\begin{aligned} A_{n+N}(\Lambda) \leqslant B_n(\Lambda) &\leqslant D_n(\Lambda)e^{c\xi_n^2}, \\ D_n(\Lambda) &\leqslant \widetilde{A}_n(\Lambda)e^{c\zeta_n}, \end{aligned} \quad (2A.26)$$

where $c > 0$ is large enough but independent of the choice of N and where we use the correcting variables

$$\begin{aligned} \xi_n &= \sup_\alpha [Y_\alpha;\ \alpha N \leqslant 2n], \\ \zeta_n &= \sup_\alpha [Z_\alpha;\ \alpha N \leqslant 2n], \quad Z_\alpha = \sum_{k \in I_\alpha} |U_k|, \end{aligned} \tag{2A.27}$$

and also assume that $n \geqslant 100N$. These variables are controlled as follows. It is clear that

$$\begin{aligned} &\mathbb{P}[Z_\alpha > \lambda] \leqslant C \exp(-c\lambda^2), \\ &\mathbb{P}[\zeta_n \geqslant \lambda] \leqslant Cn \exp(-c\lambda^2); \quad n = 1,2,\ldots,\ \alpha = 1,2,\ldots,\ \lambda > 0, \end{aligned} \tag{2A.28}$$

where C, c are independent of n, α but depend on the choice of N: in fact, $c \approx c_0/N$. On the other hand, with constants C_3, c_3 that are independent of n, α *and* N we have

$$\begin{aligned} &\mathbb{P}[Y_\alpha > \lambda] \leqslant C_3 \exp(-c_3 N\lambda^2), \\ &\mathbb{P}[\xi_n \geqslant \lambda] \leqslant C_3 n \exp(-c_3 N\lambda^2); \quad n = 1,2,\ldots,\alpha = 1,2,\ldots,\lambda > 0. \end{aligned} \tag{2A.29}$$

From (2A.29) it follows that for any preassigned k and $1 < q < +\infty$ we can choose N large enough so that

$$\| e^{k\xi_n^2} \|_q = O(n^{1/q}). \tag{2A.30}$$

On the other hand, for any fixed N we have from (2A.28),

$$\| e^{c\zeta_n} \|_q = O(n^{1/q}), \tag{2A.31}$$

for any $c > 0$ and $1 < q < +\infty$.

To finish the proof we use (2A.26) and for appropriate constants, independent of N, we have

$$A_{n+N}(\Lambda_1) \cdots A_{n+N}(\Lambda_p) \leqslant \widetilde{A}_n(\Lambda_1) \cdots \widetilde{A}_n(\Lambda_p) \exp(c\ \xi_n^2 + c\ \zeta_n); \quad n \geqslant 100N. \tag{2A.32}$$

Here we use Hölder and (2A.21): recall that $\widetilde{A}_n(\Lambda) = A(\Lambda; 1, n))$ and choose N appropriately so that (2A.30) applies. Our estimate (2A.23) follows.

3

NC-Groups

Overview of Chapter 3

As already explained, in this chapter we stick to amenable groups (i.e. $\lambda = 0$ in §1.3.1). Having proved the C-theorem in the previous chapter, here we shall prove the NC-theorem and we shall begin §3.2 by restating the theorem that is to be proved. In §3.2 we shall also give the easy reduction to simply connected groups and some important examples of NC-groups.

In this chapter the role played by amenability is important and, for this reason, in §3.1 we shall recall some of the definitions of that notion. It should be noted, however, that the original and historical definition, and the one that most people know, actually plays no role in this chapter and therefore we do not use it here. (That 'historical' definition says that the group G is amenable if it admits an invariant mean $M \in (L^\infty)^*$, i.e. $M\colon L^\infty(G) \to \mathbb{R}$ such that $M1 = 1$, $M \geqslant 0$ and M is invariant by, say, left translation on G.)

The next two sections §3.3 and §3.4 are special. In §3.3 essentially we reformulate the problem in terms of a terminology that is proper to Markov chains. For this, the reader does not have to know anything else except the definition of a Markov chain. Here, we consider only time-homogeneous chains that are induced by some Markovian operator T, which we take to consist of the convolution operator by some measure. In particular, we do not need to examine more general convolution products $\mu_1 * \mu_2 * \cdots$, as we did in the previous chapter.

In §3.4 we illustrate the criterion obtained in §3.3 by giving the proof of the theorem for some simple examples of groups. This section is important to understand because it contains the idea of the proof.

The actual general proof is carried out in §§3.4–3.5 and since these are broken up in a number of subsections the interconnection between these subsections is explained in the plan of the proof before we start.

This chapter, as the previous one, contains a Part 3.2 where we deal with heat diffusion kernels rather than convolution powers of measures, but there is also a third part that contains an alternative proof, or rather a variation of the proof given in §§3.4–3.5. This approach, if developed properly, will give sharp results (see (3.3) below), but the price that has to be paid is an additional algebraic lemma in §3.9.1 and, more seriously, the complications that arise from the fact that the exponential mapping on soluble groups is not necessarily bijective. From our point of view here, the reason for this alternative proof is that it adapts better to the general case in Chapter 5. The reader can defer reading the details here until they re-emerge and are used later.

Part 3.1: The Heart of the Matter

3.1 Amenability

In this first section we shall collect together the basic facts about the notion of amenability for a locally compact group that we shall need in the book. Reiter (1968) and Greenleaf (1969) are among the many references that can be used in the subject.

3.1.1 Preliminaries

Let $\mu \in \mathbb{P}(G)$ be some probability measure on the locally compact group G and let $T_\mu : L^2 \to L^2$ for $L^2(G; d^r g)$ denote the convolution operator $f \mapsto f * \mu = \int f(xy^{-1})\, d\mu(y)$. The operator norm $\|T_\mu\| \leqslant 1$ will also be denoted $\|\mu\|_{\text{op}}$ (it is easy to see that this is strictly positive, but this will not be used). The spectral radius $\|T_\mu\|_{\text{sp}} = \lim_n \|T_\mu^n\|^{1/n}$ will be denoted $\|\mu\|_{\text{sp}}$. Clearly, $\|\mu\|_{\text{sp}} \leqslant \|\mu\|_{\text{op}}$ with equality when μ is symmetric, that is, when $d\mu(x) = d\mu(x^{-1})$ for then $T_\mu^* = T_\mu$ is self-adjoint.

The following simple observation is basic in our definition of amenability. Let $\mu_1, \mu_2 \in \mathbb{P}(G)$ and let us suppose that there exist $0 < \alpha \leqslant 1$ such that we have $\mu_1 = \alpha\mu_2 + (1-\alpha)\lambda$ for some other probability measure λ. Alternatively, this says that $\mu_1 \geqslant \alpha\mu_2$ in the order relation of measures. Then it is clear that $\|\mu_2\|_{\text{op}} < 1$ implies that $\|\mu_1\|_{\text{op}} < 1$. This holds simply because these are norms dominated by the total mass norm.

3.1.2 Definition

Let G be some locally compact group. We then say that G is amenable if for all $\mu \in \mathbb{P}(G)$ we have $\|\mu\|_{\text{op}} = 1$. Note that when G is connected it is equivalent to say that for some $\mu \in \mathbb{P}(G)$ that satisfies conditions (ii) and (iii) of §2.4.1 we have $\|\mu\|_{\text{op}} = 1$. Notice that the compactness of the support is not used here.

Exercise Prove the equivalence. Assume that there exists *one* symmetric $\mu \in \mathbb{P}(G)$ that satisfies conditions (ii) and (iii) and for which $\|\mu\|_{\text{sp}} = 1$. Then by taking convolution powers of μ, and by the observations of §3.1.1, we see that the condition $\|\cdot\|_{\text{op}} = 1$ holds for *all* smooth probability measures. To see this we use conditions (ii) and (iii) and first consider compactly supported measures and then pass to the limit. Now, if instead we assume that *one* such measure exists for which $\|\mu\|_{\text{op}} = 1$, then again by §3.1.1 and condition (iii) we can find $\mu_1 \in \mathbb{P}(G)$ such that $\|\mu_1\|_{\text{op}} = 1$ as above and is in addition symmetric (i.e. $\mu_1(g) = \mu_1(g^{-1})$). Use Reiter (1968, §8.3.7) to get rid of the smoothness (but this last point will not be needed).

The definition we just gave is not the one most commonly used but is the one that best fits our purposes (see Reiter, 1968, §8.3.7). To illustrate this, observe that by the Harnack estimate of §2.5.1, if G is not amenable and if μ satisfies conditions (i), (ii) and (iii) from §2.4.1, then $\mu^{*n}(e) = O(e^{-\lambda n})$ for some $\lambda > 0$. More precisely, we have $\mu^{*n}(e) \leqslant C\|\mu\|_{\text{op}}^n$. The reason is that we can estimate the scalar product $\left|\langle T_\mu^n f, g\rangle\right| \leqslant \|\mu\|_{\text{op}}^n \, \|f\| \, \|g\|$. Here and in what follows (as we already did in §2.4), $\mu^{*n}(e)$ is an abuse of notation for $f^{(n)}(e)$ where $\mu^{*n} = f^{(n)} dg$ is as in §2.4.1. Here, μ^{*n} is the convolution product of n factors each equal to μ and we use (2.13). (There is no question any more of the abuse of notation of §2.4.1 where the factors are allowed to vary.)

3.1.3 Remarks

(i) For a general connected locally compact group, if μ is as in conditions (i), (ii) and (iii) in §2.4.1 and if μ is assumed symmetric then we can reverse the above inequality and $\|\mu\|_{\text{op}} = \overline{\lim}_n (\mu^{*n}(e))^{1/n}$. This is easy to prove by the spectral decomposition of T_μ. This last fact is folklore, though the author is unable to give a precise reference; it was met in Chapter 1 but it will not be again.

Exercise 3.1 Prove this: assume that for some $a \leqslant 1$, for every compact $K \subset G$ we have $\sup_K \mu^{*n}(g) = O(a^n)$. This implies in the spectral decomposition $T_\mu = \int_{-1}^{1} \lambda \, dE_\lambda$ that the projection of $L^2(K) = [f \in L^2;\ \operatorname{supp} f \subset K]$ on E_{-a} and $E_1 \ominus E_a$ is zero. Then use the fact that K is arbitrary. This argument also works for the actual limit.

The symmetry of the measures is essential in these types of considerations.

For, take $G = \mathbb{Z}$, which is amenable, and $\mu(-1) = \mu(0) = \varepsilon$, $\mu(1) = 1 - 2\varepsilon$. Then, for ε small, $\mu^n(0)$ decays exponentially.

(ii) As a by-product of the theory that will be developed in Chapters 4 and 5 (together with §6.1) we can see that for a connected locally compact group there exists C such that $\mu^{*n}(e) \leqslant C\|\mu\|_{\mathrm{sp}}^n$, but the proof is not trivial (see §5.2). The fact that this holds also for discrete groups is a consequence of the positivity of the functional $\mu \to \mu(e)$ in the C^*-algebra generated by $L^1(G)$ acting on $L^2(G)$. We then use the properties of these functionals (see Naimark, 1959, §10.3, but here it is $\|\mu\|_{\mathrm{op}}$ that has to be used).

3.1.4 Alternative definition of amenability

Let G be some locally compact group. We shall then define (see Weil, 1953, §11; Reiter, 1968, §3.5.1)

$$F = f * g(x) = \int_G f(xy^{-1})g(y)\, d^r y. \tag{3.1}$$

Then, by Hölder, $\|F\|_\infty \leqslant \|\check{f}\|_p \; \|g\|_q$, for any two conjugate indices $1/p + 1/q = 1$ and $\check{f}(x) = f(x^{-1})$. Then we say that G is amenable if we can find a sequence f_n, g_n continuous, real non-negative and compactly supported with $\|f_n\|_2 \leqslant 1$, $\|g_n\|_2 \leqslant 1$ such that $\check{f}_n * g_n \underset{n}{\to} 1$ uniformly on compacta. Here the $\| \ \|_2$ norms are taken in $L^2(G, d^r g)$. This is a strengthening of the condition $\langle f_n * \mu,\ g_n \rangle \to 1$ for $\mu \in \mathbb{P}(G)$; see Reiter (1968, §8.3.1).

The reader could consult Pier (1984) or Paterson (1988) to see how the above gives an 'approximate identity' in the 'Fourier algebra' and for further elaboration on this and other important aspects of amenability.

3.1.5 Lie groups

The criterion for amenability is simple when G is a real connected Lie group. Let $\mathfrak{g}$ be the Lie algebra and let $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ be the Levi decomposition of §2.1.2. Then G is amenable if and only if $\mathfrak{s}$ is of compact type (i.e. the simply connected Lie group that corresponds to $\mathfrak{s}$ is compact); see Reiter (1968, §8.7). Abusively, one then says that the algebra $\mathfrak{g}$ is amenable.

3.1.6 Quotients by amenable subgroups

Let G be a locally compact group and $\mu \in \mathbb{P}(G)$ and $H \lhd G$ some closed normal subgroup. Let $\pi \colon G \to G/H$ be the canonical projection and let $\check{\mu} = \check{\pi}(\mu)$ be the direct image measure. Then under the assumption that H is amenable we have $\|\mu\|_{\text{op}} = \|\check{\mu}\|_{\text{op}}$ and similarly for the spectral norm. The definition §3.1.4 can be used for this and the argument is identical to the argument that we shall later spell out in §4.4.4. This is related to the notion of transference (Coifman and Weiss, 1977).

We finally recall that closed subgroups and quotients of amenable groups are amenable; see Reiter (1968), Greenleaf (1969). Furthermore, if $H \lhd G$ is a closed subgroup such that both H and G/H are amenable, then so is G.

3.2 The NC-Theorem. A Reduction and Examples

3.2.1 The NC-theorem

Theorem 3.2 (The NC-theorem) *Let G be some real connected amenable* NC-*group and let μ be some symmetric probability measure* (*i.e.* $d\mu(x) = d\mu(x^{-1})$) *that satisfies conditions* (i), (ii) *and* (iii) *of* §*2.4.1. Then there exist positive constants C, c such that*

$$\mu^{*n}(e) \geqslant Cn^{-c}; \quad n \geqslant 2. \tag{3.2}$$

(The abuse of notation $\mu^{*n}(e)$ is as in the previous section.) The conclusion holds in particular if $\mu = \varphi\, dg$ with $\varphi \in C_0^\infty$. The reason for $n \geqslant 2$ and the abuse of notation $\mu(e)$ is as in §2.4.1.

In fact, a precise result can be proved (see Varopoulos, 1999b – sharp results are given there but unfortunately the paper is difficult and not very reader-friendly; see also the final note in §3A.6 for this problem): *with G and μ as above there exists $\nu = \nu(G,\mu) \geqslant 0$ that can be explicitly computed from the geometry of the roots of G and from μ such that*

$$cn^{-\nu} \leqslant \mu^{*n}(e) \leqslant Cn^{-\nu}; \quad n \geqslant 2 \tag{3.3}$$

for some C, c. The proof is an elaboration of the methods of this chapter. We shall give a proof of (3.2) but not of (3.3) – see §1.10. Estimate (3.3) is the generalisation of the local central limit theorem of $\mathbb{R}^d$ where we have $\nu = d/2$. As is well known, the symmetry of the measure is not essential for the local central limit theorem (see (Feller, 1968); (Woess, 2000)). It suffices that μ is *centred*, that is, $\int_{\mathbb{R}^d} x\, d\mu = 0$. The same extension can also be given here for the appropriate natural definition of centred measures; see Varopoulos (2000a). This extension will not be examined in this book. The index ν in (3.3) in general is not a rational number, let alone a half-integer. For ν to be a half-integer and independent of the particular measure, it is necessary and sufficient that in the radical of the group, the subspace spanned by $\mathscr{L}$ in V^* (notation of §2.2.1) has dimension 0 or 1 (see Varopoulos, 1996b).

3.2.2 A reduction

Let $\pi\colon G_1 \to G_2$ be a surjective homomorphism of locally compact groups and $\mu \in \mathbb{P}(G_2)$ be some probability measure on G_2 that satisfies conditions (i), (ii) and (iii) of §2.4.1. Then we can lift the measure μ to $\mu_1 \in \mathbb{P}(G_1)$ which satisfies those conditions on G_1 and its image on G_2 is $\check{\pi}(\mu_1) = \mu$. To see this we can use $\sigma\colon G_2 \to G_1$ a Borel section (i.e. an inverse of π such that $\pi \circ \sigma$ is the identity) for which $\sigma(K)$ is relatively compact for each compact subset $K \subset G_1$. This allows us to identify G_1 with $\sigma(G_2) \times H$, with $H = \ker \pi$, as Borel spaces, and to define $\mu_1 = \check{\sigma}(\mu) \otimes \nu$ for some appropriate smooth compactly supported $\nu \in \mathbb{P}(H)$. This $\mu_1 \in \mathbb{P}(G_1)$ satisfies the required conditions provided that σ has been chosen to be smooth in some neighbourhood of the identity, something that clearly is always possible. Observe also that by taking $\frac{1}{2}(\mu_1(x) + \mu_1(x^{-1}))$ we may assume that μ_1 is symmetric if μ is. From this and the Harnack estimate of §2.5 it follows that the NC-theorem holds for G_2 if it holds for G_1.

By taking the simply connected cover of a Lie group (see Varadarajan, 1974, §2.6) we conclude from the above observation that in proving the NC-theorem we may assume that G in the NC-theorem is simply connected. This observation is no longer applicable in the proof of the sharp central limit theorem (3.3), because the exponent ν changes from G_2 to G_1: this is one of the many technical difficulties in the proof of (3.3).

3.2.3 Examples and a special class of groups

We shall give here some examples of NC-groups on which we can easily apply the ideas that will be developed in this chapter. As we shall see, these examples are not very far from the general situation.

(i) NA-groups The simplest but also the most important examples of soluble simply connected groups Q are the ones for which N, the nilradical, is a semidirect factor as in §2.3.2. For these groups there exists $A \subset Q$, a closed subgroup that is isomorphic to a Euclidean space such that $Q = NA, N \cap A = \{e\}$. An alternative way of giving the definition is to say $Q \cong N \leftthreetimes A$. The groups considered in the examples in §2.3.2 are of that type with N Abelian. These groups are called NA-groups. Important examples of NA-groups with N non-Abelian occur in the theory of symmetric spaces and semisimple groups (see Helgason, 1978 and §4.3.1 later).

(ii) Algebraic groups Algebraic groups are only a sideshow in the theory. The next few lines can therefore be ignored by the reader who so wishes. For the definitions and for any of the results that will be used we shall use Chevalley (1951), which is the original historical reference for the subject; see also Varadarajan (1974, §2.1). Let K be some infinite field and V some n-dimensional vector space over K. Then $\mathrm{GL}(V)$ is the full linear group of V and $G \subset \mathrm{GL}(V)$ is called *algebraic* if there exist $p_1, \ldots, p_N$, a finite number of polynomials on the vector space of all the matrices $M_{n\times n}(K)$ over V such that $s \in G$, if and only if $p_j(s) = 0$, $j = 1, \ldots, N$. Such a group is called irreducible if the ideal of polynomials on $M_{n\times n}$ that vanish on G is a prime ideal.

To every algebraic group $G \subset \mathrm{GL}(V)$ we can associate its Lie algebra $\mathfrak{g}$ which is a finite-dimensional Lie subalgebra of $\mathfrak{gl}(V)$, the algebra of all K-linear transformations on V under the multiplication $[A,B] = AB - BA$. A number of complications occur however. To avoid these problems, in all our encounters with algebraic groups we shall make the assumption that the characteristic of K is zero. When the field $K = \mathbb{R}$, the algebraic groups that we have defined are but special cases of real Lie groups (possibly with a finite number of connected components; see Varadarajan (1974), Whitney (1958). Interesting new examples do occur, however, when $K = \mathbb{Q}_p$ is the field of p-adic numbers (see Chapter 6).

(iii) Real soluble algebraic groups Let $Q \subset \mathrm{GL}(V)$ be an irreducible algebraic group and let us assume that its Lie algebra $\mathfrak{q}$ is soluble (see §6.4.2 later on). When $K = \mathbb{R}$ this is essentially our old friend, a connected soluble real Lie group. More precisely, there exist $N, A \subset Q$, $N \cap A = \{e\}$ and $NA = Q$ and where N, A are irreducible algebraic groups, with A Abelian, and N is nilpotent (in the sense that its Lie algebras are). To conclude, if $K = \mathbb{R}$ and if Q is a soluble irreducible algebraic group, then the connected component of the identity in Q is an NA-group in the sense of (i) above.

3.3 The Principle of the Proof, an Example and the Plan

3.3.1 Convolutions of the kernel

Fix G and μ as in the NC-theorem, and $d\mu(g) = \phi(g)\, d^r g$ with $\int \phi(g)\, d^r g = 1$. More generally, we shall write $d\mu^{*n}(g) = \phi_n(g)\, d^r g$. By the symmetry $d\mu(x) = d\mu(x^{-1})$ we have $d\mu^{*n}(g) = \phi_n(g) m(g)\, dg$, $\phi_n(g^{-1}) = m(g)\phi_n(g)$ for the modular function $m(g)$; see §1.1. We can write

$$\begin{aligned} f*\mu^{*n}(x) &= \int_G f(xy^{-1})\,d\mu^{*n}(y) = \int_G \phi_n(y^{-1}x)f(y)\,dy \\ &= \int_G f(xy^{-1})\phi_n(y)\,d^r y, \\ \phi_{2n}(x) &= \int \phi_n(xy^{-1})\phi_n(y)\,d^r y, \end{aligned} \tag{3.4}$$

and therefore

$$\phi_{2n}(e) = \int \phi_n(g)\phi_n(g^{-1})\,d^r g = \int \phi_n^2(g)m(g)\,dg = \int \phi_n^2(g)\,d^r g \tag{3.5}$$

(to see this make the switch $g \leftrightarrow g^{-1}$ in the first integral). All the above hold under much more general conditions, for instance when μ is a Gaussian measure as in §2.12.2.

3.3.2 The probabilistic interpretation

The map $f \to T^n f = f*\mu^{*n}$ is a left-invariant (convolution) Markovian semigroup and it gives rise to a left-invariant Markov chain $z(n)\in G, n=1,\dots$ that is determined by $\mathbb{E}_x(f(z(n))) = T^n f(x)$. Here we shall use standard Markov chain notation (see for example Chung, 1982 or Williams, 1991) and when f is the characteristic function of $A\subset G$ this says that $\mathbb{P}_x(z(n)\in A) = \mathbb{P}[z(n)\in A//z(0)=x] = \int_A \phi_n(y^{-1}x)\,dy$. For $x=e$ this becomes

$$\mathbb{P}_e(z(n)\in A) = \int_A \phi_n(y^{-1})\,dy = \int_A \phi_n(y)\,d^r y. \tag{3.6}$$

Hölder can be used, and we end up with

$$(\mathbb{P}_e(z(n)\in A))^2 \leqslant \left(\int_G \phi_n^2(y)\,d^r y\right)|A|_r; \quad n\geqslant 1, \tag{3.7}$$

where $|\ |_r$ stands for the right Haar measure of the set A (not to be confused with $|g| = d(e,g)$). Combining this with (3.5) we finally deduce the inequality that is needed for the proof of the lower estimate in the NC-theorem:

$$\phi_{2n}(e) \geqslant \mathbb{P}_e^2(z(n)\in A)|A|_r^{-1}; \quad n\geqslant 1. \tag{3.8}$$

At this point we recall that μ is stable by the involution $x\mapsto x^{-1}$ and this in (3.6) implies that $\mathbb{P}_e(z(n)\in A) = \mathbb{P}_e(z(n)\in A^{-1})$. Hence in (3.8) we can replace A by A^{-1} in the first factor of the right-hand side. Since $|A^{-1}|_r = |A|_\ell$ is the left Haar measure of A, this means in (3.8) we can consider $|A|_r$ or $|A|_\ell$ indiscriminately. This observation, however, is not essential. Using (3.8) we shall formulate the principle on which the lower estimate (3.2) and all the other lower estimates in the theory are based.

3.3.3 General criterion

Let G be some locally compact group, let μ be a symmetric measure that satisfies conditions (i), (ii) and (iii) of §2.4.1 and let $z(n) \in G$, $n = 1, \dots$ be the Markov chain defined by the semigroup $f \mapsto T^n f = f * \mu^{*n}$ as above.

Let us assume that there exist positive constants $C, c, \dots$ and a sequence of subsets $A_1, \dots, A_n, \dots \subset G$ such that

(i) $|A_n| \leqslant Cn^C$, $n = 1, \dots$ (*here we could take either the left or the right Haar measure* $|\ |$ *of the sets*);
(ii) $\mathbb{P}_e[z(n) \in A_n] \geqslant cn^{-c}$, $n = 1, 2, \dots$.

Then there exist positive constants c_1, C_1 such that the kernel $d\mu^{*n}(g) = \phi_n(g)\, d^r g$ satisfies

$$\phi_n(e) \geqslant C_1 n^{-c_1}; \quad n \geqslant 2. \tag{3.9}$$

The $\phi_{2n}(e)$ does not need all conditions (i), (ii), (iii) of §2.4.1. It is the use of Harnack and the passage from even n to odd n that needs them.

Remark The criterion has been formulated in terms of the polynomial scale $n^{\pm c}$. We could have used a different scale, for example $\exp(\pm cn^{1/3})$, and with this we can obtain, for instance, the lower estimate that shows that the estimate (2.10) of the C-theorem is optimal for amenable groups. This observation will be exploited in Part 5.3 (in Chapter 5). More will be said on this general criterion in Chapter 5.

In the remainder of this section we shall illustrate the criterion and complete the proof of (3.2) for a special class of groups of type NA.

3.3.4 Illustration of the criterion in a special group

Here we shall consider the group $ax + b$ from §2.3.2(ii) and its natural generalisation $G = \mathbb{R}^d \leftthreetimes \mathbb{R} = H \leftthreetimes K$ in §2.3.2(iii), where K acts on H by the diagonal matrices

$$\mathbb{R} \ni t \mapsto \operatorname{diag}(\exp(\ell_1 t), \dots, \exp(\ell_d t)). \tag{3.10}$$

We impose the NC-condition, which means that we can assume that $\ell_1, \dots, \ell_d \geqslant 0$. A symmetric measure $\mu \in \mathbb{P}(G)$ that satisfies conditions (i), (ii) and (iii) of §2.4.1 is then given and we use the projection $\pi\colon G \to K \cong \mathbb{R}$ to define $\check{\pi}(\mu)$, a probability measure on $\mathbb{R}$ which is the density of a bounded centred real random variable X that is not degenerate (i.e. $X \neq 0$).

3.3.5 The gambler's ruin estimate

Let $X_1, X_2, \ldots$ be independent identically distributed centred real random variables, equidistributed with X as above, and let $S_j = X_1 + X_2 + \cdots + X_j$, $j \geqslant 1$. Then there exist $C, c > 0$ that depend only on X such that

$$\mathbb{P}_0[S_j < 0;\ j = 1, 2, \ldots, n] \geqslant Cn^{-c}; \quad n \geqslant 1. \tag{3.11}$$

This is a very coarse version of the classical gambler's ruin estimate (or of the probability of life – depending on your point of view: see Feller, 1968). In this coarse form the proof takes a couple of lines that will be given in the appendix to this chapter.

Let us now use the coordinates $(h,k) \in H \leftthreetimes K = G$ in §3.3.4. The right Haar measure can then be identified with the product of the two Lebesgue measures $d^r g = dh\,dk$, and the sets $A_n = [|h| \leqslant C_1 n] \times [|k| \leqslant C_1 n]$ have Haar measure $= cn^{d+1}$. The constants C_1 will be chosen appropriately large and to apply the criterion we write $z(n) = x_1 x_2 \cdots x_n = s_n \in G$ as in (2.21), where $x_j = h_j k_j \in H \leftthreetimes K$ are independent random variables with values in G and distribution μ. The $k_j = X_j \in K = \mathbb{R}$ are independent real symmetric random variables with distribution $\check{\mu} \in \mathbb{P}(\mathbb{R})$. With our previous notation we can identify $\dot{s}_j = X_1 + \cdots + X_j = S_j$, the image by the projection $G \to K$. Furthermore, $|h^s| = |shs^{-1}| \leqslant |h|$ for $h \in H$, $s \in \mathbb{R}$, $s < 0$ by the NC-condition that gives $\ell_j \geqslant 0$.

For the random walk $S_j \in \mathbb{R}$ we can use the gambler's ruin estimate (3.11) to deduce that $\dot{s}_1, \dot{s}_2, \ldots, \dot{s}_n < 0$ with probability at least cn^{-c}. Formula (2.21) therefore implies that $z(n) \in A_n$ with probability $\geqslant cn^{-c}$. Our criterion applies and this completes the proof of (3.2) for our special group.

A more general example is provided by (iii) in §2.3.2, $\mathbb{R}^d \leftthreetimes \mathbb{R}^m = W \leftthreetimes V$ with the action defined by (2.8). Here $L_1, \ldots, L_d \in V^*$ are assumed to satisfy the NC-condition and we shall assume for simplicity that $L_j \neq 0$, $j = 1, \ldots, d$. We shall use polar coordinates $x = r\sigma$, $r > 0$, with $\sigma \in \Sigma_{m-1}$ the unit sphere in V. Then the NC-condition is equivalent to the fact that there exists a conical domain

$$\Omega = [x = r\sigma \in V;\ r > 0,\ |\sigma - \sigma_0| < \varepsilon_0] \tag{3.12}$$

(for some small positive ε_0) such that $L_j x < 0$, $j = 1, \ldots, d$, $x \in \Omega$ (see §2.3.1). Our previous example is a special case of this with $V = \mathbb{R}$ and Ω the half-line. The proof of the NC-theorem for this more general group follows verbatim the proof of the previous case where $m = 1$. The only thing that we need to generalise in the proof is the gambler's ruin estimate, and this will be explained in the next subsection.

3.3.6 The gambler's ruin estimate in a conical domain

Let $\Omega \subset V \cong \mathbb{R}^m$ be as in (3.12) and let $X_1, \dots, X_j, \dots \in V$ be independent, identically distributed, bounded, centred random variables. Let $\psi(x)\,dx$ be the distribution of these variables. For simplicity we shall assume that $\psi(x) > c$ in some neighbourhood of 0. We shall also denote by $S_n = X_1 + \cdots + X_n$ the partial sums of the corresponding random walk. There exist then C, c such that

$$\mathbb{P}_0[S_1, \dots, S_n \in \Omega] \geqslant Cn^{-c}; \quad n \geqslant 1. \tag{3.13}$$

The proof of this estimate will be given in the appendix to this chapter. In this coarse form, as we shall see, the proof is not too difficult. Where the problem becomes both difficult and interesting is when the issue is to find the exact exponent for $n^{-c}, c = c(\Omega_{\mathscr{L}})$ in (3.13), where now $\Omega = \Omega_{\mathscr{L}} = [x \in V;\, L_j x > 0]$. This is one of the problems that we have to address in order to be able to prove the sharp local central limit theorem (3.3).

3.3.7 Plan of the proof

The examples in §3.3.4 are convincing and we clearly have a correct approach. All the more so since, by the structure theorems that we explained in §3.2.3(iii), this approach can easily be applied to all real soluble algebraic groups as in §3.2.3.

These types of structure theorems do not exist for general soluble Lie groups and it is here that the Cartan subgroups come to our rescue. In §§3.4.2–3.4.3 we explain how this notion is exploited. What we do in effect in these subsections is to reduce the problem to a special class of subgroups. It should be noted that in an earlier version of the proof we avoided the use of the Cartan subgroups (see Varopoulos, 1994b). This approach is more elementary but also more messy.

In §3.4.5 we give an elementary lemma on matrices that takes care of the fact that we cannot always diagonalise the action on the semidirect products as in §3.3.4, and as a result we must resort to root space decompositions. In effect this lemma says that when we have a sequence $T_1, T_2, \dots$ of bounded upper triangular matrices then the norm of the matrix $(I + T_1) \cdots (I + T_n)$ is $O(n^c)$, that is, it grows polynomially as $n \to \infty$.

In §3.5 we finally give the proof of the theorem. The ingredients are as follows. First we give the random walk representation of (2.21), which we recall again in the present notation in (3.23). Then we adapt the notion of the conical domain and the gambler's ruin estimate in §3.5.1. Finally, in §3.5.2 we conclude the proof. This is done by applying the criterion of §3.3.3 and for this purpose we use the lemma of §3.4.5.

3.4 The Structure Theorem and Cartan Subgroups

3.4.1 An example: algebraic groups

The NC-theorem for a general NA-group can be proved with only slight modifications in the proofs of the examples that we gave in the last section. However, it is preferable to tackle the general case straight away and start from one of the main difficulties of the problem.

To understand that difficulty we shall consider $\mathfrak{g}$, a Lie algebra over a field K (think of $K = \mathbb{R}$) of characteristic 0. We shall assume that $\mathfrak{g}$ is the Lie algebra of an algebraic group G over K as defined in §3.2. In that case we have at our disposal the perfect substitute of the example $\mathbb{R}^d \leftthreetimes \mathbb{R}^m$ that we examined in the previous section. We can assert that $\mathfrak{g} = \mathfrak{n} + \mathfrak{a} + \mathfrak{s}$ is the direct vector space sum of three subalgebras that satisfy the following conditions: $\mathfrak{n}$ is an ideal and is nilpotent; $\mathfrak{a}$ is Abelian; $\mathfrak{s}$ is semisimple and $[\mathfrak{a}, \mathfrak{s}] = 0$ (see Chevalley, 1955, V §4.2).

With the help of this structure theorem the proof that we gave in the last section easily adapts and we obtain a proof of the NC-theorem for real algebraic groups.

Nothing like this holds for real Lie groups in general and we must start by finding the substitute for this that we can use for all Lie groups.

3.4.2 Cartan subalgebras

Chevalley (1955) and Jacobson (1962) are good references for Cartan subalgebras. Let $\mathfrak{g}$ be some Lie algebra over a field of characteristic 0. We then say that $\mathfrak{h} \subset \mathfrak{g}$ is a Cartan subalgebra if $\mathfrak{h}$ is nilpotent and if it is its proper normaliser. This means that if $x \in \mathfrak{g}$ and $[x, \mathfrak{h}] \subset \mathfrak{h}$ then $x \in \mathfrak{h}$. This definition implies that if $\mathfrak{g}$ is nilpotent then $\mathfrak{h} = \mathfrak{g}$. More generally, it implies that $\mathfrak{h}$ is a maximal nilpotent subalgebra of $\mathfrak{g}$.

For our purpose the only two things that count are first, that every such Lie algebra $\mathfrak{g}$ admits at least one Cartan subalgebra. And second, that if $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$ and $\pi \colon \mathfrak{g} \to \mathfrak{g}_1$ is a surjective homomorphism, then $\pi(\mathfrak{h})$ is a Cartan subalgebra of $\mathfrak{g}_1$ (see Chevalley, 1955, VI, §4.5 Proposition 17). The conclusion from the above properties of Cartan subalgebras that we shall draw is that for every soluble algebra $\mathfrak{q}$ there exists some nilpotent algebra $\mathfrak{h} \subset \mathfrak{q}$ such that $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$, where $\mathfrak{n}$ is the nilradical. We see this by projecting on $\mathfrak{q}/\mathfrak{n}$ and then the image of a Cartan subalgebra of $\mathfrak{q}$ is $\mathfrak{q}/\mathfrak{n}$ because this algebra is Abelian.

Now let $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ be the Levi decomposition of the Lie algebra $\mathfrak{g}$ where $\mathfrak{q}$ is the radical and $\mathfrak{s}$ is some (semisimple) Levi subalgebra.

Lemma 3.3 (A lemma about Lie algebras) *Let* $\mathfrak{g} = \mathfrak{q} \rtimes \mathfrak{s}$ *be some Lie algebra together with its Levi decomposition as above and let* $\mathfrak{n}$ *be the nilradical. Then there exists* $\mathfrak{h} \subset \mathfrak{q}$*, some nilpotent subalgebra, such that* $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$ *and* $[\mathfrak{s}, \mathfrak{h}] = 0$.

We shall give the proof of the lemma at the end of this section but before that we shall draw the appropriate consequences.

First of all, since $\mathfrak{h}$ and $\mathfrak{s}$ act on $\mathfrak{n}$ we can form the semidirect product $\tilde{\mathfrak{g}} = \mathfrak{n} \rtimes (\mathfrak{h} \oplus \mathfrak{s})$ and the surjective homomorphism $\tilde{\mathfrak{g}} \to \mathfrak{g}$. See the note in §3.4.3 below. The radical of $\tilde{\mathfrak{g}}$ is $\tilde{\mathfrak{q}} = \mathfrak{n} \rtimes \mathfrak{h}$ and we can consider the roots of the action of $\mathfrak{h}$ on $\mathfrak{n}$ and the corresponding Zassenhaus decomposition of $\mathfrak{n}$. Here $\mathfrak{h}$ is not Abelian but it is nilpotent and this is good enough when $K = \mathbb{R}$ and in that case one starts by complexifying. One can find a general discussion of all this in Jacobson (1962, §II.4). These roots can be identified with the composite roots given in §2.3.4 because they vanish on $\mathfrak{n} \cap \mathfrak{h}$. The key conclusion is that $\tilde{\mathfrak{g}}$ is NC if and only if $\mathfrak{g}$ is. A more precise analysis of the action of $\mathfrak{h}$ on $\mathfrak{n}$ will be given in §3.8.

From here onwards we assume that $K = \mathbb{R}$. Now let G, N, H, S be the simply connected real Lie groups that correspond to $\mathfrak{g}$, $\mathfrak{n}$, $\mathfrak{h}$, $\mathfrak{s}$; we can then form the semidirect product and the homomorphism

$$\pi\colon\ \widetilde{G} = N \rtimes (H \oplus S) \to G. \tag{3.14}$$

Since $\widetilde{G}$ is an NC-amenable Lie group if and only if G is, we finally conclude by the reduction in §3.2.2 that it suffices to prove the NC-theorem for $\widetilde{G}$. Let us recapitulate.

3.4.3 A reduction

In the proof of the NC-theorem we may assume that the group G is of the form $G = N \rtimes (H \oplus S)$ where N, H are nilpotent, S is semisimple compact and furthermore all these groups are simply connected.

Note In Varadarajan (1974, §3.14), and Hochschild (1965, III.2), one can find more details on the semidirect products of Lie groups and Lie algebras. However, since these types of constructions will be used repeatedly, let us elaborate. If $\mathfrak{a}, \mathfrak{b} \subset \mathfrak{g}$ are subalgebras such that $\mathfrak{a}$ is an ideal then the algebra $\mathfrak{b}$ acts by derivations on $\mathfrak{a}$ and therefore we can define $\mathfrak{a} \rtimes \mathfrak{b}$. By the definition of the semidirect product, the canonical mapping $\mathfrak{a} \rtimes \mathfrak{b} \to \mathfrak{g}$ follows. If, in addition, $\mathfrak{b} = \mathfrak{b}_1 + \mathfrak{b}_2$ for two commuting subalgebras, we can also construct $\mathfrak{a} \rtimes (\mathfrak{b}_1 \oplus \mathfrak{b}_2)$ for the direct sum of the two subalgebras. The fact that $[\mathfrak{b}_1, \mathfrak{b}_2] = 0$ is, of course, essential here.

3.4.4 Proof of the Lie algebras lemma

Let $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ be the Levi decomposition of $\mathfrak{g}$ where $\mathfrak{q}$ is the radical. Since $\mathfrak{n}$ is an ideal, by the semisimplicity of $\mathfrak{s}$ it follows that there exists $\ell \subset \mathfrak{q}$, a subspace stable by the $\mathfrak{s}$ action, such that $\mathfrak{n} \oplus \ell = \mathfrak{q}$ because of H. Weyl's fundamental theorem on the representation of semisimple Lie algebras (see Varadarajan, 1974, §3.13). Since, in addition, $[\mathfrak{g}, \mathfrak{q}] \subset \mathfrak{n}$, it follows that $[\mathfrak{s}, \ell] = 0$. Let

$$\mathfrak{q}_0 = [x \in \mathfrak{q};\ [\mathfrak{s}, x] = 0];$$

this is clearly a soluble subalgebra and what we have seen is that $\ell \subset \mathfrak{q}_0$. This implies that in the mapping $\alpha\colon \mathfrak{q} \to \mathfrak{q}/\mathfrak{n}$ we have $\alpha(\mathfrak{q}_0) = \mathfrak{q}/\mathfrak{n}$. Let $\mathfrak{h}_0 \subset \mathfrak{q}_0$ be some Cartan subalgebra of $\mathfrak{q}_0$. Then since $\alpha(\mathfrak{h}_0)$ is a Cartan subalgebra of an Abelian algebra, it is the whole of $\mathfrak{q}/\mathfrak{n}$. It follows that $\mathfrak{n} + \mathfrak{h}_0 = \mathfrak{q}$. This completes the proof. For another proof of a similar lemma see Alexopoulos (1992).

Remark In the geometric theory in Part II we shall need a refinement in the construction of $\mathfrak{h}$. Let $\mathfrak{m} \subset \mathfrak{h}$ be a subspace such that $\mathfrak{n} + \mathfrak{m} = \mathfrak{q}$, $\mathfrak{n} \cap \mathfrak{m} = \{0\}$ and define the subalgebra $\mathfrak{h}_1 = [\mathfrak{h}, \mathfrak{h}] + \mathfrak{m}$. The fact that $[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{n} \cap \mathfrak{h}$ (see Varadarajan, 1974, §3.8.3) implies that for this subalgebra we have $\mathfrak{h}_1 = \mathfrak{h}$ if and only if $[\mathfrak{h}, \mathfrak{h}] = \mathfrak{n} \cap \mathfrak{h}$. Furthermore, $\mathfrak{h}_1$ clearly satisfies the conditions of the lemma and in that lemma we can replace $\mathfrak{h}$ by $\mathfrak{h}_1$. Therefore, when $\mathfrak{h}_1 \subsetneq \mathfrak{h}$ we can repeat the same construction with this new $\mathfrak{h}_1$. By repeating this operation if necessary, we can thus guarantee that $\mathfrak{h}$, the subalgebra of the lemma, satisfies $[\mathfrak{h}, \mathfrak{h}] = \mathfrak{n} \cap \mathfrak{h}$.

3.4.5 A lemma in linear algebra

We shall use the notation of §3.4.2 and we recall that $\mathfrak{n}$ is the nilradical of $\mathfrak{g}$ and $\mathfrak{h} \subset \mathfrak{q}$ is some nilpotent subalgebra such that $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$. As in §3.4.3 we denote by H the simply connected Lie group that corresponds to $\mathfrak{h}$. The $\mathrm{Ad}\,x\colon \mathfrak{n} \to \mathfrak{n}$ action is as usual defined for $x \in H$ by $\mathrm{Ad}(\exp \xi) = \exp(\mathrm{ad}\, \xi)$, $\xi \in \mathfrak{h}$, for the exponential mapping $\exp\colon \mathfrak{h} \to H$ which here is a bijection (see Varadarajan, 1974, §§2.10, 3.6.2). We can complexify $\mathfrak{n}^c = \mathfrak{n} \otimes \mathbb{C}$ and choose an appropriate basis on $\mathfrak{n}^c$ so as to have a Zassenhaus decomposition of the action of $\mathrm{ad}(\mathfrak{h})$, and $\mathrm{Ad}(H)$ (see Jacobson, 1962, §II.4). This action decomposes into diagonal blocks,

$$\begin{aligned} \mathrm{Ad}\,x = \mathrm{diag}(M_1, M_2, \ldots, M_r), \quad M_j = (\exp \lambda) I + T; \\ \lambda = \lambda_j = \lambda_j(\xi),\ x = \exp \xi,\ \xi \in \mathfrak{h}; \end{aligned} \tag{3.15}$$

here I is the identity matrix and $\lambda_1, \ldots$ are the roots of that action and, as already pointed out, they can be identified with the composite roots of §2.3.4. As for $T = T_j = (t_{pq})$, it is a strictly upper-triangular matrix: that is, $t_{pq} = 0$ for $p \geqslant q$.

To prove the NC-theorem we shall use (3.15) to estimate $\mathrm{Ad}(x_1 x_2 \cdots x_n)$ for a sequence $x_1, x_2, \ldots \in H$ that satisfies

$$|x_j|_H,\ |x_j^{-1}|_H \leqslant C_0; \quad 1 \leqslant j \leqslant n, \tag{3.16}$$

that is, the x_j lie in some fixed symmetric ball of H. We then have

$$x_j = \exp(\xi_j); \quad \xi_j \in \mathfrak{h},\ |\xi_j| \leqslant C, \tag{3.17}$$

for some fixed Euclidean norm on $\mathfrak{h}$. Each diagonal block of $\mathrm{Ad}(x_1 \cdots x_n)$ then becomes a product of matrices $M^{(1)}, M^{(2)}, \ldots$ of the form $M^{(\alpha)} = \exp \lambda^{(\alpha)} I + T_\alpha$,

$$\left.\begin{aligned} &M^{(1)} \cdots M^{(n)} = \exp \lambda^{(1)} \cdots \exp \lambda^{(n)} \prod_{\alpha=1}^{n} (I + \widetilde{T}^{(\alpha)}), \\ &\lambda^{(\alpha)} = \lambda(\xi_\alpha),\ x_\alpha = \exp \xi_\alpha \in H,\ \xi_\alpha \in \mathfrak{h},\ |\xi_\alpha| \leqslant C, \\ &\widetilde{T}^{(\alpha)} = (\exp \lambda^{(\alpha)})^{-1} T_\alpha; \quad \alpha = 1, \ldots, n, \end{aligned}\right\} \tag{3.18}$$

where here $\lambda(\xi)$ is the root that corresponds to the diagonal block and the index $1 \leqslant j \leqslant r$ of the block has been suppressed to simplify notation. The conclusion is that the matrix norm $\| \ \|_{\mathfrak{n}}$ on $\mathfrak{n}^c$ satisfies

$$\|M^{(1)} \cdots M^{(n)}\| \leqslant C(\exp \lambda^{(1)} \cdots \exp \lambda^{(n)})(1+n)^D, \tag{3.19}$$

where C depends only on C_0 in (3.16) and where $D = \dim \mathfrak{n} + 10$. This is seen by expanding the product in (3.18) and using the triangular nature of the matrices.

If we use the real part $L = \mathrm{Re}\,\lambda$ of $\lambda(\xi) = L\xi + i\,\mathrm{Im}\,\lambda(\xi)$, we finally see that (3.19) can be estimated by

$$(1+n)^D \exp(L\xi_1 + L\xi_2 + \cdots + L\xi_n). \tag{3.20}$$

We shall combine all the blocks of (3.15). Furthermore, we shall fall back on our notation of §§2.1.1 and 2.3.4 and denote by $\mathscr{L} = (L_1, \ldots, L_p)$ the set of the real roots of $\mathfrak{q}$. We recapitulate the above in the following.

Lemma 3.4 *Let the notation be as above and $x_j = \exp \xi_j \in H$, $\xi_j \in \mathfrak{h}$ as in* (3.17). *Then there exists $C > 0$ such that*

$$\|\mathrm{Ad}(x_1 \cdots x_n)\|_{\mathfrak{n}} \leqslant C\Big(\sup_{1 \leqslant i \leqslant p} \exp(L_i\xi_1 + \cdots + L_i\xi_n)\Big)(1+n)^D; \quad n \geqslant 1. \tag{3.21}$$

As we shall see in the next section, the importance of this estimate is that it splits into the polynomial factor $(1+n)^D$ and an additional exponential factor that we shall be able to control using the geometry of the roots.

3.5 Proof of the NC-Theorem

Here we assume that G is simply connected and use the notation $G = N \leftthreetimes (H \oplus S)$ of (3.14) and also the notation from (2.21) in §2.7. We then have $x_j = m_j\dot{x}_j$, $m_j \in N$, $\dot{x}_j \in H \oplus S$. The 'dots', as in §2.7.1, just denote the images by $G \to H \oplus S$ and $\dot{x}_j = h_j\sigma_j$, $h_j \in H$, $\sigma_j \in S$. In the next few lines all the obvious identifications will be made. We shall use the formula (see Varadarajan, 1974, §2.13).

$$\exp(\mathrm{Ad}(x)\zeta) = (\exp\zeta)^x = x(\exp\zeta)x^{-1}; \quad x \in G,\ \zeta \in \mathfrak{g}. \tag{3.22}$$

With this notation, the random walk formula (2.21) can now be rewritten

$$s_n = x_1 \cdots x_n = m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \dot{s}_n. \tag{3.23}$$

Here we shall set $x_j \in G$ to be independent identically distributed random variables with distribution $\mu \in \mathbb{P}(G)$, the probability measure of the NC-theorem from §3.2.1, so that $s_n = z(n)$ is the Markov chain construction in §3.3.2. To complete the proof of the theorem we shall construct the sequence of subsets $A_n \subset G$ of the criterion §3.3.3 and we shall show that conditions (i) and (ii) are satisfied. This will be done in the next few lines.

As in the examples in §3.3.4, $d^r g = dm\,dh\,d\sigma$ for $m \in N$, $h \in H$, $\sigma \in S$ and the sets are defined by

$$A_n = [|m| \leqslant Cn^c] \times [|h| \leqslant Cn^c] \times S; \quad n \geqslant 1, \tag{3.24}$$

where the constants C, c are chosen appropriately large and $|\ |$ denotes the distance from the neutral element in the group. We recall here that S is compact and therefore condition (i) of the criterion is satisfied. In the next two subsections we shall show that condition (ii) is also satisfied.

3.5.1 The vector space, the roots and the conical domain

We shall write $V = H/[H,H]$; this is a Euclidean space. Here we use the exponential mapping to identify $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}] \simeq V$. Furthermore, in §3.4 the roots of the action of $\mathfrak{h}$ on the $\mathfrak{n}$ λ_j clearly vanish on $[\mathfrak{h},\mathfrak{h}]$ (since this bracket lies in

$\mathfrak{n}$: see Varadarajan, 1974, §3.8.3) and therefore their real parts $L_j \in V^*$ can be identified with linear functionals on V. We denote

$$\Omega = \Omega_{\mathscr{L}} = [x \in V;\ L_j x < 1,\ L_j \in \mathscr{L}], \tag{3.25}$$

which by the NC-condition is a non-empty open conical domain of V. We have $0 \in \Omega$ and when $\mathscr{L} = (0)$, we have $\Omega = V$. Here the L_j are the composite real roots that we defined in §2.3.4. This action of $\mathfrak{h}$ will be examined in more detail in §3.8.2 later on.

Let $\alpha\colon K = H \times S \to H \to V$ be the composition of the canonical projections. First, the map $s_n \to \dot{s}_n$ projects from G to K; and then $\alpha\colon \dot{s}_n \to \check{s}_n \in V$ projects on V. Clearly $\check{s}_n$ is a random walk as in §3.3.6 and therefore, by (3.13),

$$\mathbb{P}_e[\check{s}_j \in \Omega;\ j = 1, 2, \ldots, n] \geqslant c n^{-c}; \quad n \geqslant 1. \tag{3.26}$$

3.5.2 Condition §3.3.3(ii) and the conclusion of the proof

First we recapitulate our earlier notation, and return to (3.23):

$$\begin{aligned} \dot{x}_j &= h_j \sigma_j,\ h_j \in H,\ \sigma_j \in S,\ x_j = m_j \dot{x}_j,\ m_j \in N, \\ z(n) = s_n &= m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \dot{s}_n, \end{aligned} \tag{3.27}$$

where $z(n)$ is the Markov chain of §3.3.2. Since $|\dot{s}_j|_{H \times S} < Cj$, for condition (ii) of the criterion to hold it suffices to verify that there exist constants such that

$$\mathbb{P}_e\left[|m_j^{\dot{s}_{j-1}}|_N \leqslant C(1+j)^C;\ j = 1, 2, \ldots, n\right] \geqslant C n^{-c}. \tag{3.28}$$

But since $m_j^{\dot{s}_{j-1}} = \sigma_1 \cdots \sigma_{j-1} m_j^{h_1 \cdots h_{j-1}} (\sigma_1 \cdots \sigma_{j-1})^{-1}$, and since S is compact, it suffices to prove (3.28) with $m_j^{h_1 \cdots h_{j-1}}$ instead.

Towards that we use (3.22) to express $m_j^h = \exp(\mathrm{Ad}(h)\zeta_j)$ with $\exp(\zeta_j) = m_j$ and $\zeta_j \in \mathfrak{n}$, $h \in H$. If we combine Lemma 3.4 with (3.26) we have

$$\mathbb{P}_e\left[|\,\mathrm{Ad}(h_1 \cdots h_{j-1})\zeta_j| \leqslant C(1+j)^C;\ j = 1, \ldots, n\right] \geqslant C n^{-c} \tag{3.29}$$

for some fixed Euclidean norm as in (3.17) on $\mathfrak{n}$ and appropriate constants. But for that norm we have polynomial distortion (see §2.14) and there exist constants such that

$$|\exp(\zeta)|_N \leqslant C(1 + |\zeta|)^C; \quad \zeta \in \mathfrak{n}. \tag{3.30}$$

Condition (ii) of the criterion follows from (3.26)–(3.30) and this completes the proof of the theorem.

Part 3.2: Heat Diffusion Kernel

3.6 Statement of the Results and the Tools

As in Part 2.2, the reader could skip this section in a first reading. The setting and the notation are those of Part 2.2 and ϕ_t denotes the convolution kernel of the semigroup $T_t = e^{-t\triangle}$ with $\triangle$ as in §2.12. The connected Lie group G will again be assumed to be an amenable NC-group. We then have the following theorem.

Theorem 3.5 (NC-theorem for the heat diffusion kernel) *Let G and ϕ_t be as above*; *then there exist $C, c > 0$ such that*

$$\phi_t(e) \geqslant Ct^{-c}; \quad t \geqslant 1. \tag{3.31}$$

The organisation and structure of the proof of this estimate are identical to what we did for compactly supported measures in the first part of this chapter. The modifications needed will be given in the next subsection. Here we shall explain the slight changes that have to be made in the tools used in the proof.

3.6.1 The lifting of the semigroup

Let $\pi\colon \widetilde{G} \to \widetilde{G}/H = G$ be a projection, where H is a closed normal subgroup of the connected Lie group $\widetilde{G}$. Let $\triangle = -\sum X_j^2$ be some sub-Laplacian on G as in §2.12. Also let $\widetilde{Y}_1, \dots, \widetilde{Y}_k$ be a basis for the space of left-invariant vector fields on the connected component of H. We shall consider $\widetilde{\triangle} = -\sum \widetilde{Y}_j^2 - \sum_j \widetilde{X}_j^2$ where $\widetilde{X}_j$ is some choice of left-invariant field on $\widetilde{G}$ such that $d\pi(\widetilde{X}_j) = X_j$. It is then clear that $\widetilde{\triangle}$ is a subelliptic Laplacian on $\widetilde{G}$ (see Varopoulos et al., 1992).

Exercise 3.6 Prove this using the Hörmander condition. If the Laplacian $\triangle$ is elliptic there is essentially nothing to prove.

Let $d\tilde{\mu}_t = \widetilde{\phi}_t\, d^r\tilde{g}$ be the corresponding heat diffusion kernel of $e^{-t\tilde{\triangle}}$ as in §2.12. It is then clear that $\check{\pi}(\tilde{\mu}_t) = \mu_t$.

Exercise 3.7 Prove this formal fact using the definitions to see that $(f * \mu_t) \circ \pi = (f \circ \pi) * \tilde{\mu}_t$ for all $f \in C_0^\infty(G)$.

3.6.2 The gambler's ruin estimate

The notation is as in §3.3, with Ω a conical domain as in (3.12) or (3.25). Here we shall assume that the independent identically distributed variables $X_1, \dots \in V \cong \mathbb{R}^d$ of §3.3.6 are non-degenerate normal variables.

The same gambler's ruin estimate (3.13) then holds for the corresponding partial sums $S_j = X_1 + \cdots + X_j$. More explicitly, let us define the event

$$\mathscr{E}_n = \left[S_j \in \Omega;\ j = 1,2,\ldots,n\right]; \quad n \geqslant 1. \tag{3.32}$$

Then there exist $C_1, c_1 > 0$ such that

$$\mathbb{P}(\mathscr{E}_n) \geqslant C_1 n^{-c_1}; \quad n = 1,2,\ldots. \tag{3.33}$$

The proof will be given in the appendix to this chapter. We should also note that there is nothing special about normal variables here. It suffices to consider centred variables that have non-singular covariance and a sufficiently high moment $\mathbb{E}(|X_j|^A) < +\infty$. In our applications these variables are normal because they are obtained by the projected heat diffusion kernel on $\mathbb{R}^d$. What is needed is the very coarse estimate (3.33). It should also be noted that getting sharp results on the exponent c_1 is a difficult problem that involves quite a lot of non-trivial potential theory; see the note at the end of §3A.6.

3.7 Proof of the NC-Theorem for the Heat Diffusion Kernel

As already said, the proof follows very closely the proof of the compact support case in §3.3–§3.5.

We proceed as before and with the lifting of the heat diffusion kernel of §3.6.1 we can reduce the problem to the simply connected case. And then with the help of the Cartan subalgebras we can reduce the proof to the group $\widetilde{G} = N \leftthreetimes (H \oplus S)$ of (3.14). For that reduction, in §3.2.2 we used the Harnack estimate of §2.5. Here we use the Harnack estimate §2.12.1 instead. As we did in §3.4.3 we suppress the tilde (˜) and denote this group by $G = N \leftthreetimes (H \oplus S)$.

Lemma 3.4 will be used again. We shall consider the Markov chain $z(n) \in G$ defined by $\mathbb{E}_x(f(z(n)) = T_n f(x)$, $n = 1,2,\ldots$ for the semigroup $T_t = e^{-t\triangle}$ on G and we shall show that it satisfies the conditions of the criterion in §3.3.3 for the sets of (3.24), $A_n = [|m| \leqslant Cn^c] \times [|h| \leqslant Cn^c] \times S \subset G$, for an appropriate choice of the constants. Here the notation is as before and $x = mh\sigma \in G, m \in N$, $h \in H$, $\sigma \in S$.

The criterion in §3.3.3 applies here because at no point in the proof did we use the compactness of the support. It was only used for the Harnack principle of §2.5 and here we have Harnack anyway from (2.54). Inequality (3.9) follows and so does (3.31) for $t = 1,2,\ldots$. We then apply the Harnack estimate of §2.12.1 and (3.31) follows for all $t > 0$.

The verification of condition (i) of the criterion is automatic as before, and

$|A_n| \leqslant Cn^c$. The issue is therefore once again to verify condition (ii) of the criterion in §3.3.3 and for this we must control the 'Gaussian tail' of §2.12.2.

3.7.1 Condition (ii) of the criterion

For this we preserve all the notation of §3.5 and in particular $V = H/[H,H]$, the roots L_j and $\Omega_{\mathscr{L}} \subset V$ as in (3.25). The specific notation of §3.5.2 is preserved and we write in particular $x_j = m_j \dot{x}_j$, $z(n) = s_n = x_1 \cdots x_n \in G$, etc. for the independent identically distributed random variables $x_j \in G$ with the density $\mu_1 = \phi_1\, d^r g \in \mathbb{P}(G)$ of the heat diffusion kernel. By formula (2.21), as in (3.27) we have $z(n) = m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \dot{s}_n$. As in §3.5.1 we use $\alpha\colon K = H \times S \to V$ to define $\check{s}_n = \alpha(\dot{s}_n) \in V$ and the event $\mathscr{E}_n = [\check{s}_j \in \Omega_{\mathscr{L}};\ j = 1,2,\ldots,n]$, as in (3.26). It therefore follows as before that condition (ii) of the criterion will be satisfied as long as we can prove the analogue of (3.28). What must be proved is that for an appropriate choice of the constants we have

$$\mathbb{P}_e\left[|m_j^{\dot{s}_{j-1}}|_N \leqslant C(1+n)^c,\ |\dot{s}_j| \leqslant Cn^c;\ j = 1,\ldots,n\right] \geqslant Cn^{-c}; \quad n = 1,2,\ldots. \tag{3.34}$$

The difference from the case of bounded supports is that in that case, the condition $|\dot{s}_j| \leqslant Cn^c$ was automatic. The proof of (3.34) is a slight modification of the proof of (3.28). The modifications needed are technical, and for the convenience of the reader it is preferable to rewrite part of the proof rather than switch back and forth to §3.5.2. The notation will, however, be exactly as in §3.5.2 and it will not be recalled here.

As in §3.5.2 we shall write $m_j^{\dot{s}_{j-1}} = \sigma_1 \cdots \sigma_{j-1} m_j^{h_1 \cdots h_{j-1}} (\sigma_1 \cdots \sigma_{j-1})^{-1}$ and, as before, conjugation by elements of S can be ignored because S is compact. We are therefore left with having to prove (3.34) with $m_j^{h_1 \cdots h_{j-1}}$ instead.

Using again the end of the argument of §3.5.2 we see that for this it suffices to construct an event $\mathscr{E}'_n \subset \mathscr{E}_n$, $n \geqslant 1$ such that

$$\mathbb{P}_e(\mathscr{E}'_n) \geqslant Cn^{-c}; \quad n \geqslant 1, \tag{3.35}$$

and such that for all n and all $(x_1, x_2, \ldots) \in \mathscr{E}'_n$ we have

$$|\operatorname{Ad}(h_1 \cdots h_{j-1})\zeta_j|_{\mathfrak{n}} \leqslant C(1+n)^c, \quad |h_j|_H \leqslant Cn^c; \quad 1 \leqslant j \leqslant n, \tag{3.36}$$

where, as before, $\zeta_j \in \mathfrak{n}$ is defined by $\exp(\zeta_j) = m_j$.

Here, unlike (3.28), it is not true that $|m_j|_N$, $|\zeta_j|_{\mathfrak{n}} \leqslant C$ and this forces us to modify the definition of $\mathscr{E}_n$ in (3.26) and consider a smaller subset $\mathscr{E}'_n \subset \mathscr{E}_n$. To define $\mathscr{E}'_n$ we consider the event

$$\mathscr{B}_n = [\sup_{1 \leqslant j \leqslant n} |x_j|_G \leqslant \log(n+10)]; \quad n \geqslant 1,$$

and by the Gaussian decay of the variables x_j of the random walk in G we see that the complementary event satisfies

$$\mathbb{P}_e(\sim \mathscr{B}_n) = O(\exp(-c\log^2(n+1))); \quad n \geqslant 1, \tag{3.37}$$

for some $c > 0$. With $x_j = m_j\dot{x}_j$ as in §3.5, on the event $\mathscr{B}_n$ we have

$$|m_j|_N \leqslant C\exp(c|x_j|_G) \leqslant Cn^c, \quad |h_j|_N \leqslant Cn^c; \quad j = 1,2,\dots,n,\ n \geqslant 1, \tag{3.38}$$

for appropriate constants. The first inequality is a consequence of the distance distortion inequality (2.59) that N satisfies in G because N is contained in the nilradical of G (see §2.14.2). The second estimate in (3.38) is evident by the projection $G \to H$. Now, by the polynomial distortion of the exponential mapping of (2.60), we have

$$|\zeta_j|_{\mathfrak{n}} \leqslant C(1 + |m_j|_N)^C; \quad j = 1,2,\dots \tag{3.39}$$

for appropriate constants, where we recall that $\zeta_j \in \mathfrak{n}$ is such that $\exp(\zeta_j) = m_j$. From (3.38) it therefore follows that on $\mathscr{B}_n$ we also have

$$|\zeta_j|_{\mathfrak{n}} \leqslant Cn^c; \quad j = 1,2,\dots,n. \tag{3.40}$$

The new events are now defined by

$$\mathscr{E}'_n = \mathscr{E}_n \cap \mathscr{B}_n; \quad n \geqslant 1, \tag{3.41}$$

and (3.35) is a consequence of the gambler's ruin estimate (3.33) and (3.37). For the application of (3.33) we observe that $\check{s}_n$ is a Gaussian random walk since it is obtained by the process (a Brownian motion) of V generated by $d\alpha \circ d\pi(\Delta)$, where we use the notation of (3.14) and (3.26). This of course is also a consequence of the more general (2.55).

To finish the proof of (3.34) we use (3.37) and Lemma 3.4 and argue as at the end of §3.5.2 and the polynomial distortion of the exponential mapping of (2.60). The proof of condition (ii) of the criterion and of the theorem is finally complete. Notice that here (3.16) fails and therefore the estimate in Lemma 3.4 needs a slight and obvious modification. This problem is taken care of in the generalisation of the lemma given in §3.9.1.

Part 3.3: An Alternative Approach

3.8 Algebraic Considerations

The rest of this chapter is devoted to an alternative approach to proving the NC-theorem. Let G be some connected real Lie group and let us consider $\mathscr{R}$

the class of the closed normal subgroups H such that G/H is an R-group as in §2.2.2. If $H_1, H_2 \in \mathscr{R}$ then the Lie algebra of $G/H_1 \cap H_2$ can be identified to a subalgebra of the Lie algebra of $(G/H_1) \times (G/H_2)$ and therefore $H_1 \cap H_2 \in \mathscr{R}$. Here we use the obvious facts that subgroups, quotients and products of R-groups are R (see Guivarc'h, 1973). Furthermore, if $H \in \mathscr{R}$ then H_0, the connected component of H, also belongs to $\mathscr{R}$. All this is clear because only the Lie algebra is involved in the definition of the R-condition of §2.2.2. It follows that $\mathscr{R}$ has a minimal element which is connected and will be denoted by N_R. If G is amenable and N is its nilradical then $G/N \cong \mathbb{R}^d \times S$ locally; it follows that in that case $N_R \subset N$.

3.8.1 A structure theorem

Theorem 3.8 *Let G be some simply connected NC-amenable Lie group and let N_R be as above. Then there exists $G_R \subset G$, a closed simply connected R-group, such that $G = N_R \leftthreetimes G_R$.*

We shall give a proof of this fact in this section. The alternative proof of the NC-theorem based on the above structure theorem will be given in the next section. It is clear that in a first reading the reader could skip all of this. The idea of this alternative proof is, however, natural: the key to the proof we gave in §3.5 is the decomposition $\widetilde{G} = N \leftthreetimes (H \oplus S)$ of the overgroup $\widetilde{G} \to G$. Here we see that we have a similar decomposition for the group itself. The price that has to be paid is that G_R is not necessarily of the form (nilpotent) $\oplus$ (compact), hence the slight modifications in the arguments. This new approach is essential for the proof of the sharp local central limit theorem (3.3) and therein lies its importance.

For our considerations, the intrinsic characterisation of N_R that we gave is inessential. The part of the theorem that will be needed is that there exists $N_R \subset N$ and $G_R \subset G$ such that $G = N_R \leftthreetimes G_R$. We shall denote by $\mathfrak{g}$ the Lie algebra of G and construct $\mathfrak{n}_R \lhd \mathfrak{g}$ a nilpotent ideal and $\mathfrak{g}_R \subset \mathfrak{g}$ and R-subalgebra such that

$$\mathfrak{g} = \mathfrak{n}_R \leftthreetimes \mathfrak{g}_R. \tag{3.42}$$

This implies Theorem 3.8 because the groups involved are simply connected and therefore the local isomorphisms give rise to global ones. We shall denote by $\mathfrak{q} \lhd \mathfrak{g}$ the radical and first give the proof of (3.42) when $\mathfrak{q} = \mathfrak{g}$, that is, in the special case when the Lie algebra is soluble.

3.8.2 The Lie algebra is soluble

Here we shall elaborate further on the considerations of §§2.1–2.3 and examine analogous relevant root space decompositions of the adjoint action $\mathrm{ad}x\colon y \to [x,y]$ on the Lie algebra.

Let $\mathfrak{n} \lhd \mathfrak{q}$ be nilradical and let $\mathfrak{h} \subset \mathfrak{q}$ be some nilpotent subalgebra such that $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$ as in §3.4.2, where $\mathfrak{q}$ is a soluble algebra. We start by complexifying $\mathfrak{n} \otimes_{\mathbb{R}} \mathbb{C}$ and consider the Zassenhaus decomposition (see Jacobson, 1962, §II.4) into root spaces by the ad-action of $\mathfrak{h}$:

$$\mathfrak{n} \otimes \mathbb{C} = \hat{\mathfrak{n}}_0 \oplus \cdots \oplus \hat{\mathfrak{n}}_p. \tag{3.43}$$

The space $\hat{\mathfrak{n}}_0$ is special and could be zero, but when not zero it corresponds to the root $\mu_0 = 0$ if such a root exists. By taking complex conjugation exactly as in §2.1.1 we see that $\hat{\mathfrak{n}}_0$ is a real subspace, that is, $\hat{\mathfrak{n}}_0 = \tilde{\mathfrak{n}}_0 \otimes \mathbb{C}$ for some subspace $\tilde{\mathfrak{n}}_0 \subset \mathfrak{n}$. By the nilpotency of $\mathfrak{h}$ we also see that $\mathfrak{h} \cap \mathfrak{n} \subset \tilde{\mathfrak{n}}_0$. The subspaces $\hat{\mathfrak{n}}_j$, $1 \leqslant j \leqslant p$ in (3.43) are by definition $\neq \{0\}$ and they are the subspaces that correspond to the distinct non-zero roots $\mu_1, \ldots, \mu_p \in \mathrm{Hom}_{\mathbb{R}}[\mathfrak{h};\mathbb{C}]$.

The roots $\mu_0, \ldots, \mu_p$ will be decomposed into the equivalence classes of the equivalence relation

$$\mu_i \sim \mu_j \Leftrightarrow \mathrm{Re}\,\mu_i = \mathrm{Re}\,\mu_j; \quad i,j = 0, \ldots, p. \tag{3.44}$$

We shall block together the subspaces $\hat{\mathfrak{n}}_{i_1}, \ldots$ that correspond to the same equivalence class $L_i = \mathrm{Re}\,\mu_{i_1} = \cdots = \mathrm{Re}\,\mu_{i_\alpha}$. By the identifications that we made in §2.3.4 it follows that when $\mathfrak{q}$ is an NC-algebra these functionals satisfy the NC-condition. We can use complex conjugation again and we see as before that we obtain a real subspace

$$\hat{\mathfrak{n}}_{i_1} \oplus \cdots \oplus \hat{\mathfrak{n}}_{i_\alpha} = \mathfrak{n}_i \otimes \mathbb{C} \tag{3.45}$$

for some subspace $\mathfrak{n}_i \subset \mathfrak{n}$, with $0 \leqslant i \leqslant r$. Here $\mathfrak{n}_0$ corresponds to the real root $L_0 = 0$ if such a root exists; otherwise it is 0. We shall also define $\mathfrak{q}_R = \mathfrak{n}_0 + \mathfrak{h}$, so that $\mathfrak{n} \cap \mathfrak{q}_R = \mathfrak{n}_0 + (\mathfrak{h} \cap \mathfrak{n})$.

We finally obtain the decomposition

$$\mathfrak{n} = \mathfrak{n}_0 \oplus \cdots \oplus \mathfrak{n}_r, \ \tilde{\mathfrak{n}}_0 \subset \mathfrak{n}_0, \quad \mathfrak{n}_0 = \mathfrak{n} \cap \mathfrak{q}_R. \tag{3.46}$$

For this decomposition we have

$$[\mathfrak{n}_i, \mathfrak{h}] \subset \mathfrak{n}_i; \quad i = 0, 2, \ldots, r. \tag{3.47}$$

On the other hand, by easy and standard Lie algebra considerations (see Jacobson, 1962, §III.2, Exercise II.8) it follows that for any $i, j = 0, \ldots, r$ when

$[\mathfrak{n}_i, \mathfrak{n}_j] \neq 0$ then $[\mathfrak{n}_i, \mathfrak{n}_j] \subset \mathfrak{n}_k$ for some $0 \leqslant k \leqslant r$ and in that case we have for the corresponding real roots $L_i + L_j = L_k$. This means that if $\mathfrak{q}$ is NC then

$$\mathfrak{n}_R = \mathfrak{n}_1 \oplus \cdots \oplus \mathfrak{n}_r \tag{3.48}$$

is a subalgebra and by (3.47) it is in fact an ideal of $\mathfrak{q}$. This holds by the NC-condition because this implies that for $i, j = 1, \ldots, n$, $[\mathfrak{n}_i, \mathfrak{n}_j] \subset \mathfrak{n}_k$ for some $k \neq 0$. Furthermore, $\mathfrak{q}_R = \mathfrak{n}_0 + \mathfrak{h}$ is a subalgebra by (3.47). Note also (see Varadarajan, 1974, §3.8.3) that

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h} \cap \mathfrak{n} \subset \tilde{\mathfrak{n}}_0 \subset \mathfrak{n}_0. \tag{3.49}$$

By the definition of $\mathfrak{n}_0$, it follows that $\mathfrak{q}_R$ is an R-algebra. The required decomposition finally follows:

$$\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R. \tag{3.50}$$

Some of the above spaces could degenerate to zero. This is the case when $\mathfrak{q}$ is nilpotent because then we can take $\mathfrak{h} = 0$ and $\mathfrak{n}_R = 0$.

Exercise 3.9 Verify that $\mathfrak{n}_R$ is the smallest ideal such that $\mathfrak{q}/\mathfrak{n}_R$ is an R-algebra. (This fact will not be used in what follows.)

The algebra $\mathfrak{q}_R$ is soluble: its action on the space $\mathfrak{n}_i \otimes \mathbb{C}$ can therefore be triangulated by an appropriate basis. It therefore takes the form (see Varadarajan, 1974, §3.7.3)

$$\operatorname{ad}\zeta = \begin{pmatrix} \rho_1(\zeta) & & * \\ & \ddots & \\ 0 & & \rho_{t_i}(\zeta) \end{pmatrix}; \quad \zeta \in \mathfrak{q}_R,\ i = 0, \ldots, r, \tag{3.51}$$

where $\rho_j \in \mathrm{Hom}_{\mathbb{R}}[\mathfrak{q}_R;\ \mathbb{C}]$ and by the nilpotency of $\mathfrak{n}$, the ρ vanish on $\mathfrak{n}_0$ and $\mathfrak{h} \cap \mathfrak{n}$, and thus can be identified with linear functionals

$$\rho_j \in \mathrm{Hom}_{\mathbb{R}}[\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n},\ \mathbb{C}]; \quad \mathfrak{h}/\mathfrak{h} \cap \mathfrak{n} = \mathfrak{q}/\mathfrak{n}. \tag{3.52}$$

Furthermore, the roots $\rho_1, \ldots, \rho_{t_i}$ can be identified exactly with the equivalence class (3.44) and for $j = 1, \ldots, t_i$ we have

$$\operatorname{Re}\rho_j = L_i, \text{ the 'real root' of the space } \mathfrak{n}_i; \quad 0 \leqslant i \leqslant r. \tag{3.53}$$

3.8.3 General NC-algebras

Here $\mathfrak{g}$ is a general NC-algebra; we do not have to assume that $\mathfrak{g}$ is amenable but this will be the case in the applications. The radical $\mathfrak{q} \lhd \mathfrak{g}$ is NC-soluble and we have the Levi decomposition $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ for a semisimple algebra $\mathfrak{s}$. We shall choose $\mathfrak{h} \subset \mathfrak{q}$ to be some nilpotent algebra as above so that $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$ and also $[\mathfrak{h}, \mathfrak{s}] = \{0\}$ as in Lemma 3.3. We shall consider $\mathfrak{q}_R \subset \mathfrak{q}$ and $\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R$ as constructed in the previous subsection with the above nilpotent algebra $\mathfrak{h}$. The commutation $[\mathfrak{h}, \mathfrak{s}] = 0$ implies that in the decomposition (3.43) we have $[\hat{\mathfrak{n}}_i, \mathfrak{s}] \subset \hat{\mathfrak{n}}_i$ and therefore also that, with our previous notation,

$$[\mathfrak{n}_i, \mathfrak{s}] \subset \mathfrak{n}_i, \quad [\mathfrak{n}_R, \mathfrak{s}] \subset \mathfrak{n}_R, \quad [\mathfrak{q}_R,\ \mathfrak{s}] \subset \mathfrak{q}_R; \quad i = 0, 1, \ldots, r. \tag{3.54}$$

The bottom line is that (see (3.46))

$$\mathfrak{n} = \mathfrak{n}_R \leftthreetimes \mathfrak{n}_0, \quad \mathfrak{g} = \mathfrak{n}_R \leftthreetimes \mathfrak{g}_R, \quad \mathfrak{g}_R = \mathfrak{q}_R \leftthreetimes \mathfrak{s}, \quad \mathfrak{q}/\mathfrak{n} = \mathfrak{q}_R/\mathfrak{n}_0, \tag{3.55}$$

and here, when $\mathfrak{g}$ is amenable, $\mathfrak{g}_R$ is an R-algebra. In general, $\mathfrak{g}_R$ is not soluble and so the analogue of (3.51) does not generally hold for $\mathfrak{g}_R$. But for the spaces (3.52) on which the ρ's of (3.51) are defined, we have the following projection:

$$\pi\colon \mathfrak{g}_R \to (\mathfrak{q}_R \leftthreetimes \mathfrak{s})/\mathfrak{n}_0 = (\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n}) \oplus \mathfrak{s} \cong (\mathfrak{q}/\mathfrak{n}) \oplus \mathfrak{s}, \tag{3.56}$$

because $\mathfrak{h} \cap \mathfrak{n} = \mathfrak{h} \cap \mathfrak{n}_0$ by (3.49). We should note that the sum on the right-hand side is direct, not just semidirect, because of the basic commutation relation $[\mathfrak{g}, \mathfrak{q}] \subset \mathfrak{n}$; see Varadarajan (1974, §3.8.3).

3.8.4 The Lie groups and the Ad-mapping

Let G, Q, N_R, Q_R and S be the simply connected groups that correspond respectively to $\mathfrak{g}$, $\mathfrak{q}$, $\mathfrak{n}_R$, $\mathfrak{q}_R$ and $\mathfrak{s}$, and let N be the nilradical of Q. We have

$$G = Q \leftthreetimes S; \quad Q = N_R \leftthreetimes Q_R, \tag{3.57}$$

and as the algebra $\mathfrak{q}/\mathfrak{n}$ is Abelian, it will be identified with the group via the exponential mapping:

$$Q/N = V \text{ identified with } \mathfrak{q}/\mathfrak{n}. \tag{3.58}$$

Although inessential, note that $N_0 = N \cap Q_R$ is the analytic subgroup that corresponds to $\mathfrak{n}_0$ (see the exercise at the end of §2.12.2 or use the topological argument of Exercise 8.9 in Part II). Therefore, by (3.55) we have $N = N_R \leftthreetimes N_0 \subset N_R \leftthreetimes Q_R$ and in particular $N \cap Q_R = N_0$ is simply connected (see Varadarajan, 1974).

Now we shall use the exp mapping to obtain the analogue of the triangulation (3.51) for the group Q_R. But since for general soluble groups the exponential mapping is not bijective, we have to proceed with care as follows. The

Ad-action of Q_R on $\mathfrak{n}_i$ can be triangulated:

$$\mathrm{Ad}(x) = \begin{pmatrix} \omega_1(x) & & * \\ & \ddots & \\ 0 & & \omega_{t_i}(x) \end{pmatrix}; \quad x \in Q_R. \tag{3.59}$$

To see this we use the exponential $\mathfrak{q}_R \to Q_R$, which is a bijection on a small neighbourhood of 0 in $\mathfrak{q}_R$, to yield $\mathrm{Ad}(\exp \xi) = e^{\mathrm{ad}\,\xi}$. This implies that the triangulation (3.59) holds for $x \in \exp(\mathfrak{q}_R)$ and that $\omega_j(x) = e^{\rho_j(\xi)}$ when $x = \exp \xi$ and ρ_j is as in (3.51). Since $\exp \mathfrak{q}_R$ generates Q_R it follows that (3.59) holds for all $x \in Q_R$ and that $\omega_j \colon Q_R \to \mathbb{C}\backslash\{0\}$ are group homomorphisms. By the simple-connectedness of Q_R we can then define homomorphisms $\theta_j \colon Q_R \to \mathbb{C}$ such that $\omega_j(x) = e^{\theta_j(x)}$ and $\theta_j(\exp \xi) = \rho_j(\xi)$, $\xi \in \mathfrak{q}_R$.

Now let us denote the canonical

$$\alpha \colon Q_R \longrightarrow \frac{Q_R}{N \cap Q_R} = \frac{Q}{N} = V;$$
$$d\alpha \colon \mathfrak{q}_R \longrightarrow \frac{\mathfrak{q}_R}{\mathfrak{n} \cap \mathfrak{q}_R} = \frac{\mathfrak{q}}{\mathfrak{n}}.$$

For typographical reasons, in what follows we shall drop all the indices. For each $\omega = \omega_j$ and $\rho = \rho_j$ as above we shall define $\tilde{\rho} \in \mathrm{Hom}_{\mathbb{R}}[\mathfrak{q}/\mathfrak{n}; \mathbb{C}]$ such that $\rho = \tilde{\rho} \circ d\alpha$ and then using (3.58) we can define $\tilde{\omega}(y) = e^{\tilde{\rho}(\zeta)}$ for $\zeta \in \mathfrak{q}/\mathfrak{n}$ and $y = \exp \zeta \in V$. For the definition of $\tilde{\rho}$ we use the fact that ρ vanishes on $\mathfrak{n} \cap \mathfrak{q}_R$. One can then readily verify that (Exercise 3.10 below)

$$\omega(x) = \tilde{\omega} \circ \alpha(x); \quad x \in Q_R. \tag{3.60}$$

Using this we shall then use (and abuse) the same notation as before and for $x \in Q_R$ and $\zeta \in \mathfrak{q}/\mathfrak{n}$ such that $\exp \zeta = \alpha(x)$ we shall denote $Lx = \mathrm{Re}\,\tilde{\rho}(\zeta)$.

The bottom line (via the identification (3.58)) is that this $L \in V^*$ can be identified with linear functionals on $Q/N = V$ and that $|\omega(x)| = e^{L(x)}$, $x \in Q_R$. Note that the above L is the real root L_i that corresponds to the space $\mathfrak{n}_i$ of (3.53).

Exercise 3.10 Prove (3.60). Since both sides of (3.60) are homomorphisms $Q_R \to \mathbb{C}^* = \mathbb{C}\backslash\{0\}$ it suffices to verify this for $x = \exp \xi$ for some $\xi \in \mathfrak{q}_R$ because these give a neighbourhood of the identity of Q_R. To do that let $\zeta = d\alpha(\xi) \in \mathfrak{q}/\mathfrak{n}$ and $y = \alpha(x) = \exp \zeta$. We then have $\omega(x) = e^{\theta(x)} = e^{\rho(\xi)} = e^{\tilde{\rho} \circ d\alpha(\xi)} = e^{\tilde{\rho}(\zeta)}$ and also $\tilde{\omega} \circ \alpha(x) = \tilde{\omega} \circ \alpha(\exp \xi) = \tilde{\omega} \exp(d\alpha(\xi)) = \tilde{\omega}(\exp \zeta) = e^{\tilde{\rho}(\zeta)}$.

The intertwining with S In the next section we shall make use of the following. With the usual notation $x^s = sxs^{-1}$ in the group and the ω_j as in (3.59), we have

$$\omega_j(x^s) = \omega_j(x); \quad x \in Q_R,\ s \in S, \tag{3.61}$$

and therefore also, if we use the above abuse of notation,

$$L_i(x^s) = L_i(x); \quad x \in Q_R,\ s \in S,\ 0 \leqslant i \leqslant r. \tag{3.62}$$

In this conjugation, the matrix of (3.59) on the other hand changes when we pass from x to x^s. The reason for the above is that $x^{-1}sxs^{-1} \in N \cap Q_R$. (See Varadarajan, 1974, §§3.8.3, 3.18.7; more explicitly in this second reference set $\mathfrak{a} = \mathfrak{g}$, $\mathfrak{b} = \mathfrak{q}$, $\mathfrak{h} = [\mathfrak{q}, \mathfrak{g}]$. More directly, assume, as we may, that $x = \exp\xi$ and s are close to the identity with $\xi \in \mathfrak{q}$. Then by Varadarajan, 1974, (3.18.3) and §3.8.3, $(\mathrm{Ad}\,s)\xi - \xi \in \mathfrak{n}$ and since $x^{-1}sxs^{-1} = \exp(-\xi)\exp((\mathrm{Ad}\,s)\xi)$ we can use §§2.15, 3.8.3 of Varadarajan, 1974.) Therefore $\mathrm{Ad}(x)$ and $\mathrm{Ad}(x^s)$ on $\mathfrak{n}_i$ have the same diagonal coefficients.

3.9 An Alternative Proof of the NC-Theorem

3.9.1 Products of triangular matrices

In this section we shall give a version of Lemma 3.4 that applies to the soluble group Q_R. In other words, we have to get around the root space decomposition and the diagonal blocks of (3.15). Let $M_1, \ldots \in M_{r\times r}(\mathbb{C})$, for some $r \geqslant 1$, be a sequence of complex matrices of the form $M_j = D_j + T_j$ where $D_j = \mathrm{diag}(d_1^{(j)}, \ldots, d_r^{(j)})$, and T_j is upper triangular, that is, $T = (t_{\alpha,\beta}^{(j)})$ with $t_{\alpha,\beta}^{(j)} = 0$ if $\alpha \geqslant \beta$. The bounds that will be imposed on these matrices are

$$|d_i^{(j)}|^{-1} \leqslant A, \quad \delta_j = \mathrm{Max}\left(|d_1^{(j)}|, \ldots, |d_r^{(j)}|\right) \leqslant A, \quad |t_{\alpha,\beta}^{(j)}| \leqslant A, \tag{3.63}$$

or equivalently, $\|M_j\|, \|M_j^{-1}\| \leqslant A$ for some other fixed $A \geqslant 1$. We can then write

$$M_j = \delta_j(D_j' + T_j'), \quad \|D_j'\| = 1 \tag{3.64}$$

for similar matrices D', T' that are diagonal and triangular respectively, and now

$$\left.\begin{aligned} &M_1 \cdots M_n = \delta_1\delta_2\cdots\delta_n \prod_{j=1}^{n}(D_j' + T_j') = \delta_1\delta_2\cdots\delta_n\Sigma;\\ &\Sigma = \sum_{\substack{\varepsilon_k=\pm\\ k=1,\ldots,n}} B_1^{\varepsilon_1}\cdots B_n^{\varepsilon_n}, \quad B_k^+ = D_k', \quad B_k^- = T_k'; \quad k = 1, \ldots, n. \end{aligned}\right\} \tag{3.65}$$

Since the product in each term of Σ is zero if more than r among the ε_k are negative, it follows that there exist constants C, independent of A, n and the δ_j such that we have the following generalisation of (3.19):

$$\|M_1 \cdots M_n\| \leqslant C\delta_1\delta_2\cdots\delta_n(1+n)^C A^{2r}. \tag{3.66}$$

Note that in the applications below we shall even have

$$\delta_j = |d_1^{(j)}| = \cdots = |d_r^{(j)}|. \tag{3.67}$$

3.9.2 Proof for soluble groups

We shall assume here that G is soluble and simply connected, with $\mathfrak{g}$ its Lie algebra and $\mathfrak{g} = \mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R$ and $G = N_R \leftthreetimes Q_R$ are as in §3.8.4.

We consider $x_j = m_j q_j \in G, m_j \in N_R, q_j \in Q_R, |x_j|_G \leqslant C$ and use the transformation (2.21) and the same notation as before:

$$\begin{aligned} s_n &= x_1 \cdots x_n = m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \dot{s}_n; \\ \dot{s}_j &= \dot{x}_1 \cdots \dot{x}_j = q_1 \cdots q_j,\ m_j \in N_R,\ q_j \in Q_R. \end{aligned} \tag{3.68}$$

To estimate m^q (here $m \in N_R, q \in Q_R$) we use the exponential mapping of N_R and, with $m = \exp\zeta$, $\zeta \in \mathfrak{n}_R$ and $|\zeta|_{\mathfrak{n}_R} \leqslant C$, we have $m^q = \exp((\mathrm{Ad}\,q)\zeta)$ and, as in 3.5.2,

$$|m_{j+1}^{\dot{s}_j}|_{N_R} \leqslant C\|\,\mathrm{Ad}(q_1\cdots q_j)\|_{\mathfrak{n}_R} \leqslant C \sup_{1\leqslant i\leqslant r} \|\,\mathrm{Ad}(q_1\cdots q_j)\|_{\mathfrak{n}_i} \tag{3.69}$$

for the decomposition (3.48) of $\mathfrak{n}_R$. A direct application of (3.66) gives in (3.69) the estimate

$$(1+j)^C \sup_{1\leqslant i\leqslant r} \exp(L_i(q_1) + \cdots + L_i(q_j)), \tag{3.70}$$

where, as at the end of §3.8, the L_i are the real roots as in (3.44) with the identifications (3.52), (3.53) and the abuse of notation of §3.8.4. This is exactly the same estimate as in (3.21). With the help of this estimate we can finish the proof of the NC-theorem in this soluble case by the same argument *verbatim*: explicitly, the criterion of §3.3.3 is used on the sets

$$A_n = [m \in N_R;\ |m| \leqslant cn^c] \times [q \in Q_R;\ |q| \leqslant cn^c] \tag{3.71}$$

for appropriate constants and with $z(n) = s_n \in G$ the Markov chain on G. Condition (i) of the criterion is obvious. In considering the measures of the corresponding sets in this condition (i), note that the groups N_R and Q_R are unimodular. For condition (ii) we denote, as in (3.25),

$$\Omega_{\mathscr{L}} = [v;\ L_i v < 1,\ 1 \leqslant i \leqslant r] \subset V = Q/N. \tag{3.72}$$

The real roots L_i, as at the end of §3.8, can be identified with linear functionals on V because of (3.52), (3.53) and §3.8.4. Finally, we use the gambler's ruin estimate (3.26) and the estimate (3.70). From these, condition (ii) of the criterion follows.

3.9.3 General amenable NC-groups

Let G be such a group and let

$$G = N_R \leftthreetimes G_R; \quad G_R = Q_R \leftthreetimes S \tag{3.73}$$

be as in (3.57). We shall follow closely the proof that was given in §3.5 for the semidirect product $G = N \leftthreetimes K$, with $K = H \oplus S$.

The criterion of §3.3.3 is used again for the sets

$$A_n = [m \in N_R;\ |m| \leqslant cn^c] \times [q \in Q_R;\ |q| \leqslant cn^c] \times S \tag{3.74}$$

and condition (i) of the criterion is again obvious.

For condition (ii) of the criterion we write $x_j = m_j q_j \sigma_j = m_j \dot{x}_j$, $m_j \in N_R$, $q_j \in Q_R$, $\sigma_j \in S$ and

$$z(n) = x_1 \cdots x_n = m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \dot{s}_n; \quad \dot{s}_j = \dot{x}_1 \cdots \dot{x}_j \in G_R, \tag{3.75}$$

and we must prove the analogue of (3.28) and to do this we must estimate the $|\ |_{N_R}$ of

$$m_{j+1}^{\dot{s}_j} = (q_1\sigma_1 \cdots q_j\sigma_j) m_{j+1} (q_1\sigma_1 \cdots q_j\sigma_j)^{-1}. \tag{3.76}$$

Here, unlike what happened in (3.28) in §3.5.2, the q_j do not commute with the σ_j. This is the only complication compared to (3.28). In (3.76) we shall commute the σ_i with the q_i (i.e. 'jump over' them) so that we can rewrite (3.76) in the form

$$\tilde{\sigma}(\tilde{q}_1 \cdots \tilde{q}_j) m_{j+1} (\tilde{q}_1 \cdots \tilde{q}_j)^{-1} \tilde{\sigma}^{-1},\ \tilde{q}_k = q_k^{\tilde{\sigma}_k}; \quad \text{for appropriate } \tilde{\sigma}, \tilde{\sigma}_1, \ldots \in S. \tag{3.77}$$

We shall use the fact that S is compact and apply (3.66). This, together with (3.62), gives the analogue of (3.69), (3.70), namely

$$\|\operatorname{Ad}(\tilde{q}_1 \cdots \tilde{q}_j)\| \leqslant C(1+j)^C \sup_{1 \leqslant i \leqslant r} \exp(L_i(q_1) + \cdots + L_i(q_j)). \tag{3.78}$$

To estimate this we can use, again, the gambler's ruin estimate (3.26) in the conical domain $\Omega_{\mathscr{L}} \subset V$ of (3.72) where Q and N are the radical and the nilradical of G. Some care is needed to justify the use of the gambler's ruin estimate. We use the projection

$$\pi\colon G_R \to (Q_R \leftthreetimes S)/N_0 = Q/N \oplus S = V \oplus S, \tag{3.79}$$

induced by (3.56), where N_0 is the normal subgroup that corresponds to $\mathfrak{n}_0$ in (3.46) (abusing notation by using the same letter π in (3.79) and (3.56)). As pointed out in (3.52) and (3.56), $L_i(q)$, $q \in Q_R$, depends only on $(\pi(q))_V \in V$, the V component of $\pi(q)$ in (3.79). Furthermore, when $\dot{x}_j = q_j \sigma_j$, the V-components of $\pi(\dot{x}_j)$ are the same as the V-components of $\pi(q_j)$. These are therefore *symmetric* independent V-valued random variables. The gambler's ruin estimate can therefore be used: note that this holds despite the fact that the q_j are not necessarily symmetric random variables in Q_R. From this, condition (ii) of the criterion follows and we are done. This last point on the symmetry of the variables is subtle and it relies on the fact that on the right-hand side of (3.79) we have a direct sum and not just a semidirect one. More explicitly, using this we can exploit the fact that the random variables $\dot{x}_j$, being defined by the projection $G \to G_R$, *are* symmetric.

3A Appendix: The Gambler's Ruin Estimate

3A.1 One-dimensional case

Let $X_1, \ldots \in \mathbb{R}$ be centred independent identically distributed random variables and let $S_0 = h > 0$ be fixed and $S_j = h + X_1 + \cdots + X_j$ be the corresponding martingale. Let $\tau = \inf[n;\ S_n \leqslant 0]$ the first exit time from $]0, +\infty[$. We shall assume $\sigma^2 = \mathbb{E}(X_j^2) < +\infty$. By the martingale property (see Williams, 1991, §10.9) we then have

$$h = \mathbb{E}(S_{n\wedge\tau}) \leqslant \mathbb{E}([\tau \geqslant n]S_n) \leqslant \mathbb{P}^{1/2}(\tau \geqslant n)\mathbb{E}(S_n^2)^{1/2}, \tag{3A.1}$$

because $S_\tau \leqslant 0$ when $\tau < \infty$ (in fact $\tau < \infty$ almost surely by the zero–one law (Chung, 1982; Feller, 1968) but this information was not used) and where here and throughout, by abuse of notation, $\tau \geqslant n$ refers both to the event and to its indicator function. We have $\mathbb{E}(S_n^2) = h^2 + n\sigma^2$ and therefore

$$\mathbb{P}(\tau \geqslant n) \geqslant \frac{h^2}{h^2 + n\sigma^2}. \tag{3A.2}$$

From this, in the set-up of §3.3.5 we can conclude that

$$\mathbb{P}[X_1 + \cdots + X_j > 0;\ j = 1, \ldots, n] \geqslant cn^{-c}, \tag{3A.3}$$

as required in (3.11). To see this we fix the positive h', h'' such that $\mathbb{P}(h' < X_1 < h'') > 0$, we condition $X_1 + \cdots + X_j$ on $[h' < X_1 < h'']$ and then apply (3A.2). Already in this case the sharp estimate $\sim (\frac{h}{\sqrt{n}})$ is much harder to prove (see Feller, 1968). However, the proof that we gave here is simple and, as we shall see, it easily generalises to higher dimensions.

Most readers must have seen what we just did in a first course in probability theory. What follows, and the extensions to higher dimensions, requires some familiarity with potential theory. My advice to readers not willing to think in these terms is to skip the rest of this appendix.

3A.2 The Markov chain in $V = \mathbb{R}^d$

Let $X_1, \ldots \in V$ be independent centred identically distributed random variables with finite moments of high enough order. We shall fix $H \in V$, $S_0 = H$, $S_j = S_0 + X_1 + \cdots + X_j \in V$, $j \geqslant 1$ which is a Markov chain.

The principle that will be used is the following. We shall consider real continuous functions u on V that are subharmonic with respect to the above Markov process. This means that the process $u(S_j)$ is a submartingale. We shall also normalise $u(H) = 1$. Further, let τ be some stopping time that satisfies

$$u(S_\tau) \leqslant 0 \quad \text{when } \tau < +\infty. \tag{3A.4}$$

We then have, as in (3A.1) (by Williams, 1991, §10.9),

$$1 \leqslant \mathbb{E}(u(S_{n\wedge\tau})) \leqslant \mathbb{E}([\tau \geqslant n]u(S_n)) \leqslant (\mathbb{P}(\tau \geqslant n))^{1/2}\mathbb{E}(u^2(S_n))^{1/2}. \tag{3A.5}$$

We shall also impose the condition

$$u(x) = O(|x|^A) \quad \text{for some } A > 0. \tag{3A.6}$$

This, combined with (3A.5), gives

$$\mathbb{P}(\tau > n) \geqslant Cn^{-c}; \quad n \geqslant 1, \tag{3A.7}$$

for appropriate constants C, c, because the moment condition imposed on the variables and (3A.6) imply that the cofactor in the right-hand side of (3A.5) grows polynomially.

3A.3 Normal variables

Let $X_1, \ldots$ be assumed to be normal as in §3.6.2; then it is possible to perform a linear change of coordinates so that $X_1 + \cdots + X_j = b(j)$ for standard Brownian motion $b(t) \in V$. The process $u(S_j)$ is then a submartingale if $\triangle u \geqslant 0$, in the distribution sense, for the (appropriately normalised) Euclidean Laplacian $\triangle = c\sum \frac{\partial^2}{\partial x_i^2}$ (see Chung, 1982).

3A.4 Proof of (3.33) for normal variables

Let Ω, σ_0 be, as in (3.12), a conical domain and let the variables $X_1, \ldots$ be normal, as in §3.6.2. The continuous subharmonic function u in (3A.4) is defined in $\mathbb{R}^d$ and $u \geqslant 0, u(H) = 1, u \equiv 0$ in $V \backslash \Omega$, where $H = h\sigma_0$ for some large $h > 0$. The existence of such a function will be proved presently. With S_j as in §3A.2 we set $\tau = \inf[n;\ S_n \notin \Omega]$. We therefore obtain (3A.7) and, by the same conditioning as in §3A.3, we deduce the required (3.33).

3A.5 The construction of the subharmonic function

Normalise and set $H = (h,0,0,\ldots,0)$ and $\sigma_0 = (1,0,\ldots,0)$, the north pole of the unit sphere. We use polar coordinates $x = r\sigma$ and $\sigma = (\theta, \varphi_1, \ldots, \varphi_{d-2}) \in \Sigma_{d-1}$ for the local coordinates on the unit sphere, where $(\varphi_1, \ldots, \varphi_{d-2}) \in \Sigma_{d-2}$ and θ is the colatitude, that is, the angle with σ_0. We then use spherical harmonics $F(x)$ (i.e. homogeneous polynomials on $V = \mathbb{R}^d$ that are harmonic), and we shall in particular consider zonal harmonics (see Szegö, 1939) of the form

$$F(x) = r^k P(\theta); \quad r = |x|; \quad \sigma = x/|x| = (\theta, \varphi_1, \ldots, \varphi_{d-2}). \tag{3A.8}$$

For $k_0, \theta_0 > 0$ it is then possible to find such a zonal harmonic that satisfies

$$\begin{gathered} k \geqslant k_0; \quad P(\theta) > 0, \quad |\theta| < \theta_1; \\ P(\theta) < 0, \quad \theta \in [\theta_1, \theta_2]; \quad \theta_2 \leqslant \theta_0, \end{gathered} \tag{3A.9}$$

that is, $P(0) > 0$, then P dips and becomes negative for $|\theta| > \theta_1$.

This in particular supplies us with the subharmonic function used in §3A.4. To see this we simply set $u = F$ in $\Omega_{\theta_1} = [|\theta| < \theta_1]$ and $u = 0$ otherwise. Since Ω_{θ_1} is a conical domain that can be made arbitrarily thin, it follows that u satisfies the properties of the function in §3A.4.

3A.6 Proof of (3.13)

Here $X_1, \ldots$ are bounded variables as in §3.3.6. We shall preserve the notation Ω, $H = h\sigma_0$, u, $F = r^k P(\theta)$, θ_1, θ_2, etc. of §§3A.4 and 3A.5, and we shall define the perturbation

$$U = u + r^{k-1/2}. \tag{3A.10}$$

Furthermore, we shall normalise and, writing μ for the distribution of the variables X_j, we shall assume $\int x_i x_j \, d\mu = \delta_{ij}$ (which is the identity matrix). For the corresponding Laplacian $\widetilde{\triangle} = \mu - \delta$ (where δ is the Dirac mass at 0), this implies that

$$|(\widetilde{\triangle} - \triangle)f(x)| \leqslant C \sup_{|y-x|\leqslant A} |\nabla^3 f(y)|; \quad f \in C^\infty, \tag{3A.11}$$

by Taylor's remainder term, where A is the diameter of the support of μ, $\nabla^3 = \partial^3/\partial x_i\,\partial x_j\,\partial x_k$ denotes the third gradient, and C in (3A.11) depends only on the dimension. Here, and in what follows, $\widetilde{\triangle}$ acts as a convolution operator.

By (3A.11) it follows in particular that if p is large enough there exist C, c such that

$$\widetilde{\triangle} r^p \geqslant c r^{p-2}; \quad r > C. \tag{3A.12}$$

We shall show that this implies that for all k large enough there exists C such that

$$\widetilde{\triangle} U(x) > 0 \quad \text{for } |x| > C, \tag{3A.13}$$

and this says that U is subharmonic for the measure μ in that region. (For a formal definition of subharmonicity in the above sense see §5A.1.)

Exercise Prove this. For $x \notin \Omega_{\theta_1}$, use (3A.12), $u \geqslant 0$ and $u(x) = 0$. For $x \in \Omega_{\theta_1}$, note that $\widetilde{\triangle} u \geqslant \widetilde{\triangle} F$ so we need to show that $\widetilde{\triangle}(F + r^{k-(1/2)})(x) > 0$. But since F is harmonic, by (3A.11), $\widetilde{\triangle} F = O(r^{k-3})$ and therefore by (3A.12), it is $\widetilde{\triangle} r^{k-(1/2)}$ that is dominant here.

Now let us denote the truncated cone and the exit time by

$$\Omega(C, \theta') = [r > C;\ |\theta| < \theta']; \quad \tau = \tau_{C,\theta'} = \inf[n;\ S_n \notin \Omega(C, \theta')]. \tag{3A.14}$$

With the notation of (3A.9), we fix some $\theta_1 < \theta' < \theta_2$. Then if C in (3A.14) is large enough, there exists C_1 such that

$$U(S_\tau) \leqslant C_1 \quad \text{when } \tau = \tau_{C,\theta'} < \infty.$$

Such a C_1 is not necessarily 0 because we may exit at a point that is not far out. For $h > 0$ large enough we obtain therefore the following substitute for (3A.5):

$$\begin{aligned} U(H) &\leqslant \mathbb{E}(U(S_{n\wedge\tau})) \\ &\leqslant \mathbb{E}([\tau \geqslant n] U(S_n)) + C_1 \\ &\leqslant (\mathbb{P}(\tau > n))^{1/2} \mathbb{E}(U^2(S_n))^{1/2} + C_1, \end{aligned}$$

and for h sufficiently large we deduce $\mathbb{P}(\tau > n) = O(n^{-c})$. This, for the random variables X_j of §3.3.6, clearly gives the proof of (3.13) by the same conditioning $X_1 + X_2 + \cdots + X_{n_0} \in$ (small neighbourhood of H) for n_0 large enough.

Note The ad hoc constructions of this appendix are what one finds, more or less, in Varopoulos (1994b). Giving a systematic exposition and finding exact constants is a difficult problem which, however, is needed if we are to prove (3.3) for the local central limit theorem. This was undertaken by the author in a long series of papers over a period of 15 years (Varopoulos, 1999c, ..., Varopoulos, 2014: you can find the full list in MathSciNet if you so wish). All this however, as already said, lies outside the scope of this book and we feel that not many of the readers would care to spend time on this problem.

4
The B–NB Classification

Overview of Chapter 4

In this chapter we shall address the case of general groups that are not necessarily amenable and we shall prove the upper B-estimate of §1.3 for an arbitrary spectral gap $\lambda \geqslant 0$.

The important new aspect of the theory lies in the algebraic B–NB classification that will be given in §4.1. This builds on the previous C–NC classification of Chapter 2 but the algebra that will be needed here is less standard and in particular we shall use the Iwasawa decomposition for semisimple Lie algebras (Helgason, 1978, Chapter VI). Only the bare bones of what is needed will be recalled in §4.1 and then we shall get on immediately with the main definition. Much more information on the ambient algebraic theory will be given in Appendix A, Appendix B and Appendix C at the end of Part I of the book. These appendices are not absolutely essential for this chapter but they will certainly clarify the picture and they will also guide the reader to navigate the vast literature of semisimple groups. Consistent with the general scheme that we explained in the overview of Chapters 2 and 3, we shall start in §4.3 by identifying the special class of groups on which the theorem has to be proved first. Here, this class contains essentially all the simply connected groups. The actual generalisation to all the connected Lie groups is done in §4.6 (and again later in §5.8*) and this is quite involved. The reader should ignore this side of things in a first reading. In §4.3 we also give a general plan of the proof.

The heart of the matter lies in §§4.4–4.5 where we prove the theorem in the more general setting of a principal bundle. Principal bundles are of central importance in geometry but for us here they are just a creature that allows us to establish convenient notation and add transparency to the proof (see, however, §14.1.5 later on for a formal definition). The ad hoc definitions that we give in §4.4 are quite formal but the reader should take them seriously and spend time to understand them properly because once we have these the actual proof of the theorem follows in just a few lines in §4.5.

In the same spirit as in the previous two chapters, in §4.6 we show that the methods that we developed are good enough to prove our theorem for a special class of simply connected groups. Techniques from Lie groups (e.g. the structure of these groups) are then used to pass to the general case. As we have already pointed out, these techniques are external to the main theme and the reader could skip this section altogether. It is, however, also true that this section, especially §§4.6.1–4.6.2, gives a good idea of what one has to do to pass from the 'special' to the 'general'.

There is a second part (Part 4.2) to this chapter, as there was for the previous one. Here, as before, we prove the same theorem for the heat diffusion kernel of an invariant differential operator on the group. As for the previous chapter, the reader could skip this second part on a first reading.

Part 4.1: The Basic Theorem

4.1 The Lie Algebras

Let $\mathfrak{g}$ be some real Lie algebra, let $\mathfrak{q} \lhd \mathfrak{g}$, $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ be its radical and some Levi decomposition where $\mathfrak{s}$ is semisimple. We can uniquely decompose $\mathfrak{s}$ as the direct sum of two algebras $\mathfrak{s} = \mathfrak{s}_n \oplus \mathfrak{s}_c$ where $\mathfrak{s}_c$ is of compact type, that is, is the Lie algebra of some compact group, and $\mathfrak{s}_n$ is of non-compact type, that is, no direct factor of $\mathfrak{s}_n$ is the Lie algebra of a compact group (see Helgason, 1978, §II.6). When $\mathfrak{s}$ is of non-compact type we shall use the Iwasawa decomposition:

$$\mathfrak{s} = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k} \quad \textit{(direct vector space sum of subalgebras).} \tag{4.1}$$

Here $\mathfrak{n}_s$ is nilpotent, $\mathfrak{a}$ is Abelian, $[\mathfrak{n}_s, \mathfrak{a}] \subset \mathfrak{n}_s$ and $\mathfrak{k}$ is the Lie algebra of some compact group K. Here K could be Abelian, that is, some torus. This is an important construction that is extensively used in the theory of symmetric spaces and representation theory (see Helgason, 1978, Chapter VI; Gangoli and Varadarajan, 1980; Varadarajan, 1974). In Appendix A, Appendix B, Appendix C at the end of Part I, we shall elaborate on the aspects of the Iwasawa decomposition that are relevant to us. For our immediate use however, the only thing that will be needed is what is described in the next few lines. The reader who is not familiar with these notions is urged to take these facts for granted and proceed to an understanding of the main definitions of this chapter. We shall make the convention and agree to write the decomposition (4.1) for an arbitrary semisimple algebra $\mathfrak{s} = \mathfrak{s}_n \oplus \mathfrak{s}_c$ as follows. We write $\mathfrak{s}_n = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k}_n$, as in (4.1), where everything is taken to be zero when $\mathfrak{s}_n = \{0\}$, and we then write $\mathfrak{s} = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k}$ with $\mathfrak{k} = \mathfrak{k}_n + \mathfrak{s}_c$. With this notation we write

$$\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_s + \mathfrak{a}, \ \mathfrak{g} = \mathfrak{r} + \mathfrak{k}, \tag{4.2}$$

and it is clear that $\mathfrak{r}$ is a soluble algebra since $[\mathfrak{q}, \mathfrak{s}] \subset \mathfrak{q}$.

The definition of the subalgebra $\mathfrak{r}$ is not unique, since the $\mathfrak{s}$, $\mathfrak{n}_s$, $\mathfrak{a}$ are not uniquely defined, and $\mathfrak{r}$ is not in general an ideal. Observe, however, that by the above convention for amenable algebras (cf. §3.1.5) we have $\mathfrak{r} = \mathfrak{q}$. We shall nonetheless call $\mathfrak{r}$ the Iwasawa radical and we have the following result.

Proposition 4.1 *Let $\mathfrak{g}$ be some real Lie algebra and let $\mathfrak{r}_1, \mathfrak{r}_2 \subset \mathfrak{g}$ be two Iwasawa radicals. Then there exists $x \in G$ the simply connected Lie group that corresponds to $\mathfrak{g}$ such that $(\mathrm{Ad}x)\mathfrak{r}_1 = \mathfrak{r}_2$; that is, $\mathfrak{r}_1$, $\mathfrak{r}_2$ are conjugate by an inner automorphism of $\mathfrak{g}$.*

The proof of this is based on the fact that the same assertion holds for the Levi subalgebra (see Varadarajan, 1974, §3.14.2). This proof, together with a short overview of the Iwasawa decomposition, will be given in Appendix A, Appendix B and Appendix C at the end of Part I of the book. But these algebraic proofs can in fact be bypassed altogether because, as we shall see, this proposition is inessential for the proof of our main classification theorem.

Definition 4.2 We call $\mathfrak{g}$ a B-*algebra* if the Iwasawa radical $\mathfrak{r}$ is a C-algebra.

We call $\mathfrak{g}$ an NB-*algebra* if $\mathfrak{r}$ is an NC-algebra.

Similarly, if G is a connected Lie group that corresponds to $\mathfrak{g}$, we shall say that G is *algebraically*-B or *algebraically*-NB respectively.

The reader at this point could anticipate and look at §A.8 where a 'smoother' formulation of the algebraic classification is given (without proof).

Before a proof of the proposition is given, note that, a priori in Definition 4.2, an algebra could be both B and NB, and the proposition is needed to guarantee that the above is a classification. We shall, however, see in the statement of the main theorems in the next section that this proposition is not necessary, and that these theorems provide an a posteriori proof of the fact that the above is indeed a classification.

The properties that I shall state below are related to the B–NB classification of Lie algebras and are not essential for the main theorem and therefore their proofs will be deferred till Appendix A, Appendix B and Appendix C.

It is an easy exercise to see that a product algebra $\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{g}_2$ satisfies the NB-condition if both $\mathfrak{g}_1$, $\mathfrak{g}_2$ do. The following two propositions are, however, less trivial.

Proposition 4.3 *If* $\mathfrak{g}$ *is an* NB-*algebra then it is an* NC-*algebra*.

We recall that in §2.1 we defined $\mathfrak{g}$ to be an NC-algebra if its radical $\mathfrak{q}$ is a (soluble) NC-algebra.

We recall now (see §2.3.5) that the Lie group G of the Lie algebra $\mathfrak{g}$ is unimodular if the trace $\mathrm{Tr}(\mathrm{ad}x) = 0$ for all $x \in \mathfrak{g}$. Such an algebra will be called a unimodular algebra.

Proposition 4.4 *Let* $\mathfrak{g}$ *be some unimodular algebra. Then exactly one of the following conditions* (i) *or* (ii) *below holds*:

(i) $\mathfrak{g}$ *is a* B-*algebra*;
(ii) $\mathfrak{g} = \mathfrak{g}_{\mathrm{R}} \times \mathfrak{s}$ (*direct product*) *where* $\mathfrak{g}_{\mathrm{R}}$ *is an* R-*algebra* (*see* §*2.2.2*) *and* $\mathfrak{s}$ *is a semisimple algebra and then the algebra is* NB.

The notation $\times$ or $\oplus$ will be used for direct products of Lie algebras or groups.

The proposition implies that the converse of Proposition 4.3 fails. To see this it suffices to consider some semidirect product of an Abelian algebra with a semisimple algebra (of non-compact type) $\mathfrak{a} \leftthreetimes \mathfrak{s}$ that is not a direct product. Indeed, such an algebra is always unimodular because $\mathrm{Tr}(\mathrm{ad}s) = 0$ for all $s \in \mathfrak{s}$ (see §B.3).

As we have already pointed out in §1.2, and in §2.5.4, for unimodular groups a parallel, less precise but more general, theory exists.

Exercise 4.5 We saw in §2.2.1 that if $\mathfrak{q}$ is a soluble algebra and $\mathfrak{z} \subset$ is a central subalgebra then $\mathfrak{q}/\mathfrak{z}$ is a C- (resp. NC-)algebra if and only if $\mathfrak{q}$ is. Similarly, if $\mathfrak{z} \subset \mathfrak{g}$ is central, $\mathfrak{g}/\mathfrak{z}$ is B (resp. NB) if and only if $\mathfrak{g}$ is. This holds because any central subalgebra lies in the radical.

4.2 Statement of the Results

In this chapter we shall consider some general connected Lie group G and $d\mu(g) = \phi(g)\,d^r g \in \mathbb{P}(G)$ as in §1.3; that is, ϕ is continuous compactly supported and $\phi(e) \neq 0$. These conditions are satisfied automatically as we see by considering $d\mu^{*n}(g) = \phi_n(g)\,d^r g$, some convolution power of μ as in §2.4.1. It will also be convenient as before to abuse notation and write $\mu^{*n}(e) = \phi_n(e)$, and to avoid confusion with some of the notation that was used in Chapter 2, we stress that, here, all the factors in the convolution powers are identical. We shall use the notation of §3.1 and denote $\|\mu\|_{\mathrm{op}} = e^{-\lambda}$. The essential part of the classification of §§1.3 and 1.10 is then contained in the following two theorems.

Theorem 4.6 (B-theorem) *Let G and $\mu \in \mathbb{P}(G)$, $\lambda \geqslant 0$ be as above. Let us assume that G is algebraically-*B*. Then there exist $C, c > 0$ such that*

$$\mu^{*n}(e) \leqslant C\exp(-\lambda n - cn^{1/3}); \quad n \geqslant 1. \tag{4.3}$$

Theorem 4.7 (NB-theorem) *Let G and $\mu \in \mathbb{P}(G)$, $\lambda \geqslant 0$ be as above. Let us assume that G is algebraically-*NB *and that μ is a symmetric measure (i.e. that it is stable by the involution $g \mapsto g^{-1}$ in G). Then there exist $\alpha \geqslant 0$ and $C > 0$ such that*

$$\mu^{*n}(e) \geqslant Cn^{-\alpha}\exp(-\lambda n). \tag{4.4}$$

If G is compact, $\lambda = \alpha = 0$ (see §6.2). In (4.4), since μ is assumed to be symmetric, $e^{-\lambda} = \|\mu\|_{\text{op}}$ as in §3.1 and therefore $\mu^{*n}(e) \leqslant ce^{-\lambda n}$. Note that in (4.3) the measure is not assumed to be symmetric and that in §5.2 we shall prove a sharper version of (4.3) where $\|\mu\|_{\text{op}}$ is replaced by $\|\mu\|_{\text{sp}}$.

Provided that μ is symmetric, the fact that, in (4.3), we have an analogous lower estimate (see §1.3) is an easy by-product of the proof of the NB-theorem in Part 5.3. What is considerably harder, and this will not be done in this book, is to prove the full thrust of this theorem in §1.3 and to prove that the same exponent α in (4.4) can be used to give an upper and a lower estimate as asserted in §1.3. This says that in the NB case we have the generalisation of the local central limit theorem of (3.3). When G is semisimple, both (4.4) and the above local central limit theorem are a known result from Bougerol (1981). The general case of this sharp central limit theorem is unpublished work of the author. The proofs of both (3.3) in Varopoulos (1999b) and of the above local central limit theorem build of course on the methods of this book. The details, however, are so long to write down in full, that the present author at least may well never find the time and energy to do that.

Finally, as we promised, the two theorems put together prove a posteriori that the two properties, B and NB, do provide a classification of the groups and the algebras without the use of Proposition 4.1.

4.3 An Overview of the Proofs

Only the B-theorem, Theorem 4.6, will be proved in this chapter; the proof of the NB-theorem will be given in the next one. Both proofs are long, rather than difficult, because they depend on ad hoc machinery that has to be introduced. The underlying idea behind this machinery is that with the use of the Levi and Iwasawa decompositions we can generalise the notion of a connected Lie group G to the more general structure of a principal fibre bundle. Let us be more precise and concentrate first on the following.

4.3.1 Special class of connected Lie groups and the principal bundles

We shall assume that the connected group G is the semidirect product $G = Q \rtimes S$ where Q is a connected amenable group and S is semisimple of non-compact type. The semisimplicity implies that the centre of S is discrete and we shall make the additional assumption that the centre of S is finite. Any Lie group is locally isomorphic to a special group as above. To see this, one uses the Levi decomposition of Lie algebras (Varadarajan, 1974, §§3.14 and 3.15.1) and the only thing that has to be verified is the assertion about the finiteness of the centre of S. This will be done in §4.6 below. The theorems will first be proved for these special groups and then, by additional arguments that are not trivial, the results will be deduced for every other group.

For the above special groups, $S = NAK$ is the Iwasawa decomposition of S and the condition on the centre implies that K is compact (see Appendix A). We recall here that N is nilpotent and A is Abelian. Both are closed simply connected subgroups and the group multiplication induces a bijective diffeomorphism of S with $N \times A \times K$ (see Appendix A). The actual group G can then be written as a product $QS = QNAK = RK$ where $R = QNA$ is a connected closed amenable subgroup (see §3.1.6). Here Q is not a priori assumed to be simply connected but if it is, then R is simply connected. In the product decomposition $G = RK$ we shall only retain the product structure of R with K: $X = R \times K$ so that $G \ni rk = g \longleftrightarrow x = (r,k) \in X$, for $r \in R, k \in K$, and retain also the left action by R on X given by $r_1(r,k) = (r_1 r, k)$ and then X should be thought of as a (trivial) principal bundle over K where R is the structure group: $X \to K$. Since the fibre bundle is trivial we can in fact ignore that more general structure and simply say that $X = R \times K$ is a product space. There are, on the other hand, interesting situations where general (non-trivial) bundles have to be considered.

One important point that should be stressed is that the actual group structure of K is and should be ignored so that K is, say, some compact manifold or even a compact topological space (or even some Borel space assigned with some finite positive measure). We shall refer to the above creature $X = R \times K$ as an R-(trivial) principal bundle over K. In this structure R is always a connected amenable Lie group, and more often that not it will in fact be soluble.

The program now runs as follows:

Step (1) We shall generalise the statements of the theorems to R-principal fibre bundles over some compact space. This is done in §4.4.

Step (2) We shall prove the B-theorem, Theorem 4.6, in this more general context. This is done in §4.5.

Step (3) We shall deduce the theorems for our special class of groups. This generalisation provides a natural framework in which the proofs should be written and is a convenient way of formalising the notation. Without this, the notation would have been considerably harder to follow.

Step (4) Finally in §4.6 we shall deduce from Step (3), the B-theorem in full generality.

4.4 Left-Invariant Operators on an R-Principal Bundle

4.4.1 The formal definition

Here $X = R \times K$ will be a trivial principal bundle and our first task is to generalise the convolution operators $f \mapsto f * \nu$ by a positive measure on the group $G = RK$. These will be called left-invariant operators and will be linear operators

$$T : C_0^\infty(X) \to \text{space of locally bounded Borel functions on } X. \tag{4.5}$$

For our applications Tf will in fact be continuous and compactly supported in a sense that will presently become obvious. The conditions imposed on T will be positivity (i.e. $Tf \geqslant 0$ for all $f \geqslant 0$) and R-translation invariance. More explicitly, if we write $f_r(r_1, k)$ for $f(rr_1, k)$, with $r, r_1 \in R$, $k \in K$, we have $T(f_r) = (Tf)_r$, with $r \in R$. The above definition could have been given without imposing positivity but in our applications only positive operators will be considered and, furthermore, that positivity requirement eliminates awkward convergence and definition issues in the integral representation of these operators that we shall give below.

4.4.2 The coordinate representation of the operators

Let $\mu_{h,k}$, $h, k \in K$, denote a collection of positive measures on R that depends continuously (or at least Borel) on h, k. Further, let

$$f \mapsto Lf(h) = \int_K L(h, dk) f(k); \quad f \in C(K) \tag{4.6}$$

be a positive operator on $C(K)$ given by the kernel $L(h,dk)$, with $h,k \in K$. The definition of L extends to all positive Borel functions and we can define

$$\begin{aligned} Tf(r,h) &= \int_X f(rr_1^{-1},k)d\mu_{h,k}(r_1)L(h,dk) \\ &= \int_K L(h,dk)(f(\cdot,k) * \mu_{h,k})(r); \quad f \geqslant 0, \end{aligned} \tag{4.7}$$

where in the first integral we integrate in $(r_1,k) \in X$. In the above general definition these integrals could be $+\infty$, but under appropriate boundedness conditions they give rise to left-invariant operators on X. We shall use the notation

$$T = L \otimes \{*\mu\} = L(h,dk) \otimes \{*\mu_{h,k}\} \tag{4.8}$$

to express (4.7), and in the special case $f(r,k) = \varphi(r)\psi(k)$ we have

$$Tf(r,h) = \int L(h,dk)(\varphi * \mu_{h,k}(r))\psi(k). \tag{4.9}$$

All the left-invariant operators that we shall consider will be of this form (see §§4.4.6, 4.10 below for explicit computations) and in fact, under reasonable boundedness and smoothness conditions, one can easily see that every left-invariant operator is of that form (this will not be essential for us and at any rate it is an easy exercise in elementary measure theory – see Bourbaki, 1963).

The representation of such an operator as in (4.7) is not unique since we have for instance $\varphi L \otimes \{*\mu\} = L \otimes \{*\varphi\mu\}$ for any $\varphi(h,k) \geqslant 0$. We shall therefore impose the additional condition that all the measures $\mu \in \mathbb{P}(R)$ are probability measures and then the representation (4.7) becomes (essentially) unique. Such a representation will be called *normal*. Unless otherwise stated, all the representations of left operators that we shall consider in what follows will tacitly be assumed to be normal. For such a normal representation we have

$$\begin{aligned} Lf = g, \quad f \in C(K), \quad \text{if and only if} \\ T(\mathbf{1} \otimes f) = \mathbf{1} \otimes g; \quad \mathbf{1}(r) = 1, \quad r \in R. \end{aligned} \tag{4.10}$$

We should also note that with the notation (4.8) we have the following composition formula for a sequence $T_i = L_i \otimes \mu^{(i)}$, $i = 1,2,\dots$ of left-invariant operators:

$$\begin{aligned} T_1 \circ \cdots \circ T_n = \int_{k_1 \in K} \cdots \int_{k_{n-1} \in K} L_1(h,dk_1)L_2(k_1,dk_2) \cdots L_n(k_{n-1},dk) \\ \otimes \left\{ * \left(\mu^{(n)}_{k_{n-1},k} * \cdots * \mu^{(1)}_{h,k_1} \right) \right\}. \end{aligned} \tag{4.11}$$

From (4.10) it follows that the L operator that corresponds to $T = T_1 \circ \cdots \circ T_n$

is $L_1 \circ \cdots \circ L_n$ and for f as in (4.9) we have

$$Tf(r,h) = \int \cdots \int L_1(h,dk_1)\cdots L_n(k_{n-1},dk)\psi(k)\left(\varphi * \mu^{(n)}_{k_{n-1},k} * \cdots * \mu^{(1)}_{h,k_1}\right)(r). \tag{4.12}$$

This can be reformulated to say that $T = L\otimes\{*\mu_{h,k}\}$ and that $\mu_{h,k}$ is a convex combination of $(\mu^{(n)}_{k_{n-1},k} * \cdots * \mu^{(1)}_{h,k_1};\ k_1,\ldots,k_{n-1}\in K)$.

Finally, let χ be a positive multiplicative character on R, that is, $\chi(r)>0$, $\chi(r_1r_2)=\chi(r_1)\chi(r_2)$, with $r_1,r_2\in R$. A left-invariant operator $T = L\otimes\{*\mu\}$ can then be conjugated by multiplication by χ (i.e. $f\to\chi^{-1}T(\chi f)$), and for this operator we have

$$\chi T\chi^{-1} = L\otimes\{*\chi\mu\}. \tag{4.13}$$

Notice, however, that (4.13) will not in general be a normal representation.

4.4.3 Measures, adjoints and L^p-norms

As already pointed out, K, the basis of the fibre bundle, is some compact space and we shall assign to K a non-vanishing measure dk (i.e. the measure of every open subset is positive). In the case when K is a group we shall, more often than not, for reference measure take dk to be the normalised Haar measure of K.

In our concrete applications we shall always have $L(h,dk) = L(h,k)\,dk$ with $L(h,k)\in L^\infty$. The reference measure that we shall then consider on X will be fixed throughout to be $dx = d^r r\otimes dk$ for the right Haar measure $d^r r$ on R. It is then clear that for any $1\leqslant p\leqslant +\infty$ the L^p operator norms in (4.12) satisfy

$$\|T_1\circ T_2\circ\cdots\circ T_n\|_{p\to p}\leqslant \|L_1\circ\cdots\circ L_n\|_{p\to p}\sup_{k_i}\|*\mu_{k_n,k_{n+1}} * \cdots * \mu_{k_1,k_2}\|_{p\to p}, \tag{4.14}$$

where $\|\ \|_{p\to p}$ refers to the $L^p\to L^p$ operator norm on X, K and R with respect to the above reference measure, and for simplicity we suppress the j from $\mu^{(j)}$.

Exercise Prove (4.14). For simplicity assume that $n=2$, $T_1=T_2$, $T_1\circ T_2 = T^2$, $L\circ L = L^2$ and set $\|*\mu_{k,k_1} * \mu_{k'_1,k'}\|_{p\to p}\leqslant A$, $\|L^2\|_{p\to p} = B$. Then, by Minkowski,

$$\|T^2 f(\cdot,h)\|_p \leqslant \iint L(h,dk_1)L(k_1,dk)\|f(\cdot,k) * \mu_{k_1,k} * \mu_{h,k_1}\|_p$$
$$\leqslant A \iint L(h,dk_1)L(k_1,dk)\|f(\cdot,k)\|_p$$
$$\leqslant A \int L^2(h,dk)\|f(\cdot,k)\|_p;$$
$$\int \|T^2 f(\cdot,h)\|_p^p\, dh \leqslant A^p B^p \int \|f(\cdot,k)\|_p^p\, dk.$$

Then use $\| f \|^p_{L^p(X)} = \int \| f(\cdot,k) \|^p_p\, dk$, $\| T^2 f \|^p_{L^p(X)} = \int \|T^2 f(\cdot,h) \|^p_p\, dh$.

Similarly it is clear that if $T = L(h,k)\,dk \otimes \{\mu_{h,k}\}$, then the adjoint operator in $L^2(X;dx)$ is given by

$$T^* = L^*(h,k)\,dk \otimes \{*\mu^*_{h,k}\};$$
$$L^*(h,k) = L(k,h), \tag{4.15}$$
$$\mu^*_{h,k}(r) = \mu_{k,h}(r^{-1}).$$

4.4.4 The amenability of R

For the left-invariant operators $T = L \otimes \{*\mu\}$ that we shall be considering, there will exist some compact subset $C \subset R$ such that the supports of the probability measures satisfy

$$\text{support}\, \mu_{h,k} \subset C; \quad h,k \in K. \tag{4.16}$$

Under that assumption, we can assert that in (4.14) we actually have

$$\|T\|_{2\to 2} = \|L\|_{2\to 2}, \tag{4.17}$$

and in fact the same thing also holds for the L^p operator norms and also for products of operators as in (4.14). To see this we use the amenability of R. For the proof of (4.17) we proceed as follows. By the second definition of amenability in §3.1.4 we can find two sequences $0 \leqslant f_m, g_m \in C_0^\infty(R)$, $m \geqslant 1$ such that

$$\|f_m\|_2 = \|g_m\|_2 = 1,$$
$$\langle f_m * \pi, g_m \rangle_{L^2(R;d^r r)} \underset{m\to\infty}{\longrightarrow} 1, \tag{4.18}$$
$$\pi = \mu_{k_1,k_2} * \cdots * \mu_{k_{n-1},k_n}$$

and for fixed n the limit is uniform in the $k_i \in K$ where the notation is as in (4.11), (4.14) and where, as before, the j on $\mu^{(j)}$ and L_j has been suppressed.

For fixed $\varphi, \psi \in C(K)$ non-negative, with $\|\varphi\|_2, \|\psi\|_2 \leqslant 1$, we consider

$$\int_K \cdots \int_K L(h, dk_1) \cdots L(k_{n-1}, dk) \varphi(k) \psi(h) \langle f_m * \mu_{k_{n-1},k} * \cdots, g_m \rangle \, dh. \quad (4.19)$$

By (4.11), this is bounded by $\|T_1 \circ \cdots \circ T_n\|_{2 \to 2}$ for the $L^2(d^r r\, dk)$ norm. If we pass to the limit $m \to \infty$ and use (4.18), we obtain $\langle L_1 \circ \cdots \circ L_n \varphi, \psi \rangle$. We let φ, ψ vary and our assertion follows.

4.4.5 Convolution operators on a group

Here we shall go back to the special class of Lie groups considered in §4.3. As explained there, such a group can be written $G = RK$ and it can be identified with the principal bundle $X = R \times K$. The convolution $T\colon f \mapsto f * \nu$ by some $\nu \in \mathbb{P}(G)$ can then be identified with a left-invariant operator on the bundle. Since it is here that the motivation for the definitions that we gave in this section lies, we shall in the next few lines give the explicit formulas related to this operator. Another explicit formula, when $d\nu = \varphi(g)\, dg$, will be given in §4.10 below.

Clearly we can fix $g_1 \in G$ and restrict our attention to the case $\nu = \delta_{g_1^{-1}}$, where the Dirac δ is the mass at the point g_1^{-1}. The operator then becomes the right-translation operator

$$T_\nu f(g) = f * \nu(g) = f(gg_1). \quad (4.20)$$

Here $K = R \backslash G$ is the homogeneous space of left cosets $K = \{Rx;\ x \in G\}$. The right translation $g \mapsto gg_1$ induces an action $K \to K, k \to k[g_1]$. And there exists a function $\rho(g_1, k) \in R$, $k \in K$, called a cocycle, such that if $G \ni g = rk = (r,k) \in X$ we have

$$gg_1 = (r\rho(g_1,k),\ k[g_1]) \in X. \quad (4.21)$$

With an abuse of notation we can therefore say that for this operator (4.20) we have $L(h,dk) = 0$ unless $h = k[g_1]$ and we have a Dirac δ-mass for $h = k[g_1]$. As for the measures $\mu_{h,k}$, they can all be taken to be the corresponding δ-masses.

4.4.6 The Haar measure on $G = RK$

In the identification of G with $X = R \times K$, as in §4.3.1, the link between the relevant measures is supplied by $dg = d^\ell g = d^\ell r \otimes dk$, $g = rk$, for the left Haar measures of G, R and K; that is, we have

$$\int_G f(g)\, d^\ell g = \int_{R \times K} f(rk)\, d^\ell r\, dk; \quad f \in C_0^\infty(G). \quad (4.22)$$

The proof is easy. We have for sure $d^\ell g = \Phi(r,k)\, d^\ell r\, dk$, where Φ is the Jacobian. The left action $g \to rg$, $r \in R$ shows that $\Phi(r,k) = \Phi(k)$ is independent of r. The right action $g \mapsto gk$, $k \in K$ on the other hand stabilises $d^\ell g$ simply because K is compact. Hence Φ is a constant (see Helgason, 1984, §I.2; Bourbaki, 1963, Chapter 7).

If we denote now by m_G and m_R the modular functions of the groups G and R (see §1.1), we see that

$$\begin{aligned} d^r g = m_G(g)\, d^\ell g = m_G(g)\, d^\ell r \otimes dk& \\ = m_G(g) m_R^{-1}(r)\, d^r r \otimes dk\; &= \chi^2(r)\, dx; \\ g = rk \in G,\ x = (r,k) \in X,\ \chi &= m_G^{1/2}(r) m_R^{-1/2}(r), \end{aligned} \tag{4.23}$$

because $m_G \equiv 1$ on K. This means that for any operator, the operator norm of T on $L^2(G; d^r g)$ is the same as the operator norm of the operator $T_\chi = \chi T \chi^{-1}$ on $L^2(X; dx)$ simply because $f \to \chi^{-1} f$ is an isometry $L^2(X; dx) \to L^2(G; d^r g)$; here T_χ is multiplication by χ^{-1} followed by action by T, followed by multiplication by χ. On the other hand, T_χ is a left-invariant operator if T is because χ is multiplicative on R. When $Tf = f * \nu$ as in §4.4.5 we shall write (see §3.1)

$$T_\chi = L \otimes \{*\mu\}; \quad \|L\|_{2\to 2} = \|\nu\|_{\text{op}}, \tag{4.24}$$

where the second relation holds because of §4.4.4. The first illustration of this basic observation is to apply it to $\nu = \delta_{g_1{}^{-1}}$ and (4.20) where we now have

$$\begin{aligned} T_\chi f(r,k) &= \chi(\rho(g_1,k)) f(r\rho(g_1,k),\ k[g_1]), \\ L\varphi(k) &= \alpha(g_1,k)\varphi(k[g_1]); \\ \alpha(g_1,k) &= \chi(\rho(g_1,k)). \end{aligned} \tag{4.25}$$

The fact that $\|\nu\|_{\text{op}} = 1$ combined with (4.24) gives $\|L\| = 1$. This fact allows us to identify α with an old friend from representation theory. Explicitly, we see that α is the square root of the Radon–Nikodym derivative of dk under the mapping $k \to k[g_1]$, and (4.25) gives the induced representation on $R\backslash G$. These connections with representation theory will, however, not be relevant to us (see Gangoli and Varadarajan, 1980, §3.1 for more).

4.5 Proof of the B-Theorem 4.6

4.5.1 Theorem 4.6 in the principal bundle

We shall consider here $X = R \times K$ some principal bundle, as in the previous subsection, and assume that R is a C-group. On X we shall fix $T = L \otimes \{*\mu\}$, some left-invariant operator and in direct analogy with §2.4.1 we shall impose on L and the measures $\mu_{h,k}$ the following conditions:

(a) $L(h,dk) = L(h,k)\,dk$ where dk is the Haar measure of K and $L(h,k) > 0$ is continuous on $K \times K$;
(b) the measures $\mu_{h,k} \in \mathbb{P}(R)$ satisfy conditions (i), (ii) and (iii) of §2.4.1 on the group R uniformly in $h,k \in K$.

Remark If χ is a multiplicative character on R and T is as above and satisfies (a) and (b), then the conjugated operator $\chi T \chi^{-1}$ that we considered in the previous subsection also satisfies (a) and (b). The reason is that if we conjugate convolution by μ in R by multiplication by χ then we obtain convolution by $\chi \cdot \mu$. The property that $\mu \in \mathbb{P}(R)$ is, however, lost when we pass to $\chi \cdot \mu$

From what has been said, and Theorem 2.3, we can easily deduce a version of the B-theorem (Theorem 4.6) for left-invariant operators that satisfy the above conditions. More precisely, let us assume that $X = R \times K$ and T are as above. Let us fix $F(x) = f(r)\varphi(k)$, $H(x) = g(r)\psi(k)$, $x = (r,k)$, where $0 \leqslant f$, g, φ, $\psi \in C_0^\infty$ are fixed. We shall then use (4.12) together with Theorem 2.3 and the uniformity in condition (b) imposed on T. We deduce that

$$\langle T^n F, H \rangle \leqslant C \|L\|_{2\to 2}^n \exp(-cn^{1/3}) \leqslant C \|T\|_{2\to 2}^n \exp(-cn^{1/3}); \quad n \geqslant 1, \quad (4.26)$$

where the constant C also depends on F and H and the last inequality follows from (4.17). The scalar product $\langle\ ,\ \rangle$ is taken with respect to dx but, given that the supports of F, H are compact, any other measure locally equivalent to dx could have been used.

We shall exploit this estimate in the next subsection.

4.5.2 Proof of a weaker form of Theorem 4.6 for the special groups of §4.3.1

We shall consider $G = RK$ a group as in §4.3.1 and identify it with the principal bundle $X = R \times K$. We shall consider $\nu \in \mathbb{P}(G)$ some measure as in Theorem 4.6 and consider the left-invariant operators T, T_χ as defined in (4.24).

If we make the additional assumption that T_χ satisfies conditions (a) and (b) of §4.5.1 and use (4.26) we deduce that

$$\begin{aligned} |\langle F * \nu^{*n}, H \rangle| = |\langle T_\chi^n \chi F, \chi^{-1} H \rangle| &\leqslant C \|L\|_{2\to 2}^n \, \exp(-cn^{1/3}) \\ &\leqslant C \|\nu\|_{\mathrm{op}}^n \, \exp(-cn^{1/3}), \end{aligned} \quad (4.27)$$

for any $F, H \in C_0^\infty(G)$.

A simple use of the Harnack principle of §2.5.1 applied to ν finally gives the proof of the estimate (4.3) of Theorem 4.6 for our special class of groups. The problem that is left is that assumptions (a) and (b) on T_χ or T that correspond to ν are not satisfied in general. To address that issue we shall use the following.

Lemma 4.8 *Let $\nu \in \mathbb{P}(G)$ be as in Theorem 4.6. Then there exists $m \geqslant 1$ such that T^m the left-invariant operator that corresponds to $\nu^{*m} = \nu * \cdots * \nu$ satisfies conditions* (a) *and* (b) *of* §*4.5.1.*

Once this has been proved, estimate (4.3) of Theorem 4.6 follows for the measure ν^{*m}. The Harnack principle of §4.5.1 allows us to deduce the general result for ν. To see this let $p = mq + r, r = 0, 1, \ldots, m-1$; then by the Harnack principle of §2.5.1 (see (2.13), (2.16)) we can estimate $\nu^{*p}(e)$ by $\nu^{*m(q+s)}(e)$ for some $s \geqslant 1$ that depends on the already chosen m but is independent of q or r. But since this measure is $\nu^{*m} * \cdots * \nu^{*m}$ and our result holds for ν^{*m} our general result follows because $\|\nu^{*m}\|_{\text{op}} \leqslant \|\nu\|_{\text{op}}^m$.

I shall finish this section with the proof of the lemma. The strengthening of Theorem 4.6 with $\|\nu\|_{\text{sp}}$, on the other hand, needs new ideas and will be deferred until the next chapter.

From this special class of groups $G = RK$ we shall be able to deduce the general theorem for all groups in §4.6 below. That section presupposes familiarity with the structure theory of Lie groups. On the other hand, the reader could content themselves with just §4.6.1 where a proof for simply connected groups will at last be found.

4.5.3 Proof of Lemma 4.8

By the conditions on the measure ν, for any $A > 0$ we can find m such that $d\nu^{*m}(g) = \Phi(g)\,dg$ with Φ continuous and compactly supported. Furthermore, $\Phi(g) > 0$ for all $|g| < A$, that is, all $g \in G$ in the A-ball of G defined in §1.1. We shall use the representation (4.8), $T^m = L \otimes \{*\mu\}$, and then $d\mu_{h,k}(r) = f_{h,k}(r)\,dr$, with $r \in R, h, k \in K$. From this, if A is large enough, condition (a) and conditions (i) and (ii) of (b) in §4.5.1 easily follow. Furthermore, $f_{h,k}(r) \neq 0$ as long as $\Phi(krh^{-1}) \neq 0$. The reason is that for fixed $k \in K$ and δ_k, the δ-mass at k (considered as a measure on G), $\Phi(k^{-1}g)$, is the density of $\delta_k * \nu^{*m}(g)$. But we have

$$\big| |krh|_G - |r|_R \big| \leqslant C; \quad h, k \in K,\ r \in R, \tag{4.28}$$

by §2.14.2 because $K \subset G$ is compact, and this means that $f_{h,k}(r) \neq 0$ for $|r|_R < A'$ for some large A'. This gives the proof of property (iii) of §2.4.1 in condition (b) for T^m and completes the proof. This simple argument avoids the use of an explicit formula for $f_{h,k}$ in terms of Φ. This formula will, however, be given in §4.10 below.

The same proof works for $T_\chi^m = \chi T^m \chi^{-1}$ (but one might like to observe that χ cannot in general be identified with a character of G and therefore T_χ cannot be identified with a convolution operator on G).

4.6 Structure Theorems and the Reduction to the Special Class of Groups

The Levi decomposition and the principal bundle

Let $\mathfrak{g}$ be a Lie algebra; the classical Levi decomposition says that $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ where $\mathfrak{q}$ is the radical and $\mathfrak{s}$ is a semisimple algebra (see Varadarajan, 1974, §3.14; Jacobson, 1962, §III.9; and Appendix A). As we said in §4.1, we can also decompose $\mathfrak{s} = \mathfrak{s}_c \oplus \mathfrak{s}_n$ the direct sum of its compact and non-compact components. It is also possible to combine $\mathfrak{q}$ and $\mathfrak{s}_c$ and write

$$\mathfrak{g} = \mathfrak{q}_a \leftthreetimes \mathfrak{s}_n, \quad \mathfrak{q}_a = \mathfrak{q} \leftthreetimes \mathfrak{s}_c.$$

We use the index 'a' because we shall call $\mathfrak{q}_a$ the amenable radical of $\mathfrak{g}$. By definition $\mathfrak{q}_a$ is an amenable algebra (in the sense of §§2.2.2, 3.1.6; i.e. $\mathfrak{q}_a/\mathfrak{q}$ is of compact type). Clearly $\mathfrak{q}_a$ is an ideal in $\mathfrak{g}$.

Exercise (Not essential for us.) Prove that $\mathfrak{q}_a$ is the 'largest' ideal of $\mathfrak{g}$ that is amenable. As such it is uniquely determined (note that this is not the case for $\mathfrak{s}$), hence the terminology '*the* amenable radical'. For the proof use §3.1.6.

What counts for us, both for the classical Levi decomposition and the above generalisation, is that it induces a similar semidirect product decomposition on the simply connected Lie group G that corresponds to $\mathfrak{g}$:

$$G \cong Q_{\text{sol}} \leftthreetimes S \cong Q_a \leftthreetimes S_n,$$

where $Q_{\text{sol}}, S, Q_a, S_n$ are simply connected groups that correspond to $\mathfrak{q}, \mathfrak{s}, \mathfrak{q}_a, \mathfrak{s}_n$, respectively. To avoid pedantic qualifications, here and in what follows, we allow S_n to be $\{e\}$ (although it is hard to defend the point of view that $\{0\}$ is a semisimple algebra of non-compact type!).

Now let G be some connected Lie group that is not assumed to be simply connected; then Q_{sol}, the subgroup that corresponds to $\mathfrak{q} \lhd \mathfrak{g}$, is called the radical of G and is always a closed subgroup of G (even when G is not simply connected; see Varadarajan, 1974, §3.18.13). Since $S_c \subset G$ the subgroup that corresponds to $\mathfrak{s}_c$ is compact, the group $Q_{\text{sol}}S_c$ is always a closed connected normal amenable subgroup. We shall call this the amenable radical of G.

A question of notation In the literature (e.g. Varadarajan, 1974) Q_{sol} is more often than not denoted by Q. Here we prefer to use the letter Q for the amenable radical: so Q corresponds to $\mathfrak{q}_{\text{sol}} \leftthreetimes \mathfrak{s}_c$, where to stress the point, here $\mathfrak{q}_{\text{sol}}$ denotes the radical $\mathfrak{q}$ and Q_{sol} corresponds to $\mathfrak{q}_{\text{sol}}$. We shall further drop the index n and denote by S the analytic subgroup that corresponds to $\mathfrak{s}_n$.

Simply connected groups With this convention we are back to the notation that we used in §4.3.1 and for a simply connected group G we have $G = Q \leftthreetimes S$ where Q is amenable and S semisimple of non-compact type.

As explained in §4.3.1 the principal bundle emerges at once from this and from the Iwasawa decomposition of $S = NAK$, provided that K is compact. This happens when S has finite centre. In that case we have $G = RK$ where $R = QNA \cong Q \leftthreetimes (NA)$ and by §3.1.6 is an amenable group. It follows that G can be identified with the principal bundle $X = R \times K$. What should be noted is that when the group is simply connected this new notion of the amenable radical is not of great use because we could have done all the above using the actual radical Q_{sol} of the group. We have then $G = Q_{\text{sol}} \leftthreetimes S$ where now $S = S_n \oplus S_c$ is a simply connected semisimple group and is the direct product of its compact and its non-compact factor. If we proceed that way we obtain $G = Q_{\text{sol}} \leftthreetimes (NAK \oplus S_c)$ with the Iwasawa decomposition of S_n. And under the additional assumption that the centre of S is finite (this clearly implies that the centre of S_n is finite) we also have that K is compact (Helgason, 1978, Chapter VI). We can then identify G with $X = R \times L$ where R is the *soluble* group $Q_{\text{sol}}NA$ and L can be identified with the compact subgroup $K \oplus S_c \subset G$. This approach will be used in the next subsection to prove the B-theorem (Theorem 4.6) for all simply connected groups.

Some abuse of terminology In the next chapter we shall be dealing mostly with simply connected groups and systematic use of the above decomposition will be made. It will be convenient to introduce the following abuse of terminology: we shall agree to define a decomposition $S = NAK$ and call it the Iwasawa decomposition of S. This will be done for all semisimple groups that are direct sums $S = S_n \oplus S_c$ where S_n is of non-compact type and S_c is compact. If $S_n = NAK_n$ is the genuine Iwasawa decomposition then we set $K = K_n S_c$ and $S = NAK$.

A variant of the Levi decomposition In §4.6.2 we shall prove the B-theorem (Theorem 4.6) for general connected groups (simply connected or not) and for this we shall need to make use of the amenable radical. An alternative approach will be outlined in §4.6.3. We shall also come back to the problem again in §5.8* where yet another different approach will be explained.

The main tool in §4.6.2 below is the following decomposition that is valid for any connected and not necessarily simply connected group G. In the classical situation this is called the Levi decomposition of the group G (see Varadarajan, 1974, §3.18.13). This we can adapt here as follows.

Let $Q \lhd G$ denote the closed (connected) normal amenable radical and let $S \subset G$ be the analytic subgroup that is generated by the subalgebra $\mathfrak{s}_n \subset \mathfrak{g}$. Let $\widetilde{S}$ be the simply connected Lie group that corresponds to $\mathfrak{s}_n$ and $j\colon \widetilde{S} \to G$ the corresponding mapping. Let $S_L = S/(\ker j)$ and $j\colon S_L \to G$ be the induced injection. This means that j is one-to-one and this operation amounts to assigning on the subgroup $S \subset G$ the finer topology of a Lie group structure, hence the index L. The next step is to consider the semidirect product and the projection $\pi\colon Q \leftthreetimes S_L = \widetilde{G} \to G$ that is induced. The group $\widetilde{G}$ is not necessarily simply connected but its Lie algebra is $\mathfrak{q} \leftthreetimes \mathfrak{s}_n = \mathfrak{g}$ and π is a covering map. To clarify, G acts on Q by inner automorphisms on Q and so therefore does S_L by the injection j. Once $\widetilde{G}$ has been constructed, by the definition of a semidirect product, the identification of $Q \subset G$ and the injection j induce the mapping π.

A simple-minded, but also more direct, way of saying all the above is as follows. The group G can be written as the product of the two subgroups $G = QS$ and the product mapping $Q \times S \ni (q,s) \to qs \in G$ is bijective as long as $|q|_Q < \varepsilon$, $|s|_S < \varepsilon$ for ε small enough. Here $|q|_Q \approx |q|_G$ denotes the distance from the identity in Q or equivalently in G (because Q is closed). On the other hand, $|s|_S$ denotes the distance from the identity in S for its intrinsic Lie group structure, which here and in §4.6.2 below we shall denote by S_L (and not for the distance and topology induced by the inclusion $S \subset G$).

One consequence of the above local bijection that we shall need is that $Q \cap S \subset S_L$ is a discrete (closed) normal (and therefore also central) subgroup. To see this let $n \in Q \cap S$ be such that $|n|_S < \varepsilon$ small enough. Then $|n|_G$, and therefore also $|n|_Q$, is also small and since $e \cdot n = n \cdot e = n$ the above bijection implies that $n = e$ (the identity).

(Alternatively, and perhaps more directly, $A = Q \cap S$ is a closed subgroup of S_L, and A_0, the component of the identity, is an amenable normal subgroup and therefore it reduces to the identity because the Lie algebra of S is $\mathfrak{s}_n$.)

4.6.1 The use of Schur's lemma

For the reader's convenience I shall treat one difficulty at a time and consider first a connected algebraically-B group $G = Q \leftthreetimes S$ that is a semidirect product of an amenable group Q with a semisimple group S of non-compact type. All these groups are connected but we do not assume that $Z(S)$ the discrete finitely generated centre of S is finite. First, we shall prove the following.

Lemma 4.9 (i) *The centre of a connected semisimple Lie group is a finitely generated Abelian group.*

(ii) *Let G be some connected Lie group and let $S \subset G$, some analytic semisimple subgroup. Let us denote by $Z(G)$ and $Z(S)$ the centres of G and S respectively. Then there exists a subgroup $Z_1 \subset Z(S) \cap Z(G)$ that is central in G and such that the index $[Z(S)\colon Z_1] < +\infty$ is finite.*

(iii) *Alternatively, and more generally, any discrete central closed subgroup of a connected Lie group is finitely generated*; *see Hochschild (1965, Chapter XVI).*

The B-theorem (Theorem 4.6) now follows for the above group G once this lemma has been proved. To see this we quotient by Z_1 so that

$$\pi\colon G \to G_1 = G/Z_1 = Q \leftthreetimes (S/Z_1), \tag{4.29}$$

where S/Z_1 has finite centre. Let $\mu \in \mathbb{P}(G)$ satisfy properties (i), (ii) and (iii) of §2.4.1 in G. Then the image of μ by π has the same properties and $\|\check{\pi}(\mu)\|_{\text{op}} = \|\mu\|_{\text{op}}$ by §3.1.6. Furthermore, Theorem 4.6 holds on the group G_1 by the previous section. If we apply Harnack and the reduction of §2.5.2, we conclude that Theorem 4.6 also holds for the group G.

Exercise 4.10 Use the structure theorems of §4.6 and the above argument to show that we already have a proof of Theorem 4.6 for all simply connected groups.

Proof of Lemma 4.9 (i) This part is not essential for us but it is given for completeness. We have the Iwasawa decomposition $S = NAK$ where $NA \simeq \mathbb{R}^d$ topologically and $Z(S) \subset K$. Since K covers the compact group $K/Z(S) = \widetilde{K}$, we have that $Z(S)$ is a quotient of the fundamental group $\pi_1(\widetilde{K})$ (see Helgason, 1978, §VI.1).

(ii) By Weyl's fundamental theorem (Varadarajan, 1974, §3.13.1) on representations of semisimple groups, $[\mathrm{Ad}(s) \in \mathrm{GL}(\mathfrak{q});\ s \in S] = H$ is a completely reducible group of linear transformations on $\mathfrak{q}$. Thus H can be identified with $H \subset \mathrm{GL}_{\mathbb{C}}(\mathfrak{q}^c)$ and it becomes a completely reducible group of linear transformations on $\mathfrak{q}^c = \mathfrak{q} \otimes \mathbb{C}$ (Onischik and Vinberg, 1988, §5.2, problem 25, p. 247; Varadarajan, 1974, just below Theorem 3.19). It follows that $\mathfrak{q}^c = V_1 \oplus \cdots \oplus V_n$ is a direct sum of irreducible complex subspaces on each of which Schur's lemma applies (see Weil, 1953, IV §16; Warner, 1971, §1.8). It follows that for $z \in Z(S)$ the restriction of $\mathrm{Ad}\,z$ on each $V_i = W$ is $\alpha(z)I$ for $0 \neq \alpha(z) \in \mathbb{C}$ with

I the identity on W. Furthermore, $(\alpha(z))^d = \mathrm{Det}_W \operatorname{Ad} z$ for $d = \dim W$. On the other hand, $\mathrm{Det}_W \operatorname{Ad} s$, for $s \in S$, gives a homomorphism $S \to \mathbb{C}\backslash\{0\}$, and thus it is $\equiv 1$. The upshot is that $\alpha(z)$ is a dth root of unity. The lemma follows. □

In §4.6.3 we shall also use the following additional fact on a connected semisimple group S. Let $S_n, S_c \subset S$ be the analytic subgroups that correspond to the compact and non-compact components of the decomposition $\mathfrak{s} = \mathfrak{s}_n \oplus \mathfrak{s}_c$ of the Lie algebra. By the definition, S_c is compact but the subgroup $S_n \subset S$ is also closed. To see this we use our previous notation for $(S_n)_L$ and consider the projection $\pi\colon S^* = (S_n)_L \times S_c \to S$. Then π is a homomorphism and $S = S^*/(\ker\pi)$, and clearly $(S_n)_L \cap (\ker\pi) = \{e\}$. But since $\ker\pi \subset Z\big((S_n)_L\big) \times Z(S_c) = Z(S^*)$, with the notation of Lemma 4.9, and since $Z(S_c)$ is finite, it follows that $\ker\pi$ is finite. From this our conclusion follows at once.

4.6.2 A reduction

Let G be some connected Lie group. Let $Q \lhd G$ be its closed amenable radical and let $S \subset G$ be some analytic subgroup that is semisimple of non-compact type with $G = QS$, the variant of the Levi decomposition as in §4.6; here S is not in general a closed subgroup.

The argument that follows is typical of what one finds in the structure theory of Lie groups but it is important to distinguish the steps and align them in the correct order.

(a) As we saw in §4.6.1, $Q \cap S$ is a discrete central subgroup of S_L because it is a normal subgroup. By §4.6 we can then find $Z \subset Q \cap S$ such that the index $[Q \cap S \colon Z] < +\infty$ is finite and Z is central in G. Let us denote by $H = \overline{Q \cap S} \subset G$ the closure in G; then clearly the index $[H\colon \overline{Z}] < +\infty$. Indeed, H is the finite union of the closure of the cosets of Z in $Q \cap S$.
(b) We have $B = Q \cap (S \cdot \overline{Z}) \subset H$, for when $b = s \cdot z$, for $s \in S, z \in \overline{Z}$, and some $b \in B$, since $z \in H \subset Q$ we must have $s \in Q$ and therefore also $s \in Q \cap S \subset H$.
(c) We consider the quotient and the canonical projections

$$\pi\colon G \longrightarrow \frac{G}{\overline{Z}} = \overline{G}; \quad \pi = Q \longrightarrow \frac{Q}{\overline{Z}} = \overline{Q}; \quad \overline{S} = \pi(S).$$

Here $\overline{S}$ is an analytic subgroup that is semisimple of non-compact type. Furthermore, $Q \subset G$ is an amenable closed normal subgroup and we have $\overline{G} = \overline{Q} \cdot \overline{S}$, which is just the variant of the Levi decomposition of the group $\overline{G}$ that we gave in §4.6.
(d) By (b) it follows that $\overline{Q} \cap \overline{S} \subset \pi(H)$ which is a finite group because of (a).

The above (a)–(d) tell us that with the new group $\overline{G}$ we are exactly where we started for $G = QS$ but with the additional condition that $\overline{Q} \cap \overline{S}$ is finite. We also have the following reduction.

The reduction To prove Theorem 4.6 for the group G it suffices to prove it for the group $\overline{G}$. The fact that $\overline{Z}$ is Abelian, together with §3.1.6, implies that for any $\nu \in \mathbb{P}(G)$ we have $\|\check{\pi}(\nu)\|_{\text{op}} = \|\nu\|_{\text{op}}$ (and the same thing for $\| \ \|_{\text{sp}}$). We then apply the reduction of §2.5.2, and this proves the assertion.

Proof for $\overline{G}$ We can consider, as before, $\overline{S}_{\text{L}}$ the analytic subgroup $\overline{S}$ assigned with its Lie group structure, and form the semidirect product and the natural projection:

$$\widetilde{G} = \overline{Q} \leftthreetimes \overline{S}_{\text{L}} \longrightarrow \overline{G}. \tag{4.30}$$

The finiteness of $\overline{Q} \cap \overline{S}$ implies that the kernel of the projection is finite. The validity of Theorem 4.6 was proved in §4.6.1 for the group $\widetilde{G}$. By the Harnack principle of (2.17) and §3.1.6 applied to the projection (4.30) we obtain therefore the validity of Theorem 4.6 for $\overline{G}$.

4.6.3 An alternative approach to general groups

This alternative approach is interesting because it illustrates the use of principal bundles that are not trivial.

Let G be again a real Lie group and let $G/Q = S$, where Q is the closed radical and S is semisimple connected. Here Q denotes the genuine radical, that is, the one that was denoted by Q_{sol} in §4.6. Let $G = Q\Sigma$, where Σ is an analytic subgroup locally isomorphic with S (Varadarajan, 1974, §3.18.13). Let $Z \subset S$ be the centre. The aim of the next few lines is to show that for the proof of Theorem 4.6 we are allowed to make the *additional assumption* that the centre of S is finite.

Let Z be that centre and suppose that it is infinite. Since Z is discrete and finitely generated we can lift the free (i.e. infinite-order) generators $z_1, \dots, z_p \in Z$ to $\tilde{z}_1, \dots, \tilde{z}_p \in Z(\Sigma)$ the centre of Σ (i.e. $\pi(\tilde{z}_i) = z_i$, $\pi\colon\ G \to S$ being the canonical projections) and then the group $\widetilde{Z} = Gp(\tilde{z}_1 \dots, \tilde{z}_p)$ is a discrete (closed) subgroup of G. By Lemma 4.9 there exists $\widetilde{Z}_1 \subset \widetilde{Z}$ that is of finite index in $\widetilde{Z}$ and central in G. Furthermore, the projection $p\colon\ G \to \widetilde{G} = G/\widetilde{Z}_1$ is bijective on Q and takes Q onto the closed radical $\widetilde{Q}$ of $\widetilde{G}$. Denote $\widetilde{G}/\widetilde{Q} = \widetilde{S}$ and by our construction we see that now the centre of $\widetilde{S}$ is finite. If we use the covering map p and the reduction of §2.5.1 we see that, as asserted, the *additional assumption* can be made in the proofs of the B-theorem (Theorem 4.6).

The alternative approach to Theorem 4.6 is to prove it first for these *special groups of new type*, which are the groups G such that the semisimple group $G/Q = S$ has finite centre. But, unlike the conditions imposed in §§4.3.1 and 4.6.1, we do not necessarily have $G = Q \leftthreetimes S$. From what we have seen, once this has been done we have Theorem 4.6 in full generality.

Let G be such a special group and let NAK be the Iwasawa decomposition of S_n, the non-compact component of S. As we saw in §4.6.1, S_n is a closed subgroup of S. The subgroup $R = \pi^{-1}(NA)$ is a closed soluble subgroup that is a C- (resp. NC-) group if and only if G is algebraically-B (resp. NB). We cannot a priori write $G = RK$ as before but we have the natural principal fibre bundle $q\colon G \to R\backslash G = K$. Here K is the compact left-homogeneous space $\{Rg; g \in G\}$ and R acts on the left on G. The difference from the situation that we treated in §4.3 is that the bundle is in general non-trivial and the only way a priori to trivialise that bundle is to use the following.

Borel cross section We can find a Borel mapping $p\colon K \to G$ such that $p(K)$ is relatively compact and $q \circ p =$ identity on K. Using this we can again identify G with $X = R \times p(K) \simeq R \times K$ but this identification is not a priori a diffeomorphism. With the use of this Borel cross section, the formalism of left R-invariant operators and their coordinate representation as explained in §4.4 goes through with only minor changes to this non-trivial principal bundle. We shall not give the details but it would provide an interesting exercise for the reader who is prepared to write them down for themselves.

Remark Another application of the Borel section is that for G, some locally compact group, and $R \lhd G$, some normal closed amenable subgroup, it allows us to identify $L^2(G; d^r_G)$ with $L^2(R \times G/R; d^r_R \otimes d^r_{G/R})$ for the respective right Haar measures (see Bourbaki, 1963). Using this and an easy variant of the argument of §4.4.4 we see that for $\mu \in \mathbb{P}(G)$ and the projected measure $\check{\pi}(\mu) \in \mathbb{P}(G/R)$ we have $\|\mu\|_{\mathrm{op}} = \|\check{\pi}(\mu)\|_{\mathrm{op}}$ as asserted in §3.1.6.

Part 4.2: The Heat Diffusion Kernel and Gaussian Measures

4.7 Gaussian Left-Invariant Operators

All the definitions and notation of Chapter 2 concerning Gaussian measures will be preserved.

Let $X = R \times K$ be a principal bundle and let $T = L \otimes \{*\mu_{h,k}\}$ be a left-invariant operator as in §4.4. We shall use again the reference measure $dx = d^r r \otimes dk$ defined in §4.4.3. We shall replace here conditions (a), (b) of §4.5.1

and say that T is *Gaussian* if $L = L(h,k)\,dk$, where $L(\cdot,\cdot)$ is continuous and strictly positive, and if the measures $\mu_{h,k}$ are Gaussian on R as in §2.12.2 with constants that are uniform in $h,k \in K$. The composition of two Gaussian left-invariant operators as above is also Gaussian. To see this we verify first that the convolution of two Gaussian measures in a group is Gaussian. The verification of this uses the exponential estimate of the volume growth $\gamma(r)$ of §2.14.3 and will be left as an exercise for the reader.

In (4.17) we used the uniform compact support condition (4.16) and the amenability of R to deduce that

$$\|T_1 \circ \cdots \circ T_n\|_{2\to 2} = \|L_1 \circ \cdots \circ L_n\|_{2\to 2}; \quad n \geqslant 1, \tag{4.31}$$

for the $L^2(X;dx)$ and $L^2(K;dk)$ operator norms. This easily extends to products of operators that are uniformly Gaussian (i.e. uniform Gaussian constants).

In fact, for the proof of (4.17) we used the existence of $0 \leqslant f_m, g_m \in C_0^\infty(R)$ that satisfy (4.16). The argument extends to the present Gaussian situation because all that is used is the fact that for all fixed $n \geqslant 1$ and $\varepsilon > 0$ there exists $C \subset R$ a compact subset such that the measures $\pi = \mu^{(1)}_{k_1,k_2} * \cdots * \mu^{(n)}_{k_n,k_{n+1}}$ of (4.18) satisfy $\pi(R\backslash C) \leqslant \varepsilon$. (As already pointed out, the convolution product of Gaussian measures is Gaussian.) This condition is clearly satisfied for our operators $T_j = L_j \otimes \{*\mu^{(j)}\}$. Notice finally that the operator $\chi T \chi^{-1}$ defined in (4.13) remains Gaussian if the original operator is (see §4.9).

From this the analogue of (4.26) extends verbatim for Gaussian left-invariant operators T on $X = R \times K$ when R is a C-group. We can assert that there exists $c > 0$ such that if F, H are positive compactly supported functions as in (4.26) we have

$$\langle T^n F, H\rangle \leqslant C\|T\|^n_{2\to 2}\exp(-cn^{1/3}); \quad n \geqslant 1, \tag{4.32}$$

where C also depends on F, H.

4.8 The Gaussian B-Theorem

Theorem 4.11 (Gaussian B-theorem) *Let G be some connected* B*-Lie group and $\nu \in Gs(G)$. Then there exist $c > 0$ such that for every $P \subset G$ compact subset we have*

$$\nu^{*n}(P) = O(\|\nu\|^n_{\mathrm{op}}\exp(-cn^{1/3})). \tag{4.33}$$

The heat diffusion kernel and the spectral gap The typical example where the theorem applies is $\nu = \mu_t = \phi_t(g)\,d^r g$ in (4.33) for the heat diffusion kernel of $T_t = e^{-t\triangle}$ in §2.12. If we combine this with the Harnack estimate of (2.54) we obtain the upper estimate in the classification (B) of §1.3.2 for the heat diffusion kernel: we can assert that there exist constants $C, c > 0$ such that

$$\phi_t(e) \leqslant C\|T_t\|_{2\to2}\exp(-ct^{1/3}); \quad t \geqslant 1 \tag{4.34}$$

for the $L^2(G; d^r g)$ operator norm.

Here $\triangle$ is self-adjoint and therefore

$$\|T_t\|_{2\to2} = e^{-\lambda t}, \tag{4.35}$$

where

$$\lambda = \inf[(\triangle f, f);\ f \in C_0^\infty;\ \|f\|_2 = 1] \tag{4.36}$$

for the scalar product in $L^2(d^r g)$. Since $\lambda = \inf[\xi \in \mathrm{sp}\triangle]$ it is called the spectral gap. If, for $\triangle = -\sum X_j^2$ we denote $|\nabla f|^2 = \sum |X_j f|^2$, with $f \in C_0^\infty$, we also have $(\triangle f, f) = |\nabla f|^2$ because the X_j are antisymmetric operators: $(Xf, g) = -(f, Xg)$.

The above follows from elementary spectral theory and is close to the point of view of Varopoulos et al. (1992, Chapter 9).

4.9 Proof of the Gaussian B-Theorem

The way to deduce Theorem 4.11 for groups from the estimate (4.32) for left-invariant operators is identical to what was done for compactly supported measures in §4.5.2 and in §4.6 above.

We shall first consider a group G that can be written as a product $G = RK$ as in §4.3.1 with K compact; this group will be identified with the principal bundle $R \times K$. The new property that must be verified is contained in the following.

Lemma 4.12 *Let $G, X = R \times K$ be as above and let $d\nu = \psi(g)\, d^r g \in \mathrm{Gs}(G)$ be a Gaussian measure on G. Assume that ψ is continuous and denote by $Tf = f * \nu$, $f \geqslant 0$ the corresponding left-invariant operator on X. Then T is a Gaussian left-invariant operator in the sense of §4.7.*

Let us first assume the lemma and conclude the proof of the theorem from it.

Exactly as in §4.4.6 with X and G as in the lemma, we consider χ and T_χ of (4.23), (4.24). Then T_χ is also a Gaussian left-invariant operator because χ is a positive character and so satisfies $\chi^{\pm1}(r) = O(\exp|r|_R)$, $r \in R$. The analogue of (4.27) therefore holds. This implies (4.33) for the special groups of the lemma.

To pass from this to general groups nothing changes in the structure theorems and the reduction of §4.6. We shall use the general fact from §2.14.2

that the projection $\check{\pi}(\mu)$ of a Gaussian measure on G by $\pi\colon G \to G/H$ is also Gaussian. We then proceed as in (2.81), without using Harnack. We shall not rewrite the details of the proof.

4.10 An Explicit Formula and the Proof of Lemma 4.12

The notation is as in Lemma 4.12. We shall systematically denote $g = rk$, $g_i = r_i k_i$, $r, r_i \in R$, $k, k_i \in K$, and prove the following explicit formula:

$$Tf(rk) = \iint f(rr_1^{-1}k_1)\psi(k_1^{-1}r_1k)\,d^r r_1\,dk_1; \quad f \geqslant 0. \tag{4.37}$$

The lemma follows from this at once because (4.37) gives the explicit representation $T = L(k,k_1)\,dk_1 \otimes \{*\mu_{k,k_1}\}$. We have

$$\begin{aligned} L(k,k_1) &= \int_R \psi(k_1^{-1}r_1,k)\,d^r r_1, \\ \mu_{k,k_1} &= f_{k,k_1}(r_1)\,d^r r_1 \in \mathbb{P}(R), \\ f_{k,k_1}(r_1) &= (L(k,k_1))^{-1}\psi(k_1^{-1}r_1k). \end{aligned} \tag{4.38}$$

As in (4.28), the lemma is therefore a consequence of §2.14.2 which gives

$$\Big| |r_1|_R - |k_1^{-1}r_1k|_G \Big| \leqslant C. \tag{4.39}$$

Proof of (4.37) Because of (4.7) and §4.4.6 with $\nu = \varphi\,dg$, we can write

$$Tf(rk) = \iint_{R\times K} f(rkk_1^{-1}r_1^{-1})\varphi(r_1k_1)\,dr_1\,dk_1; \quad f \geqslant 0. \tag{4.40}$$

Sublemma *Let us fix $k \in K$ and consider the one-to-one correspondence $(r_1,k_1) \leftrightarrow (r_2,k_2)$ given by $g = r_1k_1 = k_2^{-1}r_2k$. Write the Jacobian as $dg = dr_1\,dk_1 = J(r_2,k_2;k)\,dr_2\,dk_2$. Then we have*

$$J(r,k;k') = J(r) = \frac{m_R(r)}{m_G(r)}, \tag{4.41}$$

where m_G (resp. m_R) is the modular function of G (resp. R).

We shall presently prove the sublemma. Once this is done we shall have from (4.40),

$$Tf(rk) = \iint f(rr_2^{-1}k_2)\varphi(k_2^{-1}r_2k)J(r_2)\,dr_2\,dk_2. \tag{4.42}$$

If we then insert $\psi(k_2^{-1}rk) = \varphi(k_2^{-1}rk)m_G(r)$ and $d^r r_2 = m_R(r_2)\,dr_2$ in (4.42) we obtain (4.37) – recall that φ and ψ are the respective densities of ν with respect to dg and $d^r g$.

Proof of the sublemma Since $dg^{-1} = d^r g$ on G, formula (4.22) combined with the involution $g \mapsto g^{-1}$ implies that for the parametrisation $g = kr, k \in K$, $r \in R$, we have $d^r g = d^r r\, dk$ and this, since $dk_2^{-1} = dk_2$, implies that $d^r g = dk_2\, d^r r_2$ for $g = k_2^{-1} r_2 k$ for k fixed. This means that

$$d^r g = dk_2\, d^r r_2 = dk_2\, m_R(r_2)\, dr_2 = m_G(g)\, dg = m_G(g) J(r_2)\, dr_2\, dk_2.$$

The formula follows since $m_G(g) = m_G(r_2)$. □

5

NB-Groups

Part 5.1

Overview of Part 5.1

In this chapter we shall complete the main classification of §1.3 by proving the lower NB-estimate. This is a difficult chapter to read because it builds on the previous three chapters and it would be futile even to start reading without a reasonably good understanding of that previous material. Also, the ad hoc tools that are explained in the appendix, although elementary, make it considerably more involved than the appendices of Chapters 2 and 3. Apart from this, no fundamentally new ideas are needed and the difficulty is mostly technical.

In §5.1 we recall known results on positive eigenfunctions of positive operators. In finite dimensions these correspond to matrices with positive entries and many readers may be familiar with the special nature of their 'positive eigenvectors'. Once this is recalled in §5.3 (especially §5.3.5), we proceed to make what is often referred to as the use of the 'ground state' (i.e. the above-mentioned positive eigenfunction) to 'reduce the spectral gap'. In short, this means that we explain a method for eliminating the exponential factor $e^{-\lambda t}$ in the lower NB-estimate $\gtrsim e^{-\lambda t} n^{-\alpha}$. We are thus led to proving just a polynomial lower estimate $\gtrsim n^{-\alpha}$ for a modified new kernel.

This new polynomial lower estimate is then proved in §§5.5–5.6 by adapting the methods of Chapter 3. The hardest point to adapt there has nothing to do with Lie groups and consists in generalising appropriately the gambler's ruin estimate in the conical domains of Chapter 3. This is done in the appendix and in the context of the principal bundles of Chapter 4.

With this lower estimate on principal bundles, in conformity with the scheme that we described in the overview of Chapters 2 and 3, we proceed in §5.7 first to find the special class of Lie groups on which we can apply this previous

NB-estimate of principal bundles, and then we have to reduce the problem to this special class. The special class of groups is the same as in the previous chapter and is relatively easy to describe since it consists of 'essentially' (but not quite!) all simply connected groups. The situation is exactly as in §4.6.

A new idea based on positive-definite functions (see Naimark, 1959) is, however, needed in §5.7 to make this reduction. This new idea is elaborated further in §5.8* where we obtain a unified approach of the reductions of both §4.6 and §5.7. Certainly the reader can skip §5.8* in a first reading but even §5.7 is not a section that should be given priority in the reading of the chapter.

5.1 The First Eigenfunction and the Sharp B-Theorem 4.6

In this section we shall introduce the key definition of the 'first' eigenfunction (sometimes called the 'ground state') and illustrate that notion by completing the proof of Theorem 4.6 from the previous chapter and proving estimate (4.3) with $\|\nu\|_{\text{sp}}$ rather than the weaker $\|\nu\|_{\text{op}}$.

Here and throughout we shall denote by $X = R \times K$ a principal bundle and we shall use throughout the reference measure $dx = d^r r \otimes dk$ as in §4.4.3. Then for the left-invariant operator $T = L \otimes \{*\mu\}$ that we have been considering, we have $\|T\|_{2\to 2} = \|L\|_{2\to 2}$ for the corresponding L^2 operator norms as in (4.17) where, to fix ideas, we assume that $L = L(h,k)\,dk$ with $L \in L^2(K \times K)$. Under that assumption L is a compact operator on $L^2(K)$ and therefore by the general theory of compact operators (Dunford and Schwartz, 1958, VII.3, VII.4) we can find $\varphi_0 \in L^2(K)$, an eigenfunction (complex valued) with a maximal in modulus eigenvalue:

$$L\varphi_0 = \alpha \|L\|_{\text{sp}} \varphi_0; \quad \alpha \in \mathbb{C},\ |\alpha| = 1, \tag{5.1}$$

where we assume that $\lim \|L^n\|^{1/n} = \|L\|_{\text{sp}} \neq 0$. The following additional assumptions will be imposed:

$$\begin{aligned} &L^\infty \ni L > 0,\ L^\infty \ni \varphi_0 > 0, \\ &L\varphi_0 = e^{-\lambda}\varphi_0,\ e^{-\lambda} = \|L\|_{\text{sp}}. \end{aligned} \tag{5.2}$$

In (5.2) the first assumption on L clearly implies that $\varphi_0 \in L^\infty$ and this together with the positivity of L will allow us to choose φ_0 as in (5.2). Note that when L is continuous then $L1 > \varepsilon_0 > 0$ and therefore $\|L\|_{\text{sp}} > 0$. Both in (5.2) and throughout, the qualification 'holds a.e.' has been dropped. The proof of (5.2) will be given in the next subsection.

5.1.1 Proof of (5.2)

The fact can be put in the more general context of positive operators on Banach lattices (Schaefer, 1974, V.5, especially Theorem 5.2 and its corollary) and in the jargon of that area $\sigma = \alpha\|L\|_{\mathrm{sp}}$ in (5.1) is an element of the *peripheral point spectrum*.

The proof that follows is done in a number of steps that are either obvious or are easy exercises for the reader. We shall normalise throughout and assume that $\|L\|_{\mathrm{sp}} = 1$ and $\|\varphi_0\|_2 = 1$.

(a) By the positivity of the operator we have

$$L|\varphi_0| \geqslant |L\varphi_0| = |\varphi_0|. \tag{5.3}$$

This means that when $L(h,k) = L(k,h)$ is a symmetric operator we have $\|\varphi_0\|_2 \geqslant \|L|\varphi_0|\,\|_2 \geqslant \|\varphi_0\|_2$ and thus $\|L|\varphi_0|\,\|_2 = \|\varphi_0\|_2$ because the operator norm satisfies $\|L\|_{\mathrm{op}} = \|L\|_{\mathrm{sp}} = 1$. Using (5.3) we are done because this implies that $L|\varphi_0| = |\varphi_0|$.

If L is not assumed symmetric the proof is slightly more involved. For simplicity, and because this is the only case that occurs in our applications, we shall assume that L is continuous. This clearly implies that φ_0 is also continuous.

(b) There exists $k_0 \in K$ such that

$$L|\varphi_0|(k_0) = |\varphi_0(k_0)|; \quad L|\varphi_0|(k_0) = |L\varphi_0(k_0)|. \tag{5.4}$$

If not, there exists $\varepsilon > 0$ such that $L|\varphi_0| \geqslant (1+\varepsilon)|\varphi_0|$. Iterating this we obtain $L^n|\varphi_0| \geqslant (1+\varepsilon)^n|\varphi_0|$, $n \geqslant 1$ and this contradicts the condition $\|L\|_{\mathrm{sp}} = 1$. The second assertion in (5.4) follows from the first and (5.3).

(c) We can replace φ_0 by $e^{i\theta}\varphi_0$, with $\theta \in \mathbb{R}$, and thus assume that in (5.4) we have $\varphi_0(k_0) \geqslant 0$. This combined with the positivity of L and the *second* assertion in (5.4) implies that $\varphi_0(k) \geqslant 0$, $k \in K$. Using the strict positivity of L we conclude that $\varphi_0(k) > 0$, with $k \in K$, and we are done.

The same argument works under the more general condition that $L \in L^\infty(K \times K)$. All we have to do is to insert the qualification 'almost everywhere' in the appropriate places. The only point that needs additional thinking is condition (b) where the first statement in (5.4) has to be replaced by the weaker statement that *the measure of the set* $[k \in K;\ L|\varphi_0|(k) \leqslant (1+\varepsilon)|\varphi_0(k)|]$ *is positive for all* $\varepsilon > 0$.

This can be proved by the same argument. Elaborating step (c) we then see that this, as before, implies that there exists $\theta \in \mathbb{R}$ and $\varphi^+ \in L^\infty(K)$ such that $\varphi^+(k) \geqslant 0$, with $k \in K$, and $\varphi_0 = e^{i\theta}\varphi^+$ and we are done.

5.2 Proof of the Sharp B-Theorem 4.6

Here we now assume as in §4.5 that R is a C-group and that $T = L \otimes \{*\mu\}$ satisfies conditions (a) and (b) of §4.5.1. This implies that (5.2) holds for the operator L. In estimate (4.26) we shall set $F = f(r)\varphi_0(k)$ and from (4.11) we obtain the following improvement:

$$|\langle T^n F, H\rangle| \leqslant C\|L\|_{\mathrm{sp}}^n \exp(-cn^{1/3}) = C\|T\|_{\mathrm{sp}}^n \exp(-cn^{1/3}); \quad n \geqslant 1. \quad (5.5)$$

Now let $G = RK$ be a B-group that can be identified with $X = R \times K$ as in §4.3.1 and $\nu \in \mathbb{P}(G)$ be as in Theorem 4.6. We shall consider the convolution operator $Tf = f * \nu$. As we explained in Lemma 4.8, by replacing if necessary ν with a large convolution power ν^{*m} we may assume that T satisfies conditions (a) and (b) of Lemma 4.8.

Now let $\chi^2 = m_G m_R^{-1}$ and $T_\chi = \chi T \chi^{-1}$ be as in (4.23) and (4.24) so that $\|\nu\|_{\mathrm{sp}} = \|T_\chi\|_{\mathrm{sp}}$ for the spectral radius of §3.1.1. estimate (5.5) implies therefore that

$$|\langle F * \nu^{*n}, H\rangle| = |\langle T_\chi^n(\chi F),\ \chi^{-1}H\rangle| \leqslant C\|\nu\|_{\mathrm{sp}}^n \exp(-cn^{1/3}); \quad n \geqslant 1. \quad (5.6)$$

This combined with the Harnack estimate of §2.5.2 gives the sharp form of (4.3) with $\|\nu\|_{\mathrm{sp}}$ for Theorem 4.6. Here the change from ν to ν^{*m} that we made in order to guarantee conditions (a) and (b) clearly makes no difference because $\|\nu^{*m}\|_{\mathrm{sp}} = \|\nu\|_{\mathrm{sp}}^m$ and because we can use the Harnack estimate of §2.5 again as we did in §4.5.2.

Remark The same argument, but even simpler, can be used to prove the statement $\mu^{*n}(e) \leqslant C\|\mu\|_{\mathrm{sp}}^n$ of §3.1.3 and this is valid without the B-condition on the group. Whether a similar estimate holds for all locally compact groups (connected or not – e.g. discrete groups) I do not know.

5.3 Symmetric Markovian Operators

5.3.1 Definition and the criterion

In this section we shall consider left-invariant operators $\widehat{T} = L \otimes \{*\mu\}$ on $X = R \times K$ that satisfy the following conditions:

(I) $\widehat{T}$ is Markovian, that is, $\widehat{T}1 = 1$ and equivalently $L1 = 1$;

(II) $\widehat{T}$ satisfies conditions (a) and (b) of §4.5.1: we recall that this means that the kernel of L is continuous and strictly positive and that the probability measures $\mu_{h,k}$ satisfy conditions (i), (ii) and (iii) of §2.4.1 uniformly in h, k;

(III) $\widehat{T}$ is symmetric with respect to $\hat{dx} = d^r r \otimes \hat{dk}$ for some smooth non-vanishing measure $\hat{dk}$ on K (as in §4.4.3), that is, $\widehat{T}^* = \widehat{T}$ for the adjoint of (4.15);

(IV) $\widehat{T}^n$ admits a continuous kernel $\hat{\phi}_n(x_1,x_2), x_1,x_2 \in X$, such that

$$\widehat{T}^n f(x_1) = \int \hat{\phi}_n(x_1,x_2) f(x_2) \hat{d}x_2; \quad f \in C_0^\infty(X). \tag{5.7}$$

Let $e = e_X = (e_R, e_K) \in X$ be some reference point that can be arbitrary, though to fix ideas we could have taken the identities of R and K. We then have

$$\begin{aligned} &\hat{\phi}_n(x_1,x_2) = \hat{\phi}_n(x_2,x_1); \\ &\hat{\phi}_{2n}(e,e) = \int_X \hat{\phi}_n(e,x)\hat{\phi}_n(x,e)\,\hat{dx} = \int \hat{\phi}_n^2(e,x)\,\hat{dx}; \quad n \geqslant 1. \end{aligned} \tag{5.8}$$

Denote by $x(n) \in X$, $n = 0,1,2,\ldots$ the Markov process generated by the semigroup $\widehat{T}^n$. Then, with the same notation as in §3.3.2,

$$\mathbf{E}_x f(x(n)) = \widehat{T}^n f(x), \quad \mathbb{P}_x[x(n) \in E] = \int_E \hat{\phi}_n(x,y)\,\hat{dy};$$
$$f \in C_0^\infty,\ E \subset X,\ x \in X,\ n \geqslant 1.$$

Combining this with Hölder and (5.8) we deduce, as in (3.8),

$$\hat{\phi}_{2n}(e,e,) \geqslant \int_E \hat{\phi}_n^2(e,x)\,\hat{dx} \geqslant \mathbb{P}_e^2[x(n) \in E]\ |E|^{-1}; \quad E \subset X,\ n \geqslant 1, \tag{5.9}$$

where $|E|$ denotes the $\hat{dx}$ measure of E. From this we obtain for $\widehat{T}$ the following generalisation of the criterion of §3.3.3.

Criterion 5.1 *Let us assume that there exist positive constants C_1, C_2, c and a sequence of subsets $E_n \subset X$, $n \geqslant 1$ such that*

(i) *the $\hat{dx}$ measures satisfy $|E_n| \leqslant C_1 n^c$, $n \geqslant 1$;*
(ii) $\mathbb{P}_e[x(n) \in E_n] \geqslant C_2 n^{-c}$, $n \geqslant 1$.

Then there exist positive constants C_0, c_0 such that

$$\hat{\phi}_{2n}(e,e) \geqslant C_0 n^{-c_0}; \quad n \geqslant 1. \tag{5.10}$$

As in §3.3, as soon as we have Harnack at our disposal (see the next section), we also have the same estimate for $\hat{\phi}_n$.

5.3.2 A modification of the criterion

As we previously indicated in the criterion of §3.3.3 we can use the scale $\exp(\pm cn^{1/3})$ in the criterion and replace conditions (i), (ii) by

(i′) the $\hat{dx}$ measure of $|E_n| \leqslant C_1 \exp(cn^{1/3})$, $n \geqslant 1$;
(ii′) $\mathbb{P}_e[x(n) \in E_n] \geqslant C_2 \exp(-cn^{1/3})$, $n \geqslant 1$.

The conclusion then is

$$\hat{\phi}_{2n}(e,e) \geqslant C_0 \exp(-c_0 n^{1/3}). \tag{5.11}$$

The proof is identical.

5.3.3 The construction of the Markovian operators

We shall start from an arbitrary left-invariant operator on $X = R \times X$ that satisfies (a) and (b) of §4.5.1 and for which, as in condition (IV) of §5.3.1, there exists a continuous kernel $\phi_n(x_1, x_2)$,

$$T^n f(x_1) = \int \phi_n(x_1, x_2) f(x_2)\, \hat{dx}_2; \quad \hat{dx} = d^r r \otimes \hat{dk}; \tag{5.12}$$

we postpone specifying the measure $\hat{dk}$ till later.We shall also fix χ some positive multiplicative character on R and define $T_\chi = \chi T \chi^{-1}$ and *assume* that T_χ is symmetric with respect to $dx = d^r r \otimes dk$ for the Haar measure dk on K. We shall then use the notation of §4.4 and write $T_\chi = L_\chi \otimes \{*\mu\}$ (this is a normal representation; see §§4.4.2, 4.4.3) with L_χ continuous and strictly positive. Note also that in the representation (5.12) we have the freedom to change $\hat{dx}_2$ to another measure $a(x_2)\, \hat{dx}_2$ and then ϕ_n changes to $\phi_n(x_1, x_2) a^{-1}(x_2)$.

5.3.4 Example

Let $G = RK$ be a group that is identified with $X = R \times K$ as in §4.3.1 and let $Tf = f * \nu$ be the left-invariant operator on X (see §4.4.5) that corresponds to $\nu = \psi\, d^r g \in \mathbf{P}(G)$ some compactly supported symmetric measure where ψ is continuous as in Theorem 4.7. Then T is self-adjoint with respect to right Haar measure $d^r g$ and by replacing ν by ν^{*m} for a large enough m as in Lemma 4.8 we may assume that T satisfies conditions (I)–(IV) of §5.3.1. In that case we shall take $\chi^2 = m_G m_R^{-1}$ as in (4.23) and then

$$dx = d^r r \otimes dk = m_R\, d^\ell r \otimes dk = m_R\, d^\ell g = \chi^{-2} d^r g. \tag{5.13}$$

For the representation of T as in (5.12) see §4.10. Then $T_\chi = \chi T \chi^{-1}$ is symmetric as in §5.3.3 because for this $\chi = m_G m_R^{-1}$ the mapping $f \to \chi f$ induces an isometry $L^2(G; d^r g) \to L^2(X; dx)$ and we use the symmetry of the right convolution operator by ν on G. By the same argument we also see that $\|\nu\|_{\text{op}} = \|T_\chi\|_{2\to 2}$ for the operator norm of T_χ on $L^2(X; dx)$ and the operator norm of the convolution operator induced by ν as in §3.1.

5.3.5 General operators

We go back to a general T and χ that satisfy the conditions of §5.3.3. We shall write $T_\chi = L_\chi \otimes \{*\mu\}$, where L_χ is continuous, and define some eigenfunction $\varphi_0 \in L^2$ and $\lambda \in \mathbb{R}$ as in (5.1) such that

$$L_\chi \varphi_0 = e^{-\lambda}\varphi_0, \quad e^{-\lambda} = \|T_\chi\|_{2\to 2}, \tag{5.14}$$

for the operator norm in $L^2(X;dx)$ ((4.17) and the symmetry of the operator is used here). By our conditions on T it follows that φ_0 is continuous and as in (5.2) it can be taken to be positive. If we apply this to the example of §5.3.4, we get $\|T_\chi\|_{2\to 2} = e^{-\lambda} = \|\nu\|_{\text{op}}$.

We shall now use the eigenfunction φ_0 of (5.14) to define the following operator and measure:

$$\widehat{T} = e^{\lambda}\varphi_0^{-1}\chi T\chi^{-1}\varphi_0 = e^{\lambda}\varphi_0^{-1}T_\chi\varphi_0; \quad \hat{d}k = \varphi_0^2\, dk; \tag{5.15}$$

then $\widehat{T}$ and $\hat{d}k$ satisfy conditions (I)–(IV) and $\widehat{T}$ is symmetric with respect to $\hat{d}x = d^r r \otimes \hat{d}k$ because T_χ is symmetric with respect to dx. The kernel of $\widehat{T}^n$ with respect to $\hat{d}x$ is

$$\hat{\phi}_n(x_1,x_2) = e^{\lambda n}\alpha^{-1}(x_1)\phi_n(x_1,x_2)\alpha(x_2); \quad \alpha = \chi^{-1}\varphi_0, \tag{5.16}$$

where ϕ_n is the kernel of T^n with respect to $\hat{d}x$ as in (5.12). We therefore have

$$\phi_n(e,e) = \hat{\phi}_n(e,e)e^{-\lambda n}. \tag{5.17}$$

From this and (5.10) we conclude that if the conditions of Criterion 5.1 are satisfied for $\widehat{T}$ then for even integers there exist constants C, c such that

$$\phi_n(e,e) \geqslant Ce^{-\lambda n}n^{-c}; \quad n \geqslant 1. \tag{5.18}$$

Remark In the jargon of the area we can describe the above by saying that we have used the ground state φ_0 to reduce the spectral gap.

5.3.6 The group case

In the example of §5.3.4, $Tf = f * \nu$ with $d\nu^{*n} = \psi_n d^r g$, we certainly have $\psi_n(e_G) \sim \phi_n(e,e)$ for the neutral element e_G of G. This follows from the two different ways of representing $T^n f(g) = \int_G f(gh^{-1})\psi_n(h)\, d^r h$ and $T^n f(x) = \int_X \phi_n(x,y) f(y)\, \hat{d}y$; see (3.4). This, if the reference points are the identities e_K and e_R of the groups R and K, gives $\psi_n(e_G) = \phi_n(e,e)\varphi_0^2(e_K)$. Under the above conditions, (5.18) implies therefore that there exist constants $C, c > 0$ such that

$$\psi_n(e_G) \geqslant Ce^{-\lambda n}n^{-c}; \quad n \geqslant 1, \tag{5.19}$$

which is exactly the estimate needed in Theorem 4.7, and similarly under conditions of the modified criterion of §5.3.2. We can conclude instead

$$\psi_n(e_G) \geqslant C\exp(-\lambda n - cn^{1/3}); \quad n \geqslant 1. \tag{5.20}$$

Once more the Harnack principle of §2.5 has to be used to obtain (5.19)–(5.20) for all $n \geqslant 1$ and not just for even integers.

5.4 Theorem 4.7 for Principal Bundles and the Harnack Estimate

For the rest of this chapter we shall fix a principal bundle $X = R \times K$ as in §4.3.1 and assume that R is a soluble simply connected NC-group, We shall fix $\widehat{T}$ some Markovian left-invariant operator that satisfies conditions (I)–(IV) of §5.3.1 and is symmetric with respect to $\hat{dx}$ the measure of §5.3.1. The kernel of $\widehat{T}^n$ with respect to $\hat{dx}$ will be denoted $\hat{\phi}_n$ as in §5.3.1. The key step for the rest of the chapter is to prove the following generalisation of Theorem 4.7.

Proposition 5.2 *Let X, $\widehat{T}$, $\hat{\phi}_n$ be as above, and let $e \in X$ be the reference point of* (5.10). *There exist then $C, c > 0$ such that*

$$\hat{\phi}_n(e,e) \geqslant Cn^{-c}; \quad n \geqslant 1. \tag{5.21}$$

This will follow from Criterion 5.1 but in the constructions that we shall make next we shall also need to adapt and use the Harnack estimate of §2.5.1 to this more general setting.

Notice that by the left invariance of the operator, the kernels are also left invariant in the sense that $\hat{\phi}_n(rx_1, rx_2) = m_R(r^{-1})\hat{\phi}_n(x_1,x_2)$ for the left action of R of X and the modular function m_R of R.

Note for the reader For the global understanding of this chapter it is important to note that §§5.1–5.3 and §§5.4–5.6 deal with very different aspects. The first lot, as we have already said, deals with the 'ground state'. The second deals with left-invariant operators on a principal bundle. Note finally that the actual convolution operator will not really be examined until §5.7.

Harnack principle

To achieve this we define on X the distance $d(x_1,x_2)$ that is the sum of the left-invariant distances on R with some distance on K. By our conditions it is then clear that $\hat{\phi}_1(x_1,x_2) = 0$ if $d(x_1,x_2) > c_0$ for some $c_0 > 0$. Also by conditions

(I)–(IV) of §5.3.1 and the definition of the composition of (4.11), for all C there exists m such that $\hat{\phi}_m(x_1,x_2) > 0$ for all $x_1,x_2 \in X$ such that $d(x_1,x_2) < C$. From these it easily follows that for all $a > 0$ there exist m and C such that

$$\hat{\phi}_1(x_1,x_2) \leqslant C\hat{\phi}_m(y_1,y_2); \quad x_1,x_2,y_1,y_2 \in X,\ d(x_1,y_1),\ d(x_2,y_2) < a. \tag{5.22}$$

As in §2.5 we shall also need a more cumbersome version of this that says that for all $a > 0$ there exists $m_0 > 1$ such that for all $m > m_0$ there exists $C > 0$ for which (5.22) holds. The only issue in (5.22) is the uniformity of the constants because (5.22) is clear if we fix, say, x_1 and allow the constants to depend on it. The reason is that then the range of the x_2, y_1, y_2 that is relevant in (5.22) is compact. But from this, if we use the left invariance of the kernel as explained above, the uniformity follows. This is similar to what we did in a special context in §2.5.

If we combine these facts with the semigroup composition property

$$\hat{\phi}_{p+q}(x_1,x_2) = \hat{\phi}_p \circ \hat{\phi}_q(x_1,x_2) = \int_X \hat{\phi}_p(x_1,y)\hat{\phi}_q(y,x_2)\,\hat{dy}; \quad x_1,x_2 \in X, \tag{5.23}$$

we deduce that for all $a > 0$ there exist m and C such that

$$\hat{\phi}_n(x_1,x_2) = \hat{\phi}_{n-1} \circ \hat{\phi}_1(x_1,x_2) \leqslant C\hat{\phi}_{n-1} \circ \hat{\phi}_m(x_1,y_2) = C\hat{\phi}_{n+m-1}(x_1,y_2);$$
$$x_1,x_2,y_1,y_2 \in X,\ d(x_1,y_1) \leqslant a,\ d(x_2,y_2) \leqslant a,\ n \geqslant 2.$$

With this we pass from (x_1,x_2) to (x_1,y_2). We can repeat the same arguments with the analogous factorisations $\hat{\phi}_n = \hat{\phi}_1 \circ \hat{\phi}_{n-1}$ and this allows us to pass from (x_1,y_2) to (y_1,y_2) as needed. Needless to stress, the important point here is that in this estimate the m and the C are independent of n. This is the Harnack estimate for $\hat{\phi}_n$ that we shall use.

5.5 The Euclidean Bundle

5.5.1 The definition

In this section we shall denote by $V = \mathbb{R}^d$ the Euclidean space and for $R = V$, $X = V \times K$, the corresponding principal bundle. As before, $\hat{dk}$ is some smooth non-vanishing measure and the reference measure is $\hat{dx} = dv \otimes \hat{dk}$, where dv is Lebesgue measure on V. We shall then denote by

$$\widehat{T}f(x) = \int_X M(x,y)f(y)\,\hat{dy}; \quad f \in C_0^\infty \tag{5.24}$$

some Markovian symmetric operator where, with the notation $x = (\gamma,h)$, $y = (\lambda,k)$, $\gamma,\lambda \in V$, $h,k \in K$, we impose $M(x,y) = M(\gamma-\lambda;h,k)$, $M(\gamma;h,k) =$

$M(-\gamma;k,h)$ where $M(\gamma;h,k)$ is continuous and compactly supported. The conditions

$$\begin{gathered}\int M(\gamma;h,k)\,d\gamma\hat{d}h = \int M(\gamma;k,h)\,d\gamma\hat{d}h = 1;\\ k\in K,\ M\geqslant 0,\ M(0;h,k)>0;\quad h,k\in K\end{gathered}\tag{5.25}$$

are also imposed and then we have in (5.24) a left-invariant operator that satisfies all conditions (I)–(IV) of §5.3.1. Denoting as before $x(n)\in X$, $n=1,2,\dots$, the Markov process generated by $\widehat{T}$, we then have

$$\mathbb{P}[x(n+1)\in E\ //\ x(n)=(\gamma,k)] = \int_E M(\lambda-\gamma;h,k)\,d\lambda\,\hat{d}h. \tag{5.26}$$

This process is a generalisation of a random walk on V and when K is a single point it is the random walk given by the measure $M(\gamma)\,d\gamma$. What will be needed for us is to generalise the gambler's ruin estimates of §3.3.6 to this process.

Remark 5.3 As in Lemma 4.8, provided we are prepared to replace $\widehat{T}$ by $\widehat{T}^m$ for a large enough m, the condition $M(0;k,h)>0$, $h,k\in K$ is a consequence of the other conditions imposed on M. To see this we use the symmetry of the operator and (4.11) and $M^2(0;h,h)=\int M(\gamma;h,k)M(-\gamma;k,h)\,d\gamma\hat{d}k>0$; then we use the Harnack estimate of §5.4 to pass to different $h,k\in K$. This property, in Markov chain terminology, implies that the corresponding Markov chains are irreducible.

5.5.2 The conical domain and the gambler's ruin estimate

Let $\mathscr{L}=(L_1,\dots,L_p)\subset V^*$ be a finite set of linear functionals on V, and set

$$\Omega=\Omega_{\mathscr{L}}=[v;\ Lv<1,\ L\in\mathscr{L}]\subset V \tag{5.27}$$

the corresponding domain. We shall assume that $\mathscr{L}$ is such that Ω is an unbounded open conical domain that contains $0\in\Omega$. As in Chapter 3, when $\mathscr{L}=(0)$ we have $\Omega=V$. The fact that Ω is unbounded is equivalent to the NC-condition for $\mathscr{L}$ in §2.2.1.

We can then define the corresponding conical domain in $\Omega\times K\subset X$. The key estimate for us is the following generalisation.

5.5.3 Gambler's ruin estimate

Let $x(n)\in X$, $n=1,2,\dots$, be the process as in §5.5.1, and let $\mathscr{L}$ and Ω be as above. Then there exist $C_1,\alpha>0$ depending on $\widehat{T}$, $\mathscr{L}$ such that

$$\mathbb{P}_0[x(j)\in\Omega\times K;\ j=0,1,\dots,n]\geqslant C_1 n^{-\alpha}. \tag{5.28}$$

The notation in (5.28) is $\mathbf{P}_0[x(0) = e_X] = 1$ with e_X for some arbitrary but fixed reference point of X as in §5.3.1, that is, the probability of the process starting at e_X. In fact, α here depends only on the geometry of $\mathscr{L}$ and the geometric invariants of $\widehat{T}$. To compute α precisely is quite difficult. This is essential if we are to obtain the sharp form of NB in §1.3.1. But even the proof of (5.28) in the above weak form is non-trivial. This will be done in the appendix to this chapter. For the time being in the proofs that follow we shall take this polynomial gambler's ruin estimate for granted.

5.6 Proof of Proposition 5.2

5.6.1 Overview of the proof

This is essentially the dénouement of this chapter because as we shall see in the next section, the proof of Theorem 4.7 follows easily from this. The proof of the proposition for the bundle $X = R \times K$ will be done in the three following successive cases:

(i) $R = N \leftthreetimes V$ where N is nilpotent and $V \cong \mathbb{R}^d$ is Euclidean;
(ii) $R = N \leftthreetimes H$ where N and H are nilpotent;
(iii) we shall use (ii) to deduce the proposition for a general soluble group R.

All the groups N, H, R are simply connected and R is NC. As a consequence the real parts of the roots of the ad-action of the Lie algebra of V or H on the Lie algebra of N satisfy the NC-condition of §2.2.1.

Item (i) is of course a special case of (ii), but the proof for (ii) is essentially identical, though the notation is simpler in case (i) so we shall start with this.

5.6.2 The Euclidean case (i)

We shall denote by $x(j) \in X$ the Markov chain that is generated by the operator $\widehat{T} = L \otimes \{*\mu_{h,k}\}$ of §5.3.1. Let $\pi : X = R \times K \to Y = V \times K$ be the projection induced by $R \to R/N = V$. Then the process $y(j) = \pi(x(j)) \in Y$ is Markovian and is generated by the operator $\widehat{T}_V = L \otimes \{*\check{\pi}(\mu_{h,k})\}$ that is symmetric with respect to $dv \otimes \hat{d}k$ (cf. (4.15)). We shall introduce the following notation: $x(j) = r_j k_j$; $r_j = n_j v_j$; $y(j) = v_j k_j$; $r_j \in R$; $n_j \in N$, $v_j \in V$; $k_j \in K$. We shall define successively $\tilde{r}_j \in R$ by $r_j = \tilde{r}_1 \tilde{r}_2 \cdots \tilde{r}_j$, $j \geqslant 1$, and write $\tilde{r}_j = \tilde{n}_j \tilde{v}_j$, for $\tilde{n}_j \in N$, $\tilde{v}_j \in V$. Note that we have $v_j = \tilde{v}_1 \cdots \tilde{v}_j$ but that n_j is not in general $\tilde{n}_1 \cdots \tilde{n}_j$.

The process $r_j, j = 1,\ldots$ is not Markovian. But if we condition on the path $(k_1, k_2, \ldots)$ in K it becomes a time-inhomogeneous random walk with transition distributions μ_{k_{j+1},k_j}. Here we use formal probabilistic language but, naively, this statement says that we fix $k_1, k_2, \ldots$ and 'see what happens' to the R-coordinate of $x(j)$. From the conditions in Proposition 5.2 and the composition formula (4.11) it follows that with starting probability $x(0) = e$ (which is the reference point of (5.28)) we have almost surely for $\mathbb{P}_0$,

$$|\tilde{r}_j|_R, \quad |\tilde{n}_j|_N, \quad |\tilde{v}_j|_V \leqslant C; \quad j = 1, 2, \ldots, \tag{5.29}$$

for some $C > 0$. Alternatively, what (5.29) says is that the increments $r_{j-1}^{-1} r_j = \tilde{r}_j$ have conditional distribution μ_{k_{j+1},k_j} with uniformly bounded supports. We shall now apply the gambler's ruin estimate of §5.5.3 to the process $y(j) \in Y$. The fact that $\widehat{T}_V$ satisfies conditions (I)–(IV) of §5.4 on Y follows from the fact that these conditions are satisfied by T on X. This implies that the conditions of §5.5.1 and Remark 5.3 are satisfied for $\widehat{T}_V$ on the Euclidean bundle Y.

We define the event

$$\mathscr{E}_s = [Lv_j < 1;\, L \in \mathscr{L},\ 1 \leqslant j \leqslant s], \tag{5.30}$$

where $\mathscr{L}$ is the set of real roots of R. Since $V = R/N$ these roots can be identified to linear functionals on V. By applying (5.28) it follows therefore that there exist $c_1, \alpha > 0$ such that

$$\mathbb{P}_0(\mathscr{E}_s) \geqslant c_1 s^{-\alpha}; \quad s \geqslant 1. \tag{5.31}$$

We shall now choose an appropriately large $C > 0$ and define the corresponding sets in R:

$$B_s = \left[r = nv, n \in N,\ v \in V;\ |n|_N \leqslant Cs^C,\ |v|_V \leqslant Cs^C\right] \subset R. \tag{5.32}$$

We shall also use the representation of (2.21) and because of the definition of the $\tilde{v}_j$ we can write

$$r_j = \tilde{n}_1 \tilde{v}_1 \tilde{n}_2 \tilde{v}_2 \cdots \tilde{n}_j \tilde{v}_j = \tilde{n}_1 \tilde{n}_2^{v_1} \cdots \tilde{n}_j^{v_{j-1}} v_j. \tag{5.33}$$

If the constants C in (5.32) are large enough we can then use (5.33) and Lemma 3.4 as in the argument of §3.5.2 to deduce that

$$\mathscr{E}_s \subset [x(j) \in B_s \times K;\ j = 1, \ldots, s] = \widetilde{\mathscr{E}_s}; \quad s = 1, 2, \ldots, \tag{5.34}$$

where to estimate the last factor in (5.33) we use (5.29). This and (5.31) therefore imply that

$$\mathbb{P}_0(\widetilde{\mathscr{E}_s}) \geqslant Cs^{-c}; \quad s = 1, 2, \ldots. \tag{5.35}$$

This is the key step in the proof and it gives condition (ii) of Criterion 5.1. On the other hand, for the Haar measure on R, $d^r r = dn\,dv$, $r = nv \in R = N \leftthreetimes V$, we see that

$$|B_s \times K| \leqslant Cs^C; \quad s = 1,2,\ldots. \tag{5.36}$$

The final conclusion is that the conditions of Criterion 5.1 are satisfied for the sets $E_s = B_s \times K$. Condition (i) follows from (5.36) and condition (ii) from (5.35). This criterion therefore completes the proof of the proposition in the case $R = N \leftthreetimes V$.

5.6.3 The case $R = N \leftthreetimes H$

The proof for this more general case is almost identical to the previous one – it is just the notation that becomes more involved. We set $x(j) = r_j k_j$; $r_j = n_j h_j$; $r_j = \tilde{r}_1 \tilde{r}_2 \cdots \tilde{r}_j$; $y(j) = h_j k_j \in Y$; $\tilde{r}_j = \tilde{n}_j \tilde{h}_j$; $r_j, \tilde{r}_j \in R$; $n_j, \tilde{n}_j \in N$, $k_j \in K$, $h_j, \tilde{h}_j \in H, h_j = \tilde{h}_1 \cdots \tilde{h}_j$ and $Y = H \times K$. The quantity $y(j)$ is the projected Markov chain by the projection $X \to Y = H \times K$ that is induced by the canonical projection $R \to H$ and, as before, for the starting probability $x(0) = e_X$ we have almost surely

$$|\tilde{r}_j|_R, \ |\tilde{n}_j|_N, \ |\tilde{h}_j|_H \leqslant C; \quad j = 1,2,\ldots, \tag{5.37}$$

for some $C > 0$. Here as in (5.29) we use (4.11) and the conditions on $\widehat{T}$. Let $\mathscr{L} = (L_1, \ldots, L_p)$ be the real roots of R. These can be identified with linear functionals on $\mathfrak{h}/[\mathfrak{h},\mathfrak{h}]$ where $\mathfrak{h}$ is the Lie algebra of H. Since $V = H/[H,H]$ is Abelian they can also be identified with linear functionals on V. We shall also denote by $p : H \to V$ the canonical projection. If we denote by $p(y(j)) = v(j) = v_j k_j \in V \times K$, $v_j \in V$, $k_j \in K$, the projected chain by the induced projection p: $Y = H \times K \to V \times K = Z$ we obtain a symmetric Markov chain in the Euclidean bundle Z. Now we shall use the gambler's ruin estimate (5.28) for that chain in Z in the conical domain $\Omega = [v \in V;\ Lv < 1,\ L \in \mathscr{L}]$ as in §5.6.2. We define the event

$$\begin{aligned} \mathscr{E}_s &= [Lh_j < 1;\ L \in \mathscr{L},\ j = 1,\ldots,s] \\ &= [Lv_j < 1;\ L \in \mathscr{L},\ j = 1,\ldots,s]; \quad s \geqslant 1. \end{aligned} \tag{5.38}$$

We then obtain from (5.28),

$$\mathbb{P}_0(\mathscr{E}_s) \geqslant Cs^{-\alpha}; \quad s = 1,2,\ldots, \tag{5.39}$$

for appropriate constants $C, \alpha > 0$.

For an appropriately large $C > 0$ we shall define the following set in R:

$$B_s = [r = nh,\ n \in N,\ h \in H;\ |n|_N \leqslant Cs^C,\ |h|_H \leqslant Cs^C] \subset R. \tag{5.40}$$

We use again the representation of (2.21),

$$r_j = \tilde{n}_1\tilde{h}_1\, \tilde{n}_2\tilde{h}_2 \cdots \tilde{n}_j\tilde{h}_j = \tilde{n}_1\tilde{n}_2^{h_1} \cdots \tilde{n}_j^{h_{j-1}} h_j. \tag{5.41}$$

This together with (5.37) and Lemma 3.4 and the argument of §3.5.2 gives

$$\mathscr{E}_s \subset [x(j) \in B_s \times K;\ j = 1, \dots, s] = \widetilde{\mathscr{E}}_s. \tag{5.42}$$

This together with (5.39) implies that

$$\mathbb{P}_0(\widetilde{\mathscr{E}}_s) \geqslant cs^{-c}; \quad s = 1, 2, \dots \tag{5.43}$$

and gives condition (ii) of the criterion. To finish the proof, as in case (i), we observe that the $\hat{d}x = dn\,dh\,\hat{d}k$ measure satisfies $|B_s \times K| \leqslant Cs^c$, $s \geqslant 1$, and this gives condition (i) of the criterion. Criterion 5.1 applies again and we are done.

5.6.4 The general case (iii)

Let $X = R \times K$ be a general bundle and let $\widehat{T}$ be a Markovian symmetric operator as in the proposition. We shall then construct $\widetilde{R} = N \leftthreetimes H$ and $\pi : \widetilde{R} \to R$ a projection as (3.14), where N and H are nilpotent and therefore $\widetilde{R}$ is as in case (ii). We shall also abuse notation and continue to write $\pi : \widetilde{X} = \widetilde{R} \times K \to R \times K$ for the induced mapping. The reference measure on $\widetilde{X}$ is $\tilde{d}x = d^r\tilde{r} \otimes \hat{d}k$, $\tilde{r} \in \widetilde{R}$, $k \in K$. We shall prove the following lemma.

Lemma 5.4 *We can lift $\widehat{T}$ to a left-invariant operator $\widetilde{T}$ on $\widetilde{X}$ in the sense that $\widetilde{T}(f \circ \pi) = \widehat{T}(f) \circ \pi$, $f \in C_0^\infty(X)$. This can be done in such a way that $\widetilde{T}$ is symmetric with respect to $\tilde{d}x$ and satisfies conditions* (I)–(IV) *of §5.3.1.*

Let us first assume the lemma and complete the proof of the proposition. To see this let $\widetilde{\phi}_n$ be the kernel of $\widetilde{T}$ on $\tilde{X}$. Then for the corresponding reference point $\tilde{e} \in \widetilde{X}$ we have $\widetilde{\phi}_n(\tilde{e}, \tilde{e}) \geqslant cn^{-c}$, $n \geqslant 1$ (note $\pi(\tilde{e}) = e$) by §5.6.3. We can now use the same reduction as in §3.2.2 by using the Harnack principle of §5.4. The conclusion is the required estimate $\hat{\phi}_n(e, e) \geqslant cn^{-c}$ for $\widehat{T}$ on X. This completes the proof of the proposition.

Proof of Lemma 5.4 The proof is an elaboration of §3.2.2. Let $\widehat{T} = L \otimes \{*\mu\}$. We can lift each individual measure $\mu_{h,k}$ to some measure $\tilde{\mu}_{h,k} \in \mathbb{P}(\widetilde{R})$ so that $\check{\pi}(\tilde{\mu}_{h,k}) = \mu_{h,k}$. To guarantee that conditions (i)–(iii) of §2.4.1 are satisfied by each $\tilde{\mu}_{h,k}$ is clearly no problem. The operator $\widetilde{T} = L \otimes \{*\tilde{\mu}\}$ so obtained is Markovian because L is. To guarantee the symmetry we can always replace $\widetilde{T}$ by $\frac{1}{2}(\widetilde{T} + \widetilde{T}^*)$ which by (4.10) and (4.15) is also Markovian since $L^* = L$ is. The only issue is condition (IV) of the continuity of the kernel.

To address that question we shall use the existence of a smooth section, that is, a continuous mapping $\sigma : R \to \widetilde{R}$ such that $\pi \circ \sigma =$ identity on R. It is here that the assumption that R is simply connected is used because this implies that $\ker \pi$ is connected and then we use Varadarajan (1974, §3.18.2) – see the remark below. We can then take $\mu^{\sigma}_{h,k} = \check{\sigma}(\mu_{h,k}) \in \mathbb{P}(\widetilde{R})$, the image by σ. The continuous dependence in h, k of $\mu^{\sigma}_{h,k}$ is then guaranteed, but these measures are supported in $\sigma(R)$ and are in general singular. To mend that difficulty we write $M = \ker \pi \subset \widetilde{R}$ and fix $\theta \in \mathbb{P}(M)$, some smooth compactly supported measure. Now $M \times \sigma(R) \to M\sigma(R) = \widetilde{R}$ is a diffeomorphism and we can use this to define $\tilde{\mu}_{h,k} = \theta \otimes \mu^{\sigma}_{h,k} \in \mathbb{P}(\widetilde{R})$ which is now smooth and depends continuously on h, k. Now, by condition (IV) on $\widehat{T}$, the measures $d\mu_{h,k} = \varphi_{h,k}(r)\, dr$ are smooth and depend continuously on h, k. With this choice of $\tilde{\mu}$ therefore, the operator $\widetilde{T}$ satisfies condition (IV) and this completes the proof of the lemma. □

Remark The existence of a smooth section (Varadarajan, 1974, §3.18.2) is a difficult technical point and it is desirable to avoid using it. We can of course fall back on the alternative proof that we give in §5.11.2 below for Gaussian measures where no liftings or sections are needed. But for the construction of the above sections it should be noted that both $\tilde{R}$ and R are simply connected soluble groups and for such groups we can use the exponential coordinates of the second kind. (cf. Varadarajan, 1974, §3.18.11 and also §7.3.1 in Part II). The required sections can then be constructed by lifting the appropriate one-parameter subgroups. This type of construction will be systematically used in the geometric theory in Part II. All the details can be found in §8.4. But there the construction is difficult because we demand much more. For our needs here, the argument used in §11.3.3 is good enough.

5.7 Proof of Theorem 4.7

5.7.1 The reduction

Let G be some connected NB-group that can be written in the form $G = RK$, $R \cap K = \{e\}$ with R some simply connected NC-soluble closed subgroup and K compact. By Proposition 5.2 and the consideration of §5.3.6 it follows that Theorem 4.7 holds for this group G. For the proof, Lemma 4.8 and the Harnack principle are used once more, exactly as it was in §4.5.2.

On the other hand, by the reduction in §3.2.2, in proving Theorem 4.7 we may assume that the group G is simply connected and therefore $G = Q \leftthreetimes S$ for a simply connected soluble NC-group Q and S semisimple. What is new here with respect to Chapter 3 is the presence of $e^{-\lambda} = \|\mu\|_{\mathrm{op}}$ of (4.4), but in the reduction §3.2.2 this clearly causes no problem because of §3.1.6.

As we previously explained in §4.6.1 we can then find $Z \subset S$ some discrete central subgroup such that $G_1 = Q \leftthreetimes (S/Z)$ is of the required form RK. What remains to be done is therefore to give a proof of the following new type of reduction that has nothing to do with Harnack.

Remark 5.5 To elaborate on the choice of Z, we write $S = S_n \times S_c$ as a direct product of it non-compact and compactor factors, S_n and S_c respectively. Then $Z \subset S_n$ is taken as a subgroup that is central in G and of finite index in $Z(S_n)$ the centre in S_n. Then if $S_n/Z = NAK_n$ is the corresponding Iwasawa decomposition, we can take $R = QNA$ and $K = K_n S_c$.

Reduction *We shall show that we can deduce the validity of Theorem 4.7 for the group G from the validity of the theorem for the group $G_1 = G/Z$.*

The proof of this reduction is interesting because it relies on the new ideas that we shall explain in the next two subsections. To make the exposition self-contained we shall start by recalling some definitions from §2.14.1.

Let G be some compactly generated locally compact group, and $H \subset G$ be some compactly generated closed subgroup. For all $h \in H$ we can then define $|h|_H$, $|h|_G$ the distance from the neutral element in two different ways, that is, with h considered as an element of G or h as an element of H. And it is clear that $|h|_G \leqslant C|h|_H + C$ for appropriate constants. The proof of this is obvious (think of the case of discrete groups). We shall say that H is *not distorted* in G if the estimate the other way round also holds and we have $|h|_H \leqslant C|h|_G + C$. This is a rare phenomenon indeed. For more details and references on the above the reader should refer back to §2.14.

Let us now go back to the notation of the previous section. We have the following result.

Lemma 5.6 (Distortion lemma) *The subgroup Z in the reduction of* §*5.7.1 is not distorted in G.*

The proof of this lemma will be given in §5.7.4 below.

5.7.2 The use of positive-definite functions

Let G be some locally compact group, let $\phi(g) \geqslant 0$ be continuous compactly supported, and let $d\mu(g) = \phi(g)\, d^r g \in \mathbb{P}(G)$ $d\mu^{*n}(g) = \phi_n(g)\, d^r g$ as in §3.3.1. Then we have

$$\phi_{2n}(x) = \int \phi_n(xy^{-1})\phi_n(y)\, d^r g, \tag{5.44}$$

and if μ is symmetric we have $\phi(x^{-1}) = \phi(x)m(x)$ for the modular function. The function $\phi_{2n}(x)$ is not positive definite in general on G since we do not even have $\phi_{2n}(x) = \phi_{2n}(x^{-1})$ (Weil, 1953, §14; Naimark, 1959, §30). But if we restrict ϕ_{2n} to some central subgroup $Z \subset G$ we obtain a positive-definite function because by (5.44) for $z_1, \dots, z_n \in Z$ we have (we use $\phi_n(x^{-1}) = m(x)\phi_n(x)$, that the z_i are central and that $m(z_i) = 1$ (cf. §2.3.5) and, in the integral below, make the change of variable $g \to gz_i$)

$$\begin{aligned}\phi_{2n}(z_i z_j^{-1}) &= \int_G \phi_n(z_i z_j^{-1} g^{-1})\phi_n(g)\, d^r g = \int_G \phi_n(z_j^{-1} g^{-1})\phi_n(z_i g)\, d^r g \\ &= \int \phi_n(z_i g)\phi_n(z_j g) m(g)\, d^r g. \end{aligned} \tag{5.45}$$

Note We recall that the complex-valued function $f(g)$ is positive definite if and only if $\sum f(g_i g_j^{-1})\lambda_i \bar{\lambda}_j \geqslant 0$, $\lambda_j \in \mathbb{C}$. What we shall use is that then $f(e) \geqslant |f(g)|$ (see Naimark, 1959).

5.7.3 Proof of the reduction in §5.7.1

Let $d\mu = \phi(g)\, d^r g \in \mathbb{P}(G)$ be as in Theorem 4.7 and let $d\check{\mu} = \check{\phi}\, d^r g_1$ be its image on G_1 by the projection $\pi : G \to G_1$. We then have

$$\begin{aligned}\check{\phi}_n(x) &= \sum_{g \in \pi^{-1}(x)} \phi_n(g); \quad x \in G_1, \\ \check{\phi}_n(e) &= \sum_{g \in Z} \phi_n(g); \quad Z = \ker \pi. \end{aligned} \tag{5.46}$$

Since ϕ_{2n} is positive definite on Z we have $\phi_{2n}(e) \geqslant \phi_{2n}(z)$, $z \in Z$, and (5.46) implies that

$$\check{\phi}_{2n}(e) \leqslant \phi_{2n}(e) A_{2n}, \tag{5.47}$$

where A_n is the number of points of the set

$$[z \in Z;\ \phi_n(z) \neq 0] = Z \cap \operatorname{supp} \phi_n. \tag{5.48}$$

Since the support of μ is compact, $\operatorname{supp} \phi_n \subset [|g|_G \leqslant Cn]$ and by the distortion lemma (Lemma 5.6),

$$|\operatorname{supp} \phi_n \cap Z| \leqslant \big| z \in Z,\ |z|_Z \leqslant Cn \big|. \tag{5.49}$$

This clearly implies that $A_n \leqslant Cn^d$, where d is the rank of the Abelian group Z. The reduction of §5.7.1 clearly follows from this and (5.47) and where Harnack is used to deal with odd integers.

For the proof of Theorem 4.7 it remains therefore to prove Lemma 5.6.

5.7.4 Proof of Lemma 5.6

The lemma is in fact contained in §11.4 where the proof relies on a general principle that is used extensively in geometric theory (see Exercise 11.9). The proof that I give below is interesting because it relies on different ideas. However, the reader may want to skip it until they study the geometric theory of Chapter 11.

Let $G = RK = Q \rtimes S$ be some simply connected group, Q be the radical and $S = NAK$ be the Iwasawa decomposition of the semisimple group S (in the terminology of §4.6.1) and $R = QNA$. We can then identify $K = R\backslash G$ with the homogeneous space $(Rg : g \in G)$. Now $Z \subset K$ is some discrete subgroup that is central in G and K/Z is compact. It follows that Z, by right multiplication, induces a discrete cocompact action on $K = R\backslash G$. We can assign $R\backslash G$ with the distance $\dot{d}(k,h)$, $k,h \in K$, induced by the left-invariant group distance d_G of §1.1 on G. We have $\dot{d}(k,h) = \inf_r d_G(rk,h)$; the balls for that distance are the balls in G by the projection $G \to R\backslash G$ and they are relatively compact. The distance $\dot{d}$ is invariant by the Z-action.

We can identify $Z \subset R\backslash G$ and since clearly $\dot{d}(k,h) \leqslant d_G(k,h)$ the lemma will follow as soon as we can show that

$$|z|_Z \leqslant C\dot{d}(e,z) + C; \quad z \in Z. \tag{5.50}$$

We shall need the following result.

Sublemma 5.7 *For all $c > 0$ there exists $C > 0$ such that the number of points $|z \in Z;\ \dot{d}(e,z) < c| \leqslant C$.*

Proof Using the fact that the action of Z on K is discrete we see that there exists $\varepsilon > 0$ small enough so that $B_\varepsilon(z_1) \cap B_\varepsilon(z_2) = \emptyset, z_1, z_2 \in Z, z_1 \neq z_2$, for the balls $B_\varepsilon(k_0) = [k \in K;\ \dot{d}(k,k_0) < \varepsilon]$. Indeed, if that were not the case, using the fact that Z is central, we would be able to find sequences $z_n \in Z$, $r_n \in R$, both tending to infinity and such that $z_n r_n \to 0$ in G, which gives a contradiction. Now the Haar measure of K as a group is invariant by the Z-action and all the above balls have the same measure. The sublemma follows. □

Proof of (5.50) Let $\dot{d}(e,z) = r$ and $e = k_0, k_1, \ldots, k_{2r} = z$ some 'geodesic' in the sense that the 'size of the edge' is $\dot{d}(k_{j+1},k_j) \leqslant 1$, $k_j \in K$. Such a geodesic can be constructed because an analogous geodesic can be constructed in G for the distance d_G. By the cocompactness of the Z action on K it follows that we can assume that $k_j = z_j \in Z$ provided that we increase the size of the edge by $\dot{d}(z_{j+1},z_j) \leqslant c$ for some appropriate $c > 0$ (this is similar to what we did in §2.14.2). By the Z-action and the sublemma it follows that $z_{j+1}^{-1} z_j \in F$ some finite set of Z.

The conclusion is therefore that for a set of generators that contains F, the word distance of z in Z from e is $\leqslant 2r$. This completes the proof of (5.50), and Lemma 5.6 follows. □

Exercise A less ad hoc proof can be given using the ideas of Example 11.9 from Part II. Indeed, (5.50) is a consequence of the fact that both on the group K, with group distance, and on $R\backslash G$ the group Z acts cocompactly as a group of isometries. Deduce from this that the identification $K = R\backslash G$ is a coarse quasi-isometry in the sense of Chapter 11. The subtle point here is that on $R\backslash G$, the distance is not Riemannian. On the other hand, one can use 'geodesics' in the above sense for these two distances.

5.8* The Global Structure of Lie Groups

In this section we shall use a different, more sophisticated method and give an alternative proof to unify the two reductions of §§4.6 and 5.7.

The key fact that was used for the reductions of both Theorems 4.6 and 4.7 from a general Lie group to the special groups RK (with R soluble, K compact, $R \cap K = \{e\}$) was the following principle (cf. (4.29)):

If $G = Q \leftthreetimes S$ is the Levi decomposition of a simply connected group G and $S = NAK$ is the Iwasawa decomposition (using the generalised terminology of §4.6.1 again) of the semisimple group S, then there exists $Z \subset K$ a discrete central (in G) subgroup such that $G = G/Z \cong R \cdot (K/Z)$, $R = QAN$ is of the required form and K/Z is compact.

Unfortunately however, this general principle does not suffice to complete the reductions and a number of ad hoc additional considerations are also needed to complete the proofs. The aim of this section is to give a more global approach to the problem.

The maximal normal torus

We shall start from an easy observation that is in some sense dual to the previous principle. Let G be some connected Lie group then there exists $T \subset G$ a maximal normal torus in G. This means that $T \cong \mathbf{T}^d$ is a normal subgroup and is maximal under these conditions. Some of these facts have already been used in §2.6.1. But it is also clear, though not essential, that (i) T is unique and (ii) T is central. All these facts are well known (see §11.3 later on and §2.6.1).

Exercise Use the theorem that every soluble compact group is a torus (see Hochschild, 1965, §III.1.3) to prove all these facts.

If we use the reductions of (2.17) and §3.2.2 we also see that Theorems 4.6 and 4.7 are 'equivalent' for G and for G/T, and therefore it suffices to prove these theorems for groups that contain no normal torus. Then all we need to prove in order to close this circle of ideas is a generalisation of Lemma 5.6 and the following result.

Lemma 5.8 *Let G be some connected Lie group that contains no normal torus. Let $\pi : \widetilde{G} = Q \leftthreetimes S \to G$ be the simply connected cover of G together with its Levi decomposition and let $\widetilde{Z} = \ker \pi$. Then $\widetilde{Z}$ is not distorted in $\widetilde{G}$.*

This presents some independent interest but the main point is that it can be combined with the positive-definite function of §5.7.2 to give the following result.

Theorem (General reduction) *Theorems 4.6 and 4.7 are equivalent for G, for G/T and for $(\widetilde{G/T})$, the universal cover of G/T.*

From this reduction we conclude that for both these theorems we may assume that the group is simply connected. For such a group we use $G = Q \leftthreetimes S$, $Z \subset S$, as at the beginning of this section, to reduce the problem to $G = RK$. We obtain thus a much more unified procedure to make the reductions of Theorems 4.6 and 4.7. Indeed, the argument of §5.7.3 works 'both ways' and gives the reduction for the upper estimate as well. We shall skip the details.

Proof of Lemma 5.8 Given that this lemma is more of an accessory than essential for our theorems, we shall be brief.

Let $N \subset Q$ be the nilradical of $\widetilde{G}$. Then $\widetilde{Z} \cap N \subset Z(N)$ the centre of N and therefore $\widetilde{Z} \cap N = \{e\}$, for otherwise $N/(\widetilde{Z} \cap N)$ and therefore also G would contain some normal torus (see Varadarajan, 1974, §3.6.4).

If we quotient by N we obtain $\widetilde{G}/N = V \times S = G_V$ for $V = Q/N$ some Euclidean space, where the product is direct rather than only semidirect because of Varadarajan (1974, §3.8.3). Then $W = \theta(\widetilde{Z})$ the image of $\widetilde{Z}$ in $\theta : \widetilde{G} \to \widetilde{G}/N$ is closed because $\pi(N) \subset G$ is the *closed* nilradical of G (this is a general fact on the analytic nilradical; see Varadarajan, 1974, §§3.18.3–3.18.3). Indeed, to say that W is not closed is to say that we can find $n_j \in N$, $z_j \in \widetilde{Z}$, such that $n_j, z_j \to \infty$ and yet $n_j z_j^{-1} \to 0$. This fact is symmetric in N and $\widetilde{Z}$ and is equivalent to the fact that $\pi(N)$ is *not* closed in G (we have already used this argument in §2.11.3). It suffices therefore to prove that the central subgroup $W \subset G_V$ is not distorted in G_V because θ is one-to-one on $\widetilde{Z}$. This, however, is easy. Indeed, $W \subset G_V$ is a central subgroup and therefore $W \subset V \times Z(S) = V_Z$ for the discrete centre $Z(S)$ of S, which by Lemma 4.9 is an Abelian group of finite rank. From this it follows that every closed subgroup of V_Z is not distorted in

V_Z. On the other hand, from Lemma 5.6, V_Z is not distorted in G_V. And we are done. □

Remark A discrete central subgroup in a non-Abelian nilpotent group as in Varadarajan (1974, §3.6.4) is in general distorted. It follows that the condition on the normal torus that we imposed on G is essential.

Part 5.2: The Heat Diffusion Kernel

Overview of Parts 5.2 and 5.3

The reader could or should skip both these parts in a first (or even a second!) reading of the book.

The main aim is to prove the same results as in Part 5.1 but with measures that are not compactly supported. To achieve this we have to overcome two technical difficulties. First, for such measures we do not have the Harnack estimate of §2.5 or again of §5.4 and then the lifting of the operators of §5.6.4 is tricky to carry out unless the support is compact. Rather than battle along with these problems and obtain maximal generality, we decided to bypass these snags and proceed as follows:

(i) In §5.11.2 we give a proof that is based on the ideas of Part 3.3 and this bypasses the lifting altogether.

(ii) To recover the Harnack estimate we have to restrict ourselves to the heat diffusion kernel of the semigroup $e^{-t\Delta}$ where we have Harnack for different reasons altogether.

In the proof that we give in §5.11.2 we have to assume that the group is NB. This works of course for the main theorem but not for the lower estimate of the B-theorem in §1.3.1. This is done in Part 5.3 and here therefore we have to do the lifting of §5.6.4.

So in Part 5.3, either we must restrict ourselves to compactly supported measures and then we can perform the lifting of §5.6.4 or we must 'lift' the heat diffusion kernel. This last point leads to additional complications related to the Gaussian decay of the new kernel. These new problems are certainly interesting but lie outside the main theme of the book. We leave it therefore until Appendix D and Appendix E to explain how one deals with these additional difficulties.

5.9 Preliminaries and the Reductions

Here G will be a connected NB-group and $\triangle$ will denote a sub-Laplacian as in §2.12. And as in (2.52), we write ϕ_t for the heat diffusion kernel of the semigroup $e^{-t\triangle}$. Theorem 4.7 for this semigroup says that there exist constants C, c such that

$$\phi_t(e) \geqslant Ce^{-\lambda t}t^{-c}; \quad t \geqslant 1, \tag{5.51}$$

where $\lambda \geqslant 0$, as in (4.36), is the spectral gap of $\triangle$.

The first observation is that in proving this we may assume that G is simply connected. To see this let $\pi : \widetilde{G} \to G$ denote the simply connected cover and $\widetilde{\triangle}$ some lifting of the Laplacian on $\widetilde{G}$ as in §3.6.1. For $\tilde{\phi}_t$ the corresponding kernel of $e^{-t\widetilde{\triangle}}$ on $\widetilde{G}$ we have, as in (5.46),

$$\phi_t(e) = \sum_{z \in \ker \pi} \tilde{\phi}_t(z). \tag{5.52}$$

Since by §3.1.6 the spectral gap of $\widetilde{\triangle}$ is λ, again we see that the validity of (5.51) for $\tilde{\phi}_t$ implies the same fact for ϕ_t.

The next reduction is as in §5.7.1 and is more subtle. The simply connected group G can be written $G = Q \leftthreetimes S$. This is the Levi decomposition where Q is the radical and S is semisimple. As we saw in §4.6.1 and Remark 5.5, we can find $Z_1 \subset S \cap Z(G)$ some discrete subgroup that is central in G and such that $S_1 = S/Z_1$ has finite centre and admits the Iwasawa decomposition $S_1 = NAK$ (in the sense of §4.6.1) where K is now compact. Again we denote $\pi : G \to G_1 = G/Z_1$ and with $\triangle$ and ϕ_t as above on G we denote by $\triangle_1 = d\pi(\triangle)$ and $\phi_t^{(1)}$ the corresponding convolution kernels. The spectral gap λ of $\triangle_1$ is again the same as for $\triangle$ by §3.1.6 and the reduction consists in showing that it suffices to prove that there exist C, c such that

$$\phi_t^{(1)}(e) \geqslant Ce^{-\lambda t}t^{-c}; \quad t \geqslant 1, \tag{5.53}$$

and that then the corresponding (5.51) follows on G. To see how (5.53) implies (5.51) we shall need to use the Gaussian estimate (see Varopoulos et al., 1992, IX.1.2)

$$\phi_t(g) \leqslant Ce^{-\lambda t} \exp\left(-\frac{|g|_G^2}{ct}\right); \quad t \geqslant 1,\ g \in Z_1, \tag{5.54}$$

where C, c are appropriate constants and $|g|_G$ is as in §1.1 and where we use the fact that for the modular function (see §2.3.5) $m(Z_1) = 1$. It follows that

$$\phi_t(z) \leqslant \phi_t^{1/2}(e)e^{-\lambda t/2} \exp\left(-\frac{|z|_G^2}{2ct}\right); \quad z \in Z_1 \tag{5.55}$$

because the restriction of ϕ_t on Z_1 is positive definite; see the note in §5.7.2. From (5.52) and (5.55) it follows therefore that

$$\phi_t^{(1)}(e) = \sum_{z\in\ker\pi} \phi_t(z) \leqslant C\phi_t^{1/2}(e)e^{-\lambda t/2} \sum_{z\in\ker\pi} \exp\left(-\frac{|z|_G^2}{ct}\right). \tag{5.56}$$

Now, by Lemma 5.6 we have $|z|_G \sim |z|_{Z_1}$ for $z \in Z_1$ and since Z_1 is a finitely generated Abelian group of rank d say, we can estimate the sum in the right-hand side of (5.56) by t^d and conclude that

$$\phi_t^{(1)}(e) \leqslant C\phi_t^{1/2}(e)e^{-\lambda t/2}t^d. \tag{5.57}$$

Our assertion follows from this. If we denote by $R = QNA$ the group that corresponds to the Iwasawa radical, then R is simply connected and soluble and the bottom line is the following result.

Theorem 5.9 (Reduction) *In proving* (5.51) *we may assume that* $G = RK$, *where* R *is simply connected soluble and* K *is a compact subgroup such that* $R \cap K = \{e\}$.

5.10 Gaussian Left-Invariant Operators on Principal Bundles

In this section we shall consider $X = R \times K$, a principal bundle as in §4.3.1, and $\widehat{T} = L \otimes \{*\mu\}$, Markovian left-invariant operators that satisfy conditions (I), (III), and (IV) of §5.3.1, but instead of condition (II) satisfy

(II)′ $\widehat{T}$ is a Gaussian left-invariant operator in the sense of §4.7.

In Appendix E we shall examine these Gaussian operators in more detail.

These conditions suffice for the considerations of §5.3 to go through and again we can construct $x(n) \in X$, $n = 1, 2, \ldots$, the Markov process. For that process the proofs given in §5.3 show that we can again use Criterion 5.1. As in §5.6 we shall be able to use this criterion to prove the following result.

Proposition 5.10 *Let* X, $\widehat{T}$, $\hat{\phi}$ *be as above, where* R *is assumed to be a simply connected soluble* NC-*group. Then there exist constants* $C, c > 0$ *such that*

$$\hat{\phi}_{2n}(e,e) \geqslant Cn^{-c}; \quad n \geqslant 1, \tag{5.58}$$

for any reference point as in §5.3.1.

The proof of this proposition will be given in §5.11 below. But before that we shall show how the above proposition implies estimate (5.51) for the special

NB-groups $G = RK$ as in the reduction theorem (Theorem 5.9). This therefore completes the proof of Theorem 4.7 for the heat diffusion kernel.

To see this we review the argument of §5.3. Briefly, we consider $\triangle$ as in §5.9 on G and we identify G with $X = R \times K$. We write $T^n f = e^{-n\triangle} f = f * \nu^{*n}$, with $\nu^{*n} = \psi_n \, d^r g$. Here it is convenient to go back to the notation of §5.3 and denote the convolution kernel by ψ. We can identify T^n on X with a left-invariant Gaussian operator as in §5.3.4 that admits a continuous kernel $\phi_n(x_1, x_2)$ as in (5.12). We then define χ as in (5.13) and the corresponding T_χ, φ_0, $\widehat{T}$, $\hat{d}k$, $\lambda \geqslant 0$ in §5.3.5. Now $\widehat{T}$ satisfies the conditions of the proposition and the relation between ϕ_n and $\hat{\phi}_n$, the kernel of $\widehat{T}^n$, is as in (5.17), (5.18). Since on the other hand $\psi_n(e_G) \sim \phi_n(e,e)$, as in §5.3.6, estimate (5.58) implies that ψ_t the kernel of $e^{-t\triangle}$ satisfies $\psi_t(e) \gtrsim e^{-\lambda t} t^{-c}$ for $t = 2, 4, \ldots$, as in (5.51). A simple use of the Harnack estimate of §2.12.1 finishes the proof of (5.51) for all $t > 1$.

5.10.1 The Gaussian Euclidean bundle

As in §5.5.1 we can specialise the definition of Gaussian Markovian left-invariant operators to the Euclidean bundles $X = V \times K$, that is, when $R = V$ is a Euclidean space $V = \mathbb{R}^d$. Then as in (5.24),

$$\widehat{T} f(x) = \int_X M(x,y) f(y) \, \hat{d}y; \quad f \in C_0^\infty, \tag{5.59}$$

with $\hat{d}y$ as in §5.5.1 and where M satisfies the same conditions as in §5.5.1 with the only modification being that $M(\gamma; h, k)$ is not compactly supported but satisfies the Gaussian estimate

$$C_1 \exp(-c_1 |\gamma|^2) \leqslant M(\gamma; h, k) \leqslant C_2 \exp(-c_2 |\gamma|^2); \quad \gamma \in V. \tag{5.60}$$

The Markov chain $x(n) \in X$ can then be defined as in (5.26):

$$\mathbb{P}[x(n+1) \in E // x(n) = (\gamma, k)] = \int_E M(\lambda - \gamma; h, k) \, d\lambda \, \hat{d}h. \tag{5.61}$$

We shall prove in §5A the same gambler's ruin estimate as in (5.28) for the conical domain $\Omega \times K$. Here $\Omega = \Omega_{\mathscr{L}} = [u \in V, \; Lu < 1; \; L \in \mathscr{L}]$ is the same conical domain (5.27) for some finite set of linear functionals $L \in V^*$ that satisfy the NC-condition. What will be proved is that the estimate of (5.28) extends to this Gaussian case also.

Remark We shall see in (5.73) below that far less than condition (5.60) is actually needed (see §§3A and 5A, and also §D.5 at the end of Part I of the book). In particular, no 'essential' lower estimates are needed.

5.11 Proof of Proposition 5.10

5.11.1 The plan of the proof

Having given the preliminaries we shall in this section give the proof of Proposition 5.10.

The most obvious way to write the proof down is to imitate the proof given in §5.6. In this approach we first give the proof for special bundles of the form $\widetilde{X} = \widetilde{R} \times K$ where $\widetilde{R} = N \leftthreetimes H$, where N, H are nilpotent, and exploit the fact that for any simply connected soluble group R we can find a group $\widetilde{R}$ as above and a projection $\widetilde{R} \to R$. The two technical points that were used in passing from X to $\widetilde{X}$ were the Harnack estimate of §5.4 and the lifting of $\widehat{T}$ on X to $\widetilde{T}$ on $\widetilde{R}$ of Lemma 5.4. These two points in the generalisation of this approach will be dealt with in Appendix D. We shall give no more details here on this construction because it will not be used immediately in what follows. Instead of adopting this approach it is more direct to generalise the approach of Part 3.3, which does not require the lifting of $\widehat{T}$ to an overgroup $\widetilde{R} \to R$.

In this approach we shall use the decomposition of the soluble simply connected NC-group $R = N_R \leftthreetimes Q_R$ of §3.9.2, where N_R is nilpotent and Q_R is a soluble R-group. We then proceed exactly as for the group $N \leftthreetimes H$ in §5.6.1(ii) and show that the conditions of Criterion 5.1 are satisfied. We have to adapt the argument of §5.6.3 and take into account the following two changes:

(i) Q_R is a soluble R-group (see §2.2.2) but not necessarily nilpotent;
(ii) $\widehat{T}$ is not compactly supported (as in condition II of §5.3.1) but it is Gaussian as in §4.7.

This approach offers, of course, a new way of giving the proof of Proposition 5.2 which, although based on the same principles, differs substantially in the detail.

5.11.2 Verification of the conditions of Criterion 5.1 on $X = R \times K$, $R = N_R \leftthreetimes Q_R$

We shall follow very closely the proof given in §§5.6.2 and 5.6.3 and start by adapting the notation there. Then as we go along we shall simply point out the changes that have to be made. In particular, a good understanding of these two sections as well as Part 3.3 is essential in this subsection. The reader could or should skip the details in a first reading.

Again, we set $x(j) = r_j k_j \in X = R \times K, r_j \in R, k_j \in K$, for the Markov chain generated by $\widehat{T}$ where now $r_j = n_j h_j$ with $n_j \in N_R, h_j \in Q_R$, and where again we write $r_j = \tilde{r}_1\tilde{r}_2\cdots\tilde{r}_j$, $\tilde{r}_j \in R$, and also $y(j) = h_j k_j \in Y$ for the projected Markov chain by the projection $X \to Y = Q_R \times K$ that is induced by the canonical projection $R \to Q_R$. We also set $\tilde{r}_j = \tilde{n}_j\tilde{h}_j, \tilde{n}_j \in N_R, \tilde{h}_j \in Q_R$. We assign the starting probability $\mathbb{P}_0$ with $x(0) = e_X = (e_R, e_K)$ a reference point as in §5.3.1. Here (5.37) no longer holds; instead we have the following Gaussian estimate. We fix $s \geqslant 1$ and consider the event

$$\mathscr{B}_s = [|\tilde{r}_j|_R \leqslant \log(s+10);\ j = 1,\ldots,s]; \tag{5.62}$$

then by the Gaussian property of $\widehat{T}$ (see §§4.7 and 5.10) and the product formula (4.11) we have

$$\mathbb{P}_0[\sim \mathscr{B}_s] = O(\exp(-c\log^2(s+10))) \tag{5.63}$$

for the complementary event since this is the union of s events that satisfy the same estimate. We now use the distance distortion of (2.59) which holds because N_R lies in the nilradical of R and we see that on $\mathscr{B}_s$ we have

$$|\tilde{n}_j|_{N_R} \leqslant Cs^c, \quad |\tilde{h}_j|_{Q_R} \leqslant C\log(s+10); \quad 1 \leqslant j \leqslant s, \tag{5.64}$$

for appropriate constants (the second estimate follows from the projection $R \to Q_R$). Expressions (5.62) and (5.64) combined are the replacement for (5.37).

As in (5.40) we denote

$$B_s = \left[r = nh,\ n \in N_R,\ h \in Q_R;\ |n|_{N_R} \leqslant Cs^C,\ |h|_{Q_R} \leqslant Cs^C\right] \subset R, \tag{5.65}$$

for some appropriate large C, and our aim is to verify that Criterion 5.1 holds for the sets

$$E_s = B_s \times K \subset X; \quad s = 1,2,\ldots. \tag{5.66}$$

Condition (i) is evident as before because for the measure $\hat{d}x = d^r r \otimes \hat{d}k$ we have $d^r r = dndh, n \in N_R, h \in Q_R$ (notice that both these groups are unimodular).

To verify condition (ii) of the criterion we proceed as before and we write the analogue of (5.41):

$$r_j = \tilde{n}_1\tilde{h}_1\tilde{n}_2\tilde{h}_2\cdots\tilde{n}_j\tilde{h}_j = \tilde{n}_1\tilde{n}_2^{h_1}\cdots\tilde{n}_j^{h_{j-1}}h_j. \tag{5.67}$$

This will be used together with the following two observations:

(i) We consider the canonical mapping $Q_R \to Q_R/N \cap Q_R = R/N = V$ where N is the nilradical of R and where then V is a Euclidean space (we have already noted in §3.8.4 that $N \cap Q_R$ is an analytic subgroup but, just as in §3.8.4, this fact is not essential here). We shall compose this mapping with the projection $X \to Y \to V \times K$ and project $x(j)$ first to $y(j) \in Y$ and then to $v(j) = v_j k_j \in$

$V \times K$, $v_j \in V$, $k_j \in K$. Here $V \times K$ is a Euclidean vector bundle as in §5.10.1 and we can apply the gambler's ruin estimate to deduce that the event

$$\mathscr{P}_s = [Lv_j < 1;\ 1 \leqslant j \leqslant s,\ L \in \mathscr{L}] \tag{5.68}$$

has probability

$$\mathbb{P}_0(\mathscr{P}_s) \geqslant cs^{-c}; \quad s \geqslant 1, \tag{5.69}$$

for appropriate constants, where $\mathscr{L}$ here is the set of real roots of the action of $\mathfrak{q}_R$ on $\mathfrak{n}_R$. They were defined in (3.53) and (3.59). For all this the reader should go back to the discussion in §§3.8.2–3.8.4.

(ii) The next observation concerns $\|\operatorname{Ad}(h_j)\|_{\mathfrak{n}_R}$ and here we use the fact that on the event $\mathscr{B}_s$, we have $|\tilde{h}_j|_{Q_R} \lesssim \log(s+1)$ and recall that $h_j = \tilde{h}_1 \cdots \tilde{h}_j$. If we use (3.66) (as we did in (3.69), (3.70)) we deduce therefore that on the event $\mathscr{B}_s \cap \mathscr{P}_s$ we have $\|\operatorname{Ad}(h_j)\|_{\mathfrak{n}_R} \leqslant Cs^c$ for appropriate constants. Notice that here in the use of (3.66) the parameter $A = O(\log s)$.

If we combine this with the polynomial distortion of the exponential mapping (2.60) and use (5.64) we finally conclude that with appropriate constants we have

$$|\tilde{n}_j^{h_{j-1}}|_{N_R} \leqslant Cs^c,\ 1 \leqslant j \leqslant s,\ \text{ almost surely on } \mathscr{P}_s \cap \mathscr{B}_s; \quad s \geqslant 1. \tag{5.70}$$

This combined with (5.64) and (5.67) gives $x(s) \in E_s$ almost surely on $\mathscr{P}_s \cap \mathscr{B}_s$, $s \geqslant 1$, provided that the constants of the definition of E_s in (5.65) and (5.66) have been chosen appropriately large. Given (5.63) and (5.69), this fact gives the required verification of condition (ii) of the criterion. And this completes the proof of the proposition. It goes without saying that, since the criterion only gives the estimate of the kernel for even time $2n$, the Harnack principle, that as we know holds (see (2.54)), has to be used at the end.

Part 5.3: Proof of the Lower B-Estimate

5.12 Statement of the Results and Plan of the Proof

The proofs in the third part of this chapter are modifications of those given in Parts 5.1 and 5.2. Here we shall consider again $\phi_n(g)\,dg = d\mu^{*n}(g)$ for a compactly supported symmetric measure on the connected Lie groups G as in Theorem 4.7 or $\phi_t(g)$, $t > 0$, which is the convolution kernel of $e^{-t\triangle}$ for a sub-Laplacian as in §5.9. As before, we shall denote $e^{-\lambda} = \|\mu\|_{\text{op}}$ (see §3.1.1) for the spectral gap of $\triangle$ as in §4.8. In the proof below, these two cases will be treated simultaneously. The novelty of the situation lies in the fact that it is independent of the B–BN classification and that G is general. We shall prove the following universal lower estimate.

Theorem 5.11 *Let G and μ, λ be as above and let $e \in G$ be the identity of G. Then there exist constants $C,c>0$ such that*

$$\phi_n(e) \geqslant C\exp(-\lambda n - cn^{1/3}); \quad n \geqslant 2. \tag{5.71}$$

For the classification as formulated in §1.10 estimate (5.71) is not essential. Therefore, the reader who wishes to stay in the mainstream of the subject could skip the proof in the rest of Part 5.3.

The proof of this theorem follows the same strategy as that of Theorem 4.7. We shall prove it first when $G = RK$ where R is a soluble, simply connected closed subgroup as in §4.5.2. Once this is done the reduction given in §5.7 and the use of positive-definite functions (as in §5.7.2) allows us to pass to general groups. The details of these reductions will not be repeated here.

For the proof for these special groups $G = RK$ we use the methods of §§5.3 and 5.4 and consider the principal bundle $X = R \times K$ with R simply connected soluble and the left-invariant Markovian operators $\hat{T}$ as in §5.3.1. These operators satisfy conditions (I)–(IV) of §5.3.1 in the compactly supported case or the conditions of §5.10 in the case of the convolution kernel of $e^{-t\triangle}$. So far nothing changes and the kernel $\hat{\phi}_n(x_1,x_2)$, $x_1,x_2 \in X$; $n \geqslant 1$ is defined as in (5.7). We shall need the following version of Proposition 5.2.

Proposition 5.12 *Let X, $\widehat{T}$, $\hat{\phi}_n$ be as above. Then there exist $C,c>0$ such that*

$$\hat{\phi}_{2n}(e,e) \geqslant C\exp(-cn^{1/3}); \quad n \geqslant 1, \tag{5.72}$$

for every reference point e as in §5.3.

Once the proposition has been proved, the theorem follows as before, where §5.3.6 and (5.20) are now used.

Here the group R is not necessarily NC and therefore the new difficulty in the proof of the proposition lies in the Gaussian case because we cannot use §3.8.1 and assume as in §5.11 that $R = N_R \leftthreetimes G_R$. We must therefore fall back on the method of §5.6 and first prove the proposition for the special groups $R = N \leftthreetimes H$ where N, H are simply connected nilpotent. As we pointed out in §5.11.1, difficulties arise when we pass from these special groups to the general soluble groups of the general case. To wit, once this special case has been proved and if R is a general simply connected soluble group, we can construct $\widetilde{R} = N \leftthreetimes H$ as in (3.14) and a projection $\widetilde{R} \to R$ and $\widetilde{X} = \widetilde{R} \times K$ and the induced projection $\widetilde{X} \to X$. From the validity of the proposition on $\widetilde{X}$ we must then deduce the validity on X. To make this reduction we must lift the Markovian operator $\widehat{T}$ on X to a Markovian operator $\widetilde{T}$ on $\widetilde{X}$. This is exactly what was done in Lemma 5.4 in the compactly supported case when $\hat{T}$ satisfied (I)–(IV) and there was no problem. The problems that arise in the lifting of the heat diffusion kernel of the operator that corresponds to $\triangle$ on the bundle X are explained in Appendix D: they are non-trivial. Having said all this, in the next section we shall give the proof of the proposition for the special case $R = N \leftthreetimes H$.

5.12.1 The Euclidean principal bundle revisited

Let $X = V \times K$, $V = \mathbb{R}^d$ be a Euclidean principal bundle and let $\widehat{T}$ be a Markovian left-invariant operator as in §§5.5.1 and 5.10.1. We shall write $|x| = |v|_V$ for $x = (v,k) \in X$ and we shall maintain the notation of §5.10.1 and denote by $M(\gamma; h, k)$ the kernel of $\widehat{T}$ as in (5.24) and (5.59). Here M will be assumed to be continuous and symmetric as in (5.25) with $M(0,h,k) > 0$, $M(-\gamma; h,k) = M(\gamma; k,h)$, but, rather than the compactness of the support in §5.5.1 or the Gaussian estimate in (5.60), we shall impose the more general condition

$$\int_V |\gamma|^D M(\gamma; h,k)\, d\gamma < C; \quad h,k \in K, \tag{5.73}$$

for C independent of h, k and some high enough moment D.

We shall again denote by $x(n) \in X$ the Markov chain generated by $\widehat{T}$. We shall consider the starting probability $\mathbb{P}_0$ where $\mathbb{P}_0[x(0) = e] = 1$ for the reference point of (5.72). In §5B we shall prove the following result.

Proposition 5.13 *Let the Markov chain and the notation be as above*; *then there exist* $C, c > 0$ *such that*

$$\mathbb{P}_0\left[|x(j)| < M;\ j = 1,\ldots,t\right] \geqslant c \exp\left(-\frac{t}{cM^2}\right); \quad t, M \geqslant C. \tag{5.74}$$

The conditions $t, M > C$ imposed are in fact not essential. The estimate gives from below the same estimate that we proved from above for sums of independent random variables in the appendix to Chapter 2. How much simpler these lower estimates are for sums of independent random variables is not clear. The probabilistically inclined reader could ponder this. At any rate here (5.74) will be assumed and in the next section we shall finish the proof of Proposition 5.12 in our special case.

5.13 Proof of Proposition 5.12. Special Case

Here, as explained in §5.11.1, we shall assume that $R = N \leftthreetimes H$ where both N and H are simply connected nilpotent groups and $\widehat{T}$ will be as in the proposition. We shall follow very closely the proof in Part 5.2 and the notation will be as in §§5.11.2 and 5.6.3. Recall for the Markov chain induced by $\widehat{T}$ that we write $x(j) = r_j k_j \in X = R \times K$, with $r_j \in R, k_j \in K$ and $r_j = n_j h_j$, with $n_j \in N$, $h_j \in H$. We then define $r_j = \tilde{r}_1 \cdots \tilde{r}_j$ with $\tilde{r}_j = \tilde{n}_j \tilde{h}_j \in R, \tilde{n}_j \in N, \tilde{h}_j \in H$ and we clearly have $h_j = \tilde{h}_1 \cdots \tilde{h}_j$ (but *not* $n_j = \tilde{n}_1 \cdots \tilde{n}_j$!). We shall use the reference point e as in the proposition and the starting probability of the Markov process $\mathbb{P}_0[x(0) = e]$. We shall modify the definition of (5.62) and define instead the events

$$\mathscr{B}_s = \left[|\tilde{r}_j|_R \leqslant cs^{1/3};\ j = 1, \dots, s \right]; \quad s \geqslant 1, \tag{5.75}$$

for appropriate constants. Exactly as before we then have

$$\begin{gathered} \mathbb{P}_0[\sim \mathscr{B}_s] = O(\exp(-cs^{2/3})), \\ |\tilde{n}_j|_N \leqslant \exp(Cs^{1/3}) \quad \text{and} \quad |\tilde{h}_j|_H \leqslant Cs^{1/3} \text{ on } \mathscr{B}_s, \end{gathered} \tag{5.76}$$

for appropriate constants.

For the proof of the proposition we shall use the modified criterion §5.3.2 for the sequence of sets

$$\begin{aligned} E_s &= B_s \times K \subset X, \\ B_s &= \big[r = nh;\ n \in N,\ |n|_N \leqslant \exp(cs^{1/3}), \\ &\qquad\qquad h \in H,\ |h|_H \leqslant \exp(cs^{1/3}) \big]; \quad s \geqslant 1, \end{aligned} \tag{5.77}$$

for appropriate constants. With $\hat{d}k$ as in §5.3 and $\hat{d}x = d^r r \otimes \hat{d}k$, the $\hat{d}x$ measure of E_s is clearly $|E_s| \sim$ (d^r measure of $B_s \subset R$) $\lesssim \exp(cs^{1/3})$. The remaining issue is therefore to prove condition (ii)$'$ of the criterion in §5.3.2 and for this, formula (5.41) is used again:

$$r_j = \tilde{n}_1 \tilde{h}_1 \cdots \tilde{n}_j \tilde{h}_j = \tilde{n}_1 \tilde{n}_2^{h_1} \cdots \tilde{n}_j^{h_{j-1}} h_j. \tag{5.78}$$

Towards that, we fall back on the notation of §5.6.3 and consider $[H,H]$ the connected closed Lie subgroup that corresponds to $[\mathfrak{h}, \mathfrak{h}]$, where $\mathfrak{h}$ is the Lie algebra of H; we shall also denote $V = H/[H,H]$ and by $p : H \to V$ the canonical projection. We then project the Markov chain $x(j)$ first to $y(j) \in H \times K$ by the projection induced by the canonical $R \to H$ and then use the projection induced by p to project to $v(j) = p(y(j)) = v_j k_j \in V \times K$, $v_j \in V$, $k_j \in K$. On the Euclidean bundle $V \times K$ we obtain thus a Markov chain as in

§5.12 that satisfies (5.73) in both the case of compactly supported measures and that of the heat diffusion kernel.

Exercise In the case of the diffusion kernel we obtain on $V \times K$ the diffusion induced by some V-invariant operator. This, together with Appendix D, will certainly imply (5.73) and this is undoubtedly the best way to prove (5.73) in that case. The disadvantage is of course that we have to use Appendix D. We can avoid this by starting from the Gaussian estimate that is known to hold for the heat diffusion kernel on G itself. From this we can then adapt the argument that was used in §2.14.4 to obtain the required result. The details are left as an exercise.

The events $\mathscr{P}_s$ of (5.68) will now be defined as

$$\mathscr{P}_s = \left[|v_j|_V \leqslant s^{1/3};\ j = 1, \ldots, s\right]; \quad s \geqslant 1. \tag{5.79}$$

The probabilistic estimate (5.74) will be used and we can readily verify that there exists $c > 0$ such that

$$\mathbb{P}_0[\mathscr{P}_s] \geqslant \exp(-cs^{1/3}); \quad s \geqslant 1. \tag{5.80}$$

We shall now use the algebraic Lemma 3.4 and the final comment of §3.7.1, and deduce that on the event $\mathscr{P}_s \cap \mathscr{B}_s$ we have

$$\| \operatorname{Ad}(h_j) \|_{\mathfrak{n}} \leqslant \exp(cs^{1/3}) \tag{5.81}$$

for appropriate c where $\| \quad \|_{\mathfrak{n}}$ indicates the operator norm on the Lie algebra $\mathfrak{n}$ of N for some fixed Euclidean norm on $\mathfrak{n}$. Then in (5.78) set $\tilde{n}_j = \exp(\xi_j)$, $\xi_j \in \mathfrak{n}$. By (5.76) and the polynomial distortion (2.60) of the exponential mapping for a nilpotent group, we then have $|\xi_j|_{\mathfrak{n}} \leqslant \exp(cs^{1/3})$ on $\mathscr{B}_s \cap \mathscr{P}_s$; and by (5.81) we have $|\operatorname{Ad} h_j(\xi_j)|_{\mathfrak{n}} \leqslant \exp(cs^{1/3})$ for appropriate constants c. The estimate

$$|\tilde{n}_j^{h_{j-1}}|_N \leqslant \exp(cs^{1/3}); \quad j = 1, \ldots, s \tag{5.82}$$

then follows by the exponential mapping (3.22) and the same polynomial distortion. This together with (5.76) inserted in (5.78) finally gives the required result, namely that on $\mathscr{B}_s \cap \mathscr{P}_s$ we have $x(s) \in E_s$ for all $s \geqslant 1$. This verifies condition (ii)$'$ of the criterion in §5.3.2 because by (5.76) and (5.80), $\mathbb{P}_0[\mathscr{B}_s \cap \mathscr{P}_s] \geqslant \exp(-cs^{1/3})$ for some c. So we are done.

5A Appendix: Proof of the Gambler's Ruin Estimate §5.5.2

5A.1 μ-subharmonic functions

This subsection consists of exercises in calculus. Clearly the reader could look at the conclusions and skip the elementary but elaborate details. I should also add that the same thing can be said about §5B.1 but that apart from these first two subsections, the rest, while being essentially self-contained, relies on rather sophisticated ideas from probability theory and homogenisation theory (see 'The final word' at the end). How deeply the reader wishes to immerse themselves in these subjects is of course their choice.

We shall consider throughout $\mu \in \mathbb{P}(\mathbb{R}^n)$ that are centred in the sense that $\int x\,d\mu(x) = 0$ and admit high enough moments: $\int |x|^D\,d\mu(x) < +\infty$ for some $D \gg 1$. We shall denote by $\sigma = (\sigma_{ij})$ the covariance matrix. We shall denote by $E(a,A)$ with $A, a > 0$ the class of measures as above for which $\operatorname{supp}\mu \subset [|x| < A]$ and which satisfies

$$a^{-1}I \leqslant \sigma \leqslant aI; \quad \sigma_{ij} = \int x_i x_j\,d\mu(x) \tag{5A.1}$$

for the identity matrix $I = (\delta_{ij})$.

Now let F be a real continuous function defined in the domain $\Omega \subset \mathbb{R}^n$. We shall then say that F is $E(a,A)$-subharmonic at $x \in \Omega$ if distance $(x, \partial\Omega) > A$, $F(x) \leqslant F * \mu(x)$ for all $\mu \in E(a,A)$. When $A = +\infty$ then we must assume that $\Omega = \mathbb{R}^n$ and that the convolution is absolutely convergent. The aim of this appendix is to construct important classes of subharmonic functions in conical domains. We shall therefore start with some notation.

Polar coordinates We denote the unit sphere by $S^{n-1} \subset \mathbb{R}^n$ and the north pole by $I = (1,0,0,\ldots,0)$. We write as usual $r = |x|$, $x \in \mathbb{R}^n$, and denote the colatitude by θ, that is, the scalar product $\langle I, x\rangle = r\cos\theta$. If $(\varphi_1, \ldots, \varphi_{n-1})$ are local coordinates that are defined in appropriate subregions of S^{n-1} and $\varphi_1 = \theta$ then $(r, \varphi_1, \ldots, \varphi_{n-1})$ are the polar coordinates that we shall use in the conical region $0 < \theta < 1/10$. For every θ_0 small enough and $\nu, k \geqslant 0$ we shall then define the function $0 \leqslant F_{\nu,k}(x) = r^\nu u(\theta) \in C^\infty$ for $0 \leqslant \theta < \theta_0$ with $u_k = u$ defined as follows:

$$u(\theta) \begin{cases} \equiv 1, & 0 \leqslant \theta \leqslant \theta_0/3, \\ = (\theta_0 - \theta)^k, & \theta \in \;]2\theta_0/3, \theta_0[; \end{cases} \tag{5A.2}$$

in between $u_k = u(\theta)$ is smooth and positive.

In the next few lines we shall give the main construction and show that *for all $a > 0$, $A < +\infty$, we can choose ν, k (and $\theta_0 > 0$) as above and $B > 0$*

so that $F = F_{\nu,k}$ *is* (A,a)*-subharmonic at every* $x \in \Omega \cap [F > B] = \Omega_B$ *with* $\Omega = [\theta < \theta_0]$. *Furthermore, we can make that choice so that* Ω_B *approximates the conical region* Ω *in the sense that there exist* $\theta_0' < \theta_0$ (*in fact as close to* θ_0 *as we like*) *and* $R > 0$ *so that*

$$\Omega(\theta_0', R) = [\theta < \theta_0';\ |r| \geqslant R] \subset \Omega_B. \tag{5A.3}$$

The same construction can also be carried out with $A = +\infty$ and $\Omega = \mathbb{R}^n$ provided that the measures admit a high enough moment. Here, when $A = +\infty$, we shall only need this construction when the family of measures satisfy (5A.1) and are in addition Gaussian with fixed constants as in §2.12.1.

In the constructions that follow we shall impose the condition $0 < \nu < k$ and this will guarantee that the boundary $\partial\Omega_B = [r, \theta;\ r^\nu u(\theta) = B]$ for large B is close to the boundary of Ω. To wit,

on the boundary, if r is large enough, $r^{\nu/k}(\theta_0 - \theta) = B^{1/k}$ and

$$\delta(x) = \text{distance}(x, \partial\Omega) \sim r(\theta_0 - \theta) = r^{1-\nu/k} B^{1/k} \to \infty, \tag{5A.4}$$

as $B \to \infty$, but also $\delta(x) = o(r)$ for fixed B. This implies the inclusion (5A.3) and the fact that distance$(\Omega_B, \partial\Omega) > A$ if B is large enough and A is finite.

To fix ideas, assume first that $A < +\infty$, and that $\mu \in E(a, A)$.

We shall use the notation of (5A.1) and the Taylor series remainder term estimate

$$(\mu - \delta)F(x) = \sum_{i,j} \sigma_{ij} H_{ij}(x) + O\left(\sup_{|x-y|<A} \left|\nabla^3 F(y)\right| \right), \tag{5A.5}$$

where $H_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}$ are the coefficients of the Hessian, $\nabla^3 = \frac{\partial^3}{\partial y_i \partial y_j \partial y_k}, \ldots$ indicates the third-order gradient for Euclidean coordinates and δ is the Dirac mass at the origin.

To compute the Hessian for $r > 0$ and $0 < \theta < \theta_0$ with θ_0 small enough we observe that

$$\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right) = M \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial}{\partial \varphi_1}, \ldots, \frac{1}{r}\frac{\partial}{\partial \varphi_{n-1}} \right) \tag{5A.6}$$

(recall: $\varphi_1 = \theta$), where the matrix M and its inverse M^{-1} are locally C^∞ in $\theta \neq 0$ and independent of r and where we restrict ourselves of course to a coordinate patch on which the local coordinates $\varphi_1, \ldots$ have been defined. It follows that the principal term in the right-hand side of (5A.5) can be written as

$$s_{0,0}\frac{\partial^2 F}{\partial r^2}+2\frac{1}{r}\sum_{i=1}^{n-1}s_{0,i}\frac{\partial^2 F}{\partial r\,\partial\varphi_i}$$
$$+\frac{1}{r^2}\sum_{i,j=1}^{n-1}s_{i,j}\frac{\partial^2 F}{\partial\varphi_i\,\partial\varphi_j}+\alpha\frac{1}{r}\frac{\partial F}{\partial r}+\frac{1}{r^2}\sum_{i=1}^{n-1}\beta_i\frac{\partial F}{\partial\varphi_i}, \tag{5A.7}$$

where here the coefficients are independent of r and are C^∞ in $\theta\neq 0$. Furthermore, the matrix $(s_{i,j})$, $i,j=0,\dots,n-1$ is symmetric and satisfies (5A.1) for some new $a>0$ as long as $|\theta|\geqslant\theta_0/10$ where θ_0 has been fixed once and for all and is small as in (5A.2). (This is clear if $r\sim 1$. Then for any large but fixed R_0, in the region $r\sim R_0$ scale $x\to R_0^{-1}x$.) For the special function $F=r^\nu u(\theta)$, (5A.7) becomes for large ν,

$$r^{\nu-2}\big[(\nu(\nu-1)s_{0,0}+\nu\alpha)u+(2\nu s_{0,1}+\beta_1)u'+s_{1,1}u''\big]$$
$$=r^{\nu-2}\big[\nu^2(s_{0,0}+O(1/\nu))u+2\nu(s_{0,1}+O(1/\nu))u'+s_{1,1}u''\big]. \tag{5A.8}$$

The computation of ∇^3F is elementary. For the estimates that we shall need we shall set $\partial_0=\frac{\partial}{\partial r}$, $\partial_i=\frac{1}{r}\frac{\partial}{\partial\varphi_i}$ and $\nabla=M\partial$ for the matrix M in (5A.6) and $\partial=(\partial_0,\partial_1,\dots)$. Then $\nabla^2F=M\partial M\partial F+M^2\partial^2F$ and

$$\nabla^3F=\partial^2M\partial F+\partial M\partial^2F+\partial^3F=\frac{1}{r^2}\partial F+\frac{1}{r}\partial^2F+\partial^3F, \tag{5A.9}$$

where in this formula we have suppressed the brackets and also, in some places, M and other factors that stay bounded for $\theta>\theta_0/10$.

For large ν for the function $F=r^\nu u(\theta)$, we can estimate (5A.9) by

$$r^{\nu-3}\big[\nu^3|u|+\nu^2|u'|+\nu|u''|+|u'''|\big]. \tag{5A.10}$$

If we specialise to the function $F_{\nu,k}$ and $\theta_0-\theta\leqslant\theta_0/3$, the principal term (5A.8) can be rewritten as

$$r^{\nu-2}(\theta_0-\theta)^{k-2}\big[(s_{0,0}+O(1/\nu))\xi^2+2k(s_{0,1}+O(1/\nu))\xi$$
$$+k(k-1)s_{1,1}\big];\quad \xi=\nu(\theta_0-\theta). \tag{5A.11}$$

If we take the inf in ξ of the term in the brackets we see that in that range this principal term is bounded from below by

$$\begin{aligned}&r^{\nu-2}(\theta_0-\theta)^{k-2}(s_{0,0}+O(1/\nu))^{-1}D,\\&D=k(k-1)s_{1,1}(s_{0,0}+O(1/\nu))-k^2(s_{0,1}+O(1/\nu))^2.\end{aligned} \tag{5A.12}$$

If we combine this with (5A.1), that is satisfied for the $s_{i,j}$, $i,j=0,1$, we deduce the following lower estimate.

Lemma *There exist $N, c > 0$ such that in the range $\theta_0 - \theta < \theta_0/3$, $\theta_0 \leqslant 10^{-10}$ and $\nu, k \geqslant N$ the principal term* (5A.7) *is bounded from below by $ck^2 r^{\nu-2}(\theta_0 - \theta)^{k-2}$, $r > 0$.*

The choice of the parameters and the end of the construction The aim here is to use the above estimates and choose the parameters $k > \nu \gg 1$ and B to guarantee the positivity of (5A.5): $\theta_0 < 10^{-10}$ has been fixed once and for all and we shall choose ν large enough, and k, say, so that $\nu < k < 2\nu$. We shall consider separately the following three ranges.

(i) *The range* $\theta_0 - \theta \leqslant \theta_0/3$ Here $k > \nu \geqslant N$ as in the lemma and arguing as in (5A.4) we see that for r large and $x = (r, \theta, \dots) \in \Omega_B$ we have $\delta(x) \approx r(\theta_0 - \theta) \geqslant r^{1-\nu/k} B^{1/k}$, and also $r^\nu \geqslant B(\theta_0 - \theta)^{-k}$ and that if ν, k have been chosen and fixed, then $\delta(x) \to \infty$, $r \to \infty$ when $B \to \infty$.

The error term (5A.10) in that range can be estimated from above by

$$r^{\nu-3}(\theta_0 - \theta)^{k-3}(\nu^3 + k^3); \quad x \in \Omega_B. \tag{5A.13}$$

Indeed, a moment's reflection shows that provided that B is large enough we can estimate the value of

$$\sup_{|y-x| \leqslant A} \left|\nabla^3 F(y)\right| \sim \left|\nabla^3 F(x)\right|; \quad x \in \Omega_B \tag{5A.14}$$

by the value at x and ignore the sup (because passing from x to y amounts to replacing r by cr and $\theta_0 - \theta$ by $\theta_0 - \theta + \frac{c}{r}$ and then we use $\delta(x) \sim r(\theta_0 - \theta) \gg 1$). It is of course here that we use the fact that $A < +\infty$.

The upshot is therefore that we first choose ν, k large and then B large enough in terms of this choice. If we use the lemma, the positivity of (5A.5) follows because the extra factor $r(\theta_0 - \theta)$ that we find in the lemma compared to what we have in (5A.13) is $r(\theta_0 - \theta) \approx \delta(x)$, and as we saw this tends to ∞.

(ii) *The range* $\theta_0/3 \leqslant \theta \leqslant 2\theta_0/3$ By (5A.8) the principal term in (5A.5) is bounded from below by $c\nu^2 r^{\nu-2}$ provided that $\nu \geqslant N_1$ is large enough. Similarly, the error term can be estimated by $\nu^3 r^{\nu-3}$. Once more we first choose ν and then B (and therefore r) large enough.

(iii) *The range* $\theta < \theta_0/3$ In this range (5A.7), (5A.8) and (5A.9) still hold (but in the local coordinates $\varphi_1, \varphi_2, \dots, \varphi_{n-1}$ that we use we cannot take $\theta = \varphi_1$) and here the principal term in (5A.8) is $\nu(\nu-1)s_{0,0}r^{\nu-2} + \alpha\nu r^{\nu-2}$ and the error term in (5A.10) is $C\nu^3 r^{\nu-3}$. The positivity of (5A.5) clearly follows for $\nu > 3$ and r large enough. Of course, here $F(x_1, \dots, x_n) = r^\nu$, for $\nu = 2$, is convex because it is the sum of the convex functions x_i^2, so clearly F is convex for $\nu \geqslant 2$. As long as the support of the measure is compact, one therefore does not need to use the above argument. This completes the construction for compactly supported measures. Note that in §5B below we shall need to compute the Hessian of another radial function.

Measures with non-compact support The above proof can easily be adapted to centred measures that satisfy (5A.1) and the more general condition $\int |x|^D \, d\mu(x) < +\infty$, for some D large enough, but that are not necessarily compactly supported.

The main new observation is that for any degree of smoothness p we can extend the function (5A.2) to be zero outside $[\theta < \theta_0] = \Omega$ and then $F_{\nu,k} \in C^p(\mathbb{R}^n)$ is sufficiently smooth provided that ν, k are large enough. Then for any measure as above we can still use the Taylor remainder formula in (5A.5) but now we shall use the integral remainder formula that gives the remainder in the following form

$$R(x) = \sum_{\alpha} \frac{3}{\alpha!} \int \frac{\partial^\alpha}{\partial x_\alpha} F(x+u) \, d\lambda_\alpha(u), \tag{5A.15}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ runs through all the multi-indices with $|\alpha| = \alpha_1 + \cdots = 3$ and λ_α is now a bounded, but no longer necessarily positive, measure on $\mathbb{R}^n$ and is defined to be the image of the measure $(1-t)^2 y^\alpha \, d\mu(y) \, dt$ on $[0,1] \times \mathbb{R}^n$ by the mapping $(t,y) \to ty$ (see Hörmander, 1983, (1.1.7)$'$).

The new measures λ_α also admit D-moments

$$\int |y|^D \, d|\lambda_\alpha|(y) < \infty. \tag{5A.16}$$

In what follows we shall estimate (5A.15) for $F = r^\nu u_k(\theta)$ as in (5A.2) and show that there exist appropriate constants such that

$$|R(x)| \leqslant C_{\nu,k} \, r^{\nu-3} (\theta_0 - \theta)^{k-3}; \quad x \in \Omega_B, \tag{5A.17}$$

provided that ν, k, B are large and appropriately chosen as before. This is the analogue of (5A.13) and once this is done the rest of the argument remains unchanged. More precisely, we again use the lemma, (5A.8) and the considerations of (ii) and (iii) when $\theta \leqslant 2\theta_0/3$.

In the integrands of (5A.15) for $|x| = r$ large enough and ν, k large, we can use polar coordinates and arguing as in (5A.14) we see that there exist constants that depend on ν, k such that we can use (5A.10) and we estimate (5A.15) by

$$\begin{gathered}\int (r + O(\xi))^{\nu-3} U(\theta + O(\xi)/r) \, d\sigma(\xi); \\ 0 \leqslant U(\theta) \leqslant C_{\nu,k} (\theta_0 - \theta)^{k-3} \quad \text{for } x = (r, \theta, \ldots) \in \Omega,\end{gathered} \tag{5A.18}$$

where U is taken to be $\equiv 0$ for $\theta \geqslant \theta_0$ and where σ is an appropriate positive measure that satisfies the same moment condition (5A.16) and is independent of ν, k. We can compute and estimate (5A.18) by

$$\int_{\mathbb{R}^d} (r+|\xi|)^{\nu-3} \left(|\theta_0-\theta| + \frac{|\xi|}{r} \right)^{k-3} d\sigma(\xi).$$

If we expand the powers in the integrand we see that we can estimate by $r^{\nu-3-\alpha-\beta}(\theta_0-\theta)^{k-3-\beta}\int|\xi|^{\alpha+\beta}\,d\sigma(\xi) \lesssim r^{\nu-3}(\theta_0-\theta)^{k-3} r^{-\alpha}(r(\theta_0-\theta)^{-\beta}$, with $\alpha,\beta \geqslant 0$, because the integral can be controlled by the moment condition. Therefore if we use the fact that $r(\theta_0-\theta) \approx \delta(x)$ is bounded from below, we are done.

The bottom line is that we have constructed some function in $\mathbb{R}^d$ that is subharmonic on Ω_B for our measures provided they admit a high enough moment. This in particular holds for the Gaussian measures as in §2.12.2.

5A.2 Harmonic coordinates on the Euclidean principal bundle and the gambler's ruin estimate

Here we shall go back to the Euclidean vector bundle of §5.5 and maintain the notation that was introduced there but drop the caret and write dh rather than $\hat{d}h$. We denote by $x(n) = (v(n),k(n)) \in X = V \times K$, $v(n) \in V = \mathbb{R}^d$ the Markov chain that was introduced there by the left-invariant operator $\widehat{T}$ and write $D_n = v(n+1) - v(n)$. By (5.26) we have therefore

$$\begin{aligned}\mathbb{P}[D_n \in d\xi // x(n){=}(v,k)] &= \mu_k(d\xi) \\ &= \left(\int_K M(\xi;h,k)\,dh\right) d\xi; \quad v \in V,\ k \in K,\end{aligned} \tag{5A.19}$$

where $//$ denotes, as usual, conditional probability or conditional expectation. When M is compactly supported as in §5.5.1, the measures μ_k ($k \in K$) are uniformly compactly supported and when M satisfies the Gaussian estimate (5.60), the measures μ_k are Gaussian as in §2.12.2 with uniform constants.

These measures can be used to compute the following conditional expectations with respect to the σ-fields $\mathscr{F}_n$ generated by $x(1),\ldots,x(n)$. By (5.26) we have for any real continuous function F on V,

$$\begin{aligned}\mathbf{E}(F(v(n{+}1))//\mathscr{F}_n) &= \mathbf{E}_{x(n)}(F(v(n)+D_n)) \\ &= \int F(v+\xi)\mu_k(d\xi) \\ &= F * \check{\mu}_k(v); \quad x(n) = (v,k),\ \check{\mu}_k(\xi) = \mu_k(-\xi),\end{aligned} \tag{5A.20}$$

provided that the integrals involved are absolutely convergent. (For legibility, in (5A.20) we have set $v = v(n)$, $k = k(n)$.) This is certainly the case for compactly supported measures and in the Gaussian case, provided that F does not grow too fast at infinity.

By setting, in (5A.20), $F(t) = t_i$ the coordinate functions, it follows that the process $v(1), v(2), \ldots$ is a martingale if and only if all the measures μ_k, for $k \in K$, are centred and satisfy $\int u\, d\mu_k(u) = 0$. An equivalent way of saying this is that the coordinates $(v_1, \ldots, v_d) \in V$ considered as functions on X are harmonic functions with respect to the Markov chain (cf. Williams, 1991, Chung, 1982).

Similarly, for a more general function F the process $F(v(n))$ is a submartingale as long as F is μ-subharmonic for all the measures $\check{\mu}_k$, that is,

$$F(v) \leqslant F * \check{\mu}_k(v); \quad v \in V,\ k \in K. \tag{5A.21}$$

In what follows we shall consider a Euclidean principal bundle and aim to prove the gambler's ruin estimate in the compactly supported case of §5.5.1 and in the Gaussian case of (5.60) and (5.61). We shall consider these two cases simultaneously. We shall make the following assumptions.

Harmonicity assumption *We shall assume that the coordinate functions $v_1, \ldots, v_d$ on X as defined above are harmonic for the Markov chain.*

Assertion *Under these assumptions we shall show that the gambler's ruin estimate* (5.28) *and its Gaussian analogue in* (5.60) *and* (5.61) *both hold good.*

Before we give the proof, we need to observe that from the above assumptions and the definition (5A.19) it follows from (5.29) (see also Remark 5.3) that there exists $a > 0$ for which (5A.1) holds for the measures μ_k, $k \in K$, uniformly in k.

Exercise Verify this.

In the rest of this subsection we shall use the subharmonic functions that we have constructed in §5A.1 to prove the above assertion. Then in §5A.3 below, we shall show that it is possible to modify the original coordinates of V so that the new coordinates are harmonic functions. The notation of §5A.1 will be preserved.

The function $F = F_{v,k}$ has been defined on $\Omega = [\theta < \theta_0]$ and for large enough B we write $\Omega_B = \Omega \cap [F > B]$ and define $\Phi = F$ on Ω_B and $\equiv B$ outside and this is a globally μ-subharmonic function for all the measures $\mu = \mu_k$, $k \in K$ of (5A.19).

For the proof of our assertion we shall use the same argument as in the appendix of Chapter 3. We use the starting probability for the Markov process with $x(0) = (\rho I, e_K) = H$, $\rho I = (\rho, 0, 0, \ldots, 0)$ and e_K some reference point, say the identity of the group K. We denote $\tau = \inf[j;\ x(j) \notin \Omega_B \times K]$. As in (3A.14) by the submartingale property we have

$$\begin{aligned}\Phi(H) &\leqslant \mathbf{E}(\Phi(x(n \wedge \tau))) \\ &\leqslant \mathbf{E}([\tau > n]\,\Phi(x(n))) + B \\ &\leqslant (\mathbb{P}[\tau > n])^{1/2}(\mathbf{E}(\Phi^2(x(n))))^{1/2} + B. \end{aligned} \tag{5A.22}$$

By (5A.19) and (5A.20) we also have

$$\begin{aligned}\Phi^2(x(n)) &\leqslant Cn^{2\nu}\sum_{j=1}^{n} D_j^{2\nu}; \\ \mathbf{E}(D_j^{2\nu}) &= \mathbf{E}\mathbf{E}(D_j^{2\nu}//x(j)) \leqslant C\sup_{k\in K}\int |x|^{2\nu}\,d\mu_k(x). \end{aligned} \tag{5A.23}$$

This, together with the conditions imposed on the measures, implies that the cofactor of $\mathbb{P}(\cdot)$ on the right-hand side of (5A.22) is $O(n^{\nu+1})$. If now ρ in the choice of H is large enough, the inequality (5A.22) implies the required estimate

$$\mathbb{P}(\tau > n) \geqslant cn^{-c}; \quad n \geqslant 1, \tag{5A.24}$$

for appropriate constants. From this the gambler's ruin estimate (5.28) and its Gaussian variant of §5.10.1 follows. To see this we observe that by the conditions on the operator $\widehat{T}$ of §5.5 or §5.11.1, for arbitrary ε_0 there exists n_0 such that $\mathbb{P}[|x(n_0) - H| < \varepsilon_0] > 0$. We can then condition on this event as in §3A.1 and our assertion follows.

5A.3 The change of coordinates on a Euclidean bundle and the correctors

Let $X = V \times K$, $V = \mathbb{R}^d$ be a Euclidean bundle and $\widehat{T}$ be as in §5.5 with the coordinates (v,k), $v = (v_1,\dots,v_d) \in V$ and let $\chi(k) = (\chi_1(k),\dots,\chi_d(k)) \in V$, $k \in K$ be some continuous section. We can then perform the change of coordinates and set $u_j = v_j - \chi_j(k)$, $j = 1,2,\dots,d$, $k \in K$. This induces a bijective mapping

$$X \ni (v,k) \leftrightarrow (u = v - \chi, k) \in X \tag{5A.25}$$

that commutes with left translation by V. We shall now introduce a notion that is basic in homogenisation theory (cf. Jikov et al., 1991) and also for random walks in an inhomogeneous environment; see Varopoulos (2001, §6) for more on this circle of ideas and the way they are used in the present context.

Definition We say that $\chi_1, \ldots, \chi_d$ are *correctors* for the operator $\widehat{T}$ if the new coordinates $u_1, \ldots, u_d$ are harmonic functions for $\widehat{T}$, that is, $\widehat{T} u_j = u_j$.

We shall presently show that for M as in §5.5.1 that is either compactly supported or satisfies the Gaussian estimates (5.60), such correctors always exist. Once this is done the gambler's ruin estimate follows in full generality because we can use these new harmonic coordinates in the previous subsection and observe that $|u_j - v_j| \leqslant C,\ j = 1, \ldots, d$.

The Fredholm alternative and the existence of the correctors We shall use the fact that $\widehat{T}$ is Markovian and denote

$$L(k,h) = \int_V M(\gamma, k, h)\, d\gamma; \quad k,\ h \in K \tag{5A.26}$$

and by $L : C(K) \to C(K)$ as in (4.10) the Markovian operator that is induced by the kernel (5A.26). The harmonicity of u (here we drop the indices for the coordinates) is equivalent to

$$\int M(u - \omega; h, k)(\omega - \chi(k))\, d\omega\, dk = u - \chi(h) \tag{5A.27}$$

and since $\widehat{T}$ is Markovian this is equivalent to

$$L\chi(h) - \chi(h) = \int M(\omega; h, k)\, \omega\, d\omega\, dk = \theta(h). \tag{5A.28}$$

It follows that to construct the correctors it suffices to solve the above equation in $L^2(K; dk)$ because if $\chi \in L^2$ is a solution of (5A.28) then by the continuity of the kernel $L(h,k)$, it follows that χ is automatically continuous. The advantage of using the Hilbert space L^2 is that the kernel L induces a compact operator on L^2 and we can use standard Fredholm theory (see Hörmander, 1985, §19.1) to obtain a solution of (5A.28) in L^2. But for this we must verify that in the Hilbert space we have

$$\theta \perp \operatorname{Ker}(I - L^*). \tag{5A.29}$$

Here $L = L^*$ and the invariant functions $Lf = f$, $f \in L^2$ have to be constant by the properties of $L(h,k)$ (i.e. harmonic functions are constant; to see this, as in the classical situation one considers the max f that is attained and then uses the irreducibility of the Markov chain of Remark 5.3). The orthogonality relation (5A.29) amounts therefore to

$$\int M(\gamma; h, k)\gamma\, d\gamma\, dh\, dk = 0,$$

which is again a consequence of the symmetry $M = M^*$ (see §5.5.1). This completes the proof.

5B Appendix: Proof of (5.74)

5B.1 The Hessian and preliminaries

Let F be some smooth real function on $\mathbb{R}^d$; we recall that for $x \in \mathbb{R}^d$ the Hessian $\mathrm{Hess}(F;x) = H_x$ is a symmetric bilinear form on T_x, the tangent space at x, which for the Euclidean basis $\frac{\partial}{\partial x_j}$ of T_x can be expressed in matrix form $(F_{ij}$, $i, j = 1, \ldots, d)$ with $F_{i,j} = \frac{\partial^2 F}{\partial x_i \partial x_j}$; that is, $H_x(\xi, \zeta) = F_{i,j}\xi_i\zeta_j$ with the summation convention.

For our purposes we shall use the Hessian in the Taylor series remainder term for $\mu \in \mathbb{P}(\mathbb{R}^d)$ some centred measure that admits second moment $\int |x|^2 \, d\mu(x) < +\infty$. If $F \in C^2$ with bounded second derivatives we then have

$$\begin{aligned} F * (\mu - \delta)(x) &= 2\sum_{i,j} \int_0^1 \int_{\mathbb{R}^d} (1-t) F_{i,j}(x+ty) y_i y_j \, d\mu(y) \, dt \\ &= 2\int (1-t) H_{x+ty}(y,y) \, d\mu(y) \, dt \\ &= \int_{|y|<A} \cdots \quad + r(A), \end{aligned} \tag{5B.1}$$

where the remainder term $r(A)$ tends to 0 as $A \to \infty$. Here δ is as usual the Dirac mass at 0 (see Hörmander, 1983, equation (1.1.7)′).

We shall now define $F(x) = \psi(r)$ some radial function that satisfies $\psi \equiv 1$ for $|x| = r < 1/4$, $\psi \equiv 0$ for $r > 1$, $\psi > 0$ smooth for $r < 1$ and specified to be $\psi(r) = (1-r)^2$ for $1/2 < r < 1$.

If we denote the radial field by $\rho = \frac{x}{|x|}$ (which is $\frac{\partial}{\partial r}$ for the polar coordinates of §5A.1), we shall need to compute and verify that H, the Hessian of $F(x)$, satisfies

$$\begin{aligned} &H_x(\zeta, \xi) = 2\langle \rho, \xi \rangle \langle \rho, \zeta \rangle + B(\xi, \zeta), \\ &|B(\xi, \zeta)| \leqslant C(1-r)|\xi| \, |\zeta|; \quad \xi, \zeta \in T_x, \ r < 1, \end{aligned} \tag{5B.2}$$

for a numerical constant C.

Remark The Hessian $F \to \mathrm{Hess}(F) = D^2F$ acts as a second derivative under composition with $f\colon \mathbb{R} \to \mathbb{R}$ and we have $D^2(f \circ F) = f'' \, dF \otimes dF + f' \, D^2F$. We then apply this successively: first to $F_0(x) = r^2 = \sum x_j^2$ where $\frac{1}{2}\,\mathrm{Hess}(F_0) =$ the identity matrix, then to $r = \sqrt{F_0}$ and finally to $F = \psi(r)$. Alternatively, the computation of the Hessian of $(1-r^2)^2 = \psi(r)(1+r)^2$ is even simpler and from this (5B.2) easily follows. If you are a fan of differential geometry, in Greene and Wu (1979) you will find a general version of (5B.2).

In what follows, abusing notation, denoting $F(x)$ by $\psi(x)$ we shall scale ψ and define $\psi_M(x) = \psi(\frac{x}{M})$; it is then clear that $\mathrm{Hess}(\psi_M;x) = M^{-2}\,\mathrm{Hess}(\psi; \frac{x}{M})$

$= M^{-2}H_{\frac{x}{M}}$ and we shall combine this with the scaled version of (5B.1) that says, explicitly,

$$M^2(\mu-\delta)*\psi_M = \int_{|y|\leqslant A}(1-t)H_{\frac{x+ty}{M}}(y,y)\,d\mu(y)\,dt + r(A), \tag{5B.3}$$

where in the integral we have the Hessian of ψ at the point $\frac{x+ty}{M}$. What is, however, relevant there is that in (5B.3), $r(A)\to 0$ uniformly in $M\geqslant 1$ because this depends only on the second moment condition of μ; moreover, since for the proof of (5B.1) we need $F\in C^2$, we first prove (5B.3) for a modified function: that is, $(1-r)^\alpha$ in $r<1$ near $r=1$ for $\alpha>2$. Then we let $\alpha\to 2$.

5B.2 The subharmonic functions

Here we shall consider centred probability measures $\mu\in\mathbb{P}(\mathbb{R}^d)$ that admit a second moment and satisfy the conditions (5A.2) for some $a>0$.

The key fact that will be verified is that for all $a>0$ there exist ε_0, C_0 such that if $0<\varepsilon<\varepsilon_0$ and if M is such that $(1-\varepsilon)M\geqslant C_0$ then

$$\mu*\psi_M(x)\geqslant\psi_M(x); \quad |x|>(1-\varepsilon)M. \tag{5B.4}$$

We choose ε small and $M, A, M\geqslant 100A$ large, and observe that the range in which (5B.4) needs proving is $(1-\varepsilon)M\leqslant|x|\leqslant M$. By (5B.2), (5B.3) there then exists $C>0$ such that

$$\left.\begin{aligned} &M^2(\mu-\delta)*\psi_M(x)\\ &\quad = 2\int_{\substack{|x+ty|<M\\|y|<A}}(1-t)\left[\left\langle\frac{x+ty}{|x+ty|},y\right\rangle^2 + R|y|^2\right]d\mu(y)\,dt + r(A),\\ &|R|\leqslant C\left(1-\frac{|x+ty|}{M}\right)\leqslant C\left(\varepsilon+\frac{A}{M}\right), \end{aligned}\right\} \tag{5B.5}$$

and $\left|\frac{x+ty}{|x+ty|}-\frac{x}{|x|}\right|$ in the integral is $\left|\frac{\mathbf{1}+\tau}{|\mathbf{1}+\tau|}-\mathbf{1}\right|$ where $\mathbf{1}=\frac{x}{|x|}$ is a vector of length 1 and $\tau = t\frac{y}{|x|}$ and thus can be estimated by $|\tau|\leqslant 2\frac{A}{M}$ and is small. The conclusion is that

$$\left.\begin{aligned} &M^2(\mu-\delta)*\psi_M(x) = 2I(x) + \text{ error},\\ &I(x) = \int_{\substack{|x+ty|<M\\|y|<A}}(1-t)\left\langle\frac{x}{|x|},y\right\rangle^2 d\mu(y)\,dt,\\ &|\text{error}|\lesssim \varepsilon+\frac{A}{M}+r(A). \end{aligned}\right\} \tag{5B.6}$$

What remains to be seen is that in our range $(1-\varepsilon)M\leqslant|x|\leqslant M$ we can guarantee that there exists $c_0>0$ such that $I(x)>c_0$ for any choice of ε, M, A, provided that ε is small enough and M, A are large enough. To see this, by rotational symmetry we shall, as we may, assume that $x = (x_1, 0, \dots, 0)$. Then if $a > 0$ is given and μ satisfies (5A.1) we claim that we can choose some A large and then M can be chosen in terms of A large enough so that we have

$$I(x) \gtrsim \int_{\substack{y_1<0\\ |y|<A}} y_1^2 \, d\mu(y) \gtrsim c_0(a), \tag{5B.7}$$

where $c_0(a) > 0$ depends on a. This is of course a purely geometric property that follows from condition (5A.1) for centred measures that admit a high enough moment. A picture should be drawn by the reader at this point. But to illustrate the issue, suppose that $\operatorname{supp}\mu$ is compact. Then M is chosen so large that even in the worst case when $x_1 = M$, the integration region of $I(x)$ gets very close to $[y_1 < 0] \cap [|y| < A]$ and in fact contains $D_\eta = [y_1 < \eta] \cap [|y| < A]$ for an arbitrary preassigned η. In that case, if η is small enough, condition (5A.1) implies that

$$\int_{D_\eta} d\mu(y) \geqslant c(a) > 0 \quad \text{and therefore} \quad \int_{D_\eta} y_1^2 d\mu(y) \geqslant \eta^2 c(a), \tag{5B.8}$$

for a constant that depends on a and the support of the measure. We use for this the fact that μ is centred. If the support of μ is not compact the argument is a trifle more involved. But we shall skip the details.

If we combine this lower estimate $I(x) \geqslant c_0$ with (5B.6), our assertion (5B.4) follows.

We shall now suppose that M and ε have been chosen so that (5B.4) holds. On the other hand, for every $x \in \mathbb{R}^d$ we have

$$|(\mu - \delta) * \psi_M(x)| \leqslant CM^{-2}; \quad x \in \mathbb{R}^d, \tag{5B.9}$$

for some C that depends only on ψ. In the complementary region from the range of (5B.4) we have therefore

$$(\mu - \delta) * \psi_M(x) \geqslant -c_1 M^{-2} \psi_M(x); \quad |x| \leqslant (1-\varepsilon)M, \tag{5B.10}$$

where c_1 depends only on ψ and ε but not on M. Combining (5B.4) and (5B.10) we finally conclude that (5B.10) holds for all $x \in \mathbb{R}^d$.

What we have done in this subsection can be reformulated as follows.

Lemma *Let $a, B > 0$ be given; then there exist $M_0, c_1 > 0$ such that for every $\mu \in \mathbb{P}(\mathbb{R}^d)$ that is centred and satisfies* (5A.1) *and $\int |x|^2 d\mu(x) \leqslant B$ we have*

$$\mu * \psi_M(x) \geqslant e^{-c_1 M^{-2}} \psi_M(x); \quad x \in \mathbb{R}^d,\ M \geqslant M_0. \tag{5B.11}$$

5B.3 Proof of estimate (5.74)

We consider the Euclidean principal bundle $X = V \times K$, $V = \mathbb{R}^d$, as in §5.12, and $x(n) = (v(n), k(n)) \in X$, $n \geqslant 0$, the Markov chain generated by the left-invariant operator $\widehat{T}$ as in §§5.12, 5.10.1 and 5.5.1. We shall follow the same method and same notation as in §5A and in proving our estimate (5.74) we shall use the correctors of §5A.3 and assume, as we may, that for $x = (v,k) \in X$ the V component v is given by harmonic coordinates according to the definition in §5A.2. In what follows we shall denote by $F = \psi_M$ the function that satisfies (5B.11) as in the previous subsection. Furthermore, we shall put on the Markov chain the starting probability for which $x(0) = (0, e_K)$ with e_K the identity of K as in (5.74). We shall denote by $\mathscr{F}_n$ the σ-field on the path space that is generated by $x(0), x(1), \ldots, x(n)$.

Now, exactly as in the exercise of §5A.2, the measures $\mu_k \in \mathbb{P}(V)$, with $k \in K$, of the left-invariant operator $\widehat{T}$ are centred because of the harmonicity of the coordinates and, furthermore, in the case when $\widehat{T}$ is Gaussian, they satisfy $\int |x|^2 \, d\mu_k(x) < +\infty$. Condition (5A.1) is also satisfied for some $a > 0$ in both the Gaussian case and the compactly supported case for $\widehat{T}$. All this holds uniformly in $k \in K$. By §5B.2 the M in ψ_M can be chosen sufficiently large for (5B.11) to hold for μ_k uniformly in K.

Using (5B.11), all this can be reinterpreted in probabilistic language as follows:

$$\mathbf{E}(F(v(n+1))//\mathscr{F}_n) \geqslant e^{-c_0 M^{-2}} F(v(n)); \quad n \geqslant 0. \tag{5B.12}$$

The argument that follows is a standard one from discrete Markov and martingale theory (see Chung, 1982; Williams, 1991).

We shall introduce the stopping time $T = \inf[j;\ |v(j)| \geqslant M]$ and denote $v^*(n) = v(n \wedge T)$, $n \geqslant 0$. Then because F vanishes outside a ball of radius M we have

$$\begin{aligned} \mathbf{E}(F(v^*(n))//\mathscr{F}_{n-1}) &= \mathbf{E}(F(v(n))[T \geqslant n]//\mathscr{F}_{n-1}) \\ &= [T \geqslant n]\,\mathbf{E}(F(v(n)//\mathscr{F}_{n-1}); \quad n \geqslant 1, \end{aligned} \tag{5B.13}$$

for the indicator function $[T \geqslant n]$ that is in $\mathscr{F}_{n-1}$. This holds because $[T \geqslant n]$ is the complement of the union $\bigcup([T \leqslant j];\ j = 1, \ldots, n-1)$ and so we have $[T \geqslant n] \in \mathscr{F}_{n-1}$. Using (5B.12) we deduce therefore that (5B.13) is bounded from below by

$$e^{c_0 M^{-2}} [T \geqslant n] F(v(n-1)). \tag{5B.14}$$

But if we check separately for $[T \geqslant n]$ and $[T \leqslant n-1]$ the fact that $F(x) = 0$

for $|x| \geqslant M$ implies that

$$[T \geqslant n]F(v(n-1)) = F(v^*(n-1)), \tag{5B.15}$$

and if we combine (5B.13), (5B.14) and (5B.15), we finally conclude that

$$\mathbf{E}(F(v^*(n))//\mathscr{F}_{n-1}) \geqslant e^{c_0 M^{-2}} F(v^*(n-1)); \quad n \geqslant 1.$$

After iteration this gives

$$\mathbf{E}F(v^*(n)) \geqslant \exp\Big(-c_0 \frac{n}{M^2}\Big), \tag{5B.16}$$

since $F(0) = 1$. On the other hand, since $0 \leqslant F \leqslant 1$, by the definition of T and ψ_M we have

$$\mathbf{E}F(v^*(n)) \leqslant \mathbb{P}[T \geqslant n] = \mathbb{P}[\sup|v(j)| < M;\ 1 \leqslant j < n]. \tag{5B.17}$$

If we combine (5B.16) and (5B.17) the required estimate (5.74) follows.

Exercise (in elementary probability theory) In the above we have tacitly used the fact that $T < \infty$ a.s. This is a consequence of the well-known zero–one law. By inserting the event $[T < \infty]$ in the appropriate places of the argument, the reader is invited to avoid the use of this fact.

A final word on this appendix

The results in this appendix would have appeared to be less ad hoc if we had chosen to treat continuous time processes rather than random walks (cf. §E.4 for more on this). The existence of the correctors, in particular in §5A.3, are part of a general construction in homogenisation theory (see Jikov et al., 1991). To a certain extent, that is the point of view that was adopted in the earlier papers in the subject (see Varopoulos, 1996b; in this paper all the ideas of Chapters 4 and 5 are already in place but they are presented in a way that is not always reader friendly).

Note finally that if the proof of the full thrust of the NB-condition of §1.3 (i.e. the generalisation of the local central theorem of (3.3) as explained in §4.2) is ever to see the light of the day we must first refine the ideas of this appendix and of the homogenisation theory involved, to the ‘bitter end’. This explains my pessimism about it happening anytime soon. The reader could return to the end of Chapter 3 and find more references in that direction.

6

Other Classes of Locally Compact Groups

Overview of Chapter 6

In this chapter we shall examine briefly how the B–NB classification extends to other classes of locally compact groups.

If we restrict ourselves to connected or compact groups this is easy to do (see §§6.1–6.2 below). For this we use the well-known fact that every connected group can be appropriately approximated by a Lie group (see Montgomery and Zippin, 1955). On the other hand, among totally disconnected groups, there is a class of 'Lie groups' over a totally disconnected locally compact infinite field K (e.g. K the field of p-adic numbers $\mathbb{Q}_p$; see Weil, 1995; Cassels, 1986) where our classification should also work.

Let us start with an example and consider the group of affine motions on K: $x \to ax + b$ with $a \in K^* = K\backslash\{0\}$ the multiplicative group of K (this is denoted by $K^\times$ in Weil, 1995). This group generalises the example in §2.2 and can be written as $K \leftthreetimes K^*$. In our classification below this will turn out to be an NC-group. Similar examples, as in §2.2, can be given by other semidirect products of the form $(K \oplus K) \leftthreetimes K^*$, say, and so on. One should, however, note that the additive group of K is *not* compactly generated. All these are examples of affine algebraic groups over K. (We shall recall the definition of this notion in §6.3 below.)

For algebraic groups a formal notion of connectedness exists as provided by algebraic geometry. The groups that we shall be considering will be connected in this sense although they are totally disconnected for the locally compact topology. For these (connected) algebraic groups it seems realistic to think that a B–NB classification exists and that it can be expressed in equivalent terms:

$$\text{analytic} \leftrightarrows \text{algebraic} \leftrightarrows \text{geometric},$$

as explained in Chapter 1.

Using the methods of Chapters 2 and 3 there are some special cases of such groups that can be dealt with very easily. This will be done in §§6.4–6.8 below. More explicitly, §§6.4–6.8 consist of a series of indications of how to navigate the classical work Chevalley (1951, 1955) and how to adapt the methods of Chapters 2 and 3 in the context of some of the algebraic groups of that work. In writing these sections we have taken the point of view that only the more dedicated readers, especially those that are interested in p-adic groups, will be reading this chapter, and for these readers the indications that we give should be good enough. Alternatively, filling in the details in the material of this chapter could serve as a motivation for learning more on locally compact fields and basic algebraic geometry. The author of this book confesses that the little he knows about these beautiful subjects was learned while writing this chapter!

To complete the classification in general for all algebraic groups (over any locally compact field), although plausible, is likely to be a difficult program. In §6.9 below we indicate some of the concrete difficulties that one has to face.

It is interesting to see how all this fits with the general scheme that we described in the overviews of Chapters 2 and 3. In terms of this it is (iii), the third part of the scheme, that causes problems here. In other words, we have the results for a special class of groups but the passage to the general case remains problematic.

6.1 Connected Locally Compact Groups

Let G be some general locally compact group and assume that G is connected. By a basic result in the area (Montgomery and Zippin, 1955, IV.6) there exists then $H \subset G$ some compact normal subgroup such that $G/H = \check{G}$ is a connected real Lie group. Let $\pi : G \to \check{G}$ be the canonical projection and for any $\mu \in \mathbb{P}(G)$ let us as usual denote by $\check{\mu} = \check{\pi}(\mu)$ the direct image of μ by π. As we saw in §2.5.2, if μ satisfies conditions (i), (ii) and (iii) of §2.4.1 on G, the measure $\check{\mu}$ satisfies the same conditions on $\check{G}$ and furthermore, by (2.17), $\mu^{*n}(e)$ and $\check{\mu}^{*n}(e)$ behave identically as $n \to \infty$. The classification B–NB of §1.3 for the group G follows therefore from the classification of $\check{G}$.

Historical note: The above structure theorem is part of the program that led to the solutions of Hilbert's fifth problem (see Montgomery and Zippin, 1955).

6.2 Compact Groups and a Generalisation

Let G be a compact group, connected or not. Then the NB-condition holds in the very strong sense that if $\mu \in \mathbb{P}(G)$ is symmetric and satisfies conditions (i), (ii) and (iii) of §2.4.1, then, for $\mu^{*n} = \phi_n(g)\, dg$ and the normalised Haar measures dg, we have that the ϕ_n converge uniformly for some continuous function ϕ with $\phi(e) \neq 0$. The proof of this easy fact is entirely alien to the spirit of the book; therefore we shall content ourselves with helping the reader to reconstruct a proof, if they so wish, in the exercise below.

By combining the compact and the connected cases we can consider G_0 to be the connected component of a general locally compact group and then under the assumption that G/G_0 is compact, G has the same B–NB classification as G_0. The proof of this is a straightforward use of the methods of Chapters 2–5 together with the above result on compact groups.

Exercise Fill in the details in what follows. We can certainly extract subsequences of the μ^{*n} that converge weakly. These all converge to the same measure $\nu = \phi\, dg$, and ϕ is continuous. The continuity of ϕ and the uniform convergence follow because the ϕ_n have a common modulus of continuity. For the uniqueness of the limit we use first a direct argument in the same spirit as we did in §5.1.1 to show that if $f * \mu = -f$ then $f = 0$. This combined with the spectral decomposition of the symmetric operator T_μ of §3.1 implies that T_μ^n (as $n \to \infty$) converges to the projection on the eigenspace $[f \in L^2;\ Tf = f]$. This does it. Since the limit is a symmetric measure and $\nu * \nu = \nu$, it also follows that $\phi(e) \neq 0$. It is of course also easy to see that ϕ is the normalised characteristic function of the open subgroup generated by the support of μ. This is not the correct proof since, for instance, it does not work for measures that are not symmetric while the same fact does hold. But for our purposes here it is good enough. For more, see Kawada and Ito (1940).

What does lie in the spirit of the book is the generalisation of connectedness as supplied by algebraic geometry and which is used in the theory of algebraic groups. The possibility of extending the B–NB classification to such groups will be the subject of the remainder of this chapter.

6.3 On a Class of Locally Compact Groups

Let K be some locally compact non-discrete field. Such fields are completely classified, and when, in addition, they are of characteristic zero, they are exactly $\mathbb{R}$, $\mathbb{C}$ and the fields that are finite extensions of $\mathbb{Q}_p$, the field of the p-adic numbers. In that last case they are assigned with a non-Archimedean norm (i.e. $|x+y| \leqslant \max(|x|,|y|)$). More such fields exist but then the characteristic is positive (see Weil, 1995, Chapter 1 and, for a more thorough exposition, Cassels, 1986).

Let V be some finite-dimensional vector space over K and let $G \subseteq \mathrm{GL}(V)$ be some irreducible algebraic group (see Chevalley, 1951, §II.3). This means that there exists $\mathscr{I}$, a prime ideal of the ring of polynomial functions on the vector space of matrices on V over K, such that $g \in G$ if and only if $p(g) = 0$ for all $p \in \mathscr{I}$. The fact that $\mathscr{I}$ is prime reflects the fact that G cannot be decomposed into two subsets that are closed in the above sense (i.e. $G = A_1 \cup A_2$ where A_1, A_2 are the zeros of some polynomial ideal). This is a notion of connectedness and although these groups may be disconnected for the topology induced by K and V, they *are* connected in the sense of algebraic geometry. When $K = \mathbb{R}, \mathbb{C}$ the (topologically) connected component G_0 of an irreducible algebraic group G is a real connected Lie group and furthermore G/G_0 is finite and we have nothing new (Whitney, 1958; Varadarajan, 1974, §2.1). On the other hand, when K is non-Archimedean we obtain a new class of locally compact groups that is not necessarily compactly generated. (An example of this is the additive group of K itself. By the non-Archimedean property, every compact subset is contained in some compact subgroup. On the other hand, K can be realised as a group of 2×2 'unipotent' matrices.) Nonetheless, the classification of §1.3 has a chance to hold for those groups. For the rest of this chapter we shall concentrate on those groups that are topologically totally disconnected.

The serious problem in extending the B–NB classification to these groups is that we do not in general have at our disposal the necessary ready-made structure theorems (e.g. the Levi decomposition (see §2.1.2) and Iwasawa decomposition (see §4.1 and Appendix A)) that were essential in the proofs in Chapters 2–5. New methods have to be devised therefore and the problem is not trivial. There are special cases, however, where adequate structure theorems exist. We shall examine these special groups in the next subsection and then proceed to give the B–NB classification for these groups. For the proofs, it suffices to adapt the proofs of the real case and the task turns out to be quite easy.

The remainder of this chapter should be considered as a series of exercises that help to illustrate the methods of Chapters 2–3. The reader could of course skip all this. It is also true, however, that interesting open problems exist in this area (see §6.9 below) but these would involve the modern theory of algebraic groups in a serious way and in particular the highly developed theory of reductive groups (see Borel and Tits, 1965).

6.4 A Review of Some Results from Algebraic Groups

6.4.1 General definitions

We shall follow the original historical reference (Chevalley, 1951, 1955) because we feel that for our needs it is by far the most accessible to those readers who have had no exposure to modern algebraic geometry. The only problem with this reference is that like all good old-fashioned books, one may find it difficult to navigate and find the exact statements that one needs. For this reason an effort has been made below to give precisely the necessary cross references. There are at least two other, more recent books on algebraic groups (Onischik and Vinberg, 1988; Humphreys, 1975) into which the reader can plunge without previous knowledge of algebraic geometry. Unfortunately the theory is developed there over algebraically closed fields and to go back to the original field, say $\mathbb{Q}_p$, is not a meagre affair (e.g. Humphreys, 1975, Chapter XII); see the final remark in §6.9. In fact, the book by Chevalley is the only one, as far as I know, that treats the case of general fields directly (i.e. without first passing through the algebraic closure!). In a final analysis, this was the reason that made me disregard all the other inconveniences involved and use the book as the main reference in this chapter.

Let $G \subset \mathrm{GL}(V)$ be some irreducible algebraic group and in what follows the characteristic of the field will be assumed to be 0. The Lie algebra $\mathfrak{g}$ of G can then be defined and is a subalgebra of $E = \mathfrak{gl}(V)$, the space of all linear transformations on V, that is, $E = \mathrm{End}_K(V)$ under the bracket operation $[a,b] = ab - ab \in E$, with $a, b \in E$. There exists then a one-to-one correspondence between irreducible algebraic groups G and a class of subalgebras (not all the subalgebras in general) of $\mathfrak{gl}(V)$. That correspondence respects inclusions. The characteristic 0 is essential for this fact (see Chevalley, 1951, §II.8, p. 146).

Let $G \subset \mathrm{GL}(V)$ be as above and let $\mathfrak{g}$ be its Lie algebra; we assume that $\mathfrak{g}$ is *solvable*. Note that we shall be using the terms *soluble* and *solvable* interchangeably. The definition of solubility is as in (Chevalley, 1955, §V.1), and this is exactly as the definition that we gave in §2.1. Let $\mathfrak{n} \subset \mathfrak{g}$ be the set of all the nilpotent transformations (i.e. $a^r = 0$ for some $r \geqslant 1$) that lie in $\mathfrak{g}$. Then $\mathfrak{n}$ is an ideal of $\mathfrak{g}$ (see Chevalley, 1955, §V.2.2) and $\mathrm{ad}\, x\colon\ \mathfrak{n} \to \mathfrak{n}$, with $x \in \mathfrak{n}$, is a nilpotent transformation (see Chevalley, 1955, §IV.4.2, Corollary 4 or Bourbaki, 1972, Chapter 1, §5.4). This means that $\mathfrak{n}$ is a nilpotent ideal in the sense of §2.1 (see Varadarajan, 1974, §3.5) and therefore $\mathfrak{n}$ is contained in the nilradical used in §2.1 but it need not be the whole nilradical (e.g. dimension $\mathfrak{g} = 1$ then the nilradical of §2.1 is the whole $\mathfrak{g}$, and $\mathfrak{n}$ as above could be 0).

6.4.2 Soluble groups

We shall restrict ourselves to the case when $\mathfrak{g}$ is a solvable algebra and recall the following basic structure theorem (Chevalley, 1955, §V.3.5, Proposition 21): G can be written as a semidirect product of two irreducible algebraic subgroups $N, A \subset G$, that is, N is normal in G, $G = NA$ and $N \cap A = \{e\}$. The subgroup A is Abelian and consists of semisimple transformations on V. Furthermore, the Lie algebra $\mathfrak{g}$ of G is the direct vector space sum of $\mathfrak{n}$ and $\mathfrak{a}$ the Lie algebras of N and A, and $\mathfrak{n}$ is exactly those elements of $\mathfrak{g}$ that are nilpotent transformations of V. Moreover, as already pointed out, $\mathfrak{n}$ is an ideal in $\mathfrak{g}$ for its Lie algebra structure. An additional accessible modern reference for the above is Onischik and Vinberg (1988, Chapter 6, no. 6), but this refers only to $K = \mathbb{C}$ and not even $K = \mathbb{R}$, let alone $\mathbb{Q}_p$.

(i) **The** Ad**-action** The reader who is not too familiar with algebraic group theory should note the following facts.

For any invertible $s \in \mathrm{GL}(V)$ we can consider $I_s = X \to sXs^{-1}$, with $X \in \mathrm{End}_K(V) = E$. This gives a representation of $\mathrm{GL}(V)$ on the space of endomorphisms of V. On the other hand, if V^* is the dual we can identify $E = V \otimes V^*$. From that definition it follows that if s is a semisimple transformation on V then I_s is semisimple as a linear transformation on E. This is elementary to verify but a proof is spelled out in Chevalley (1955, §IV.4.2 Cor. 1 and 3): the relevant pages are 31 and 79; see also Varadarajan (1974, §3.1) for the more general theory of replicas. What counts here is that if $g \in G$ then $I_g\mathfrak{g} \subset \mathfrak{g}$ (see Chevalley, 1951, §II.9, Prop. 6 and the few lines that follow it).

Using the characterisation of $\mathfrak{n}$ as nilpotent transformations of V, it follows that $I_g\mathfrak{n} \subset \mathfrak{n}$. Therefore we can conclude that if $a \in A$, the transformation $\mathrm{Ad}\,a\colon \mathfrak{n} \to \mathfrak{n}$ (i.e. $\xi \to a\xi a^{-1}$ with $\xi \in \mathfrak{n}$) is a semisimple transformation. The reason why the same notation Ad as before (see Varadarajan, 1974, §2.13) is used for this transformation is explained in Chevalley (1951, §II.9, Def. 2, Prop. 7) and Varadarajan (1974, §2.13.14).

(ii) **The semisimple factor** A The factor A in the semidirect product decomposition of G consists of commuting semisimple transformations in $\mathrm{GL}(V)$. This means that we can find some finite extension $\overline{K}$ of K such that for some appropriate basis of $\overline{V} = V \otimes_K \overline{K}$ all the matrices that represent the elements of A belong to $\overline{D} \subset \mathrm{GL}(\overline{V})$, the group of diagonal matrices for that basis. On the other hand, A is a closed subgroup of $\mathrm{GL}(V)$ for the locally compact topology and since also $\mathrm{GL}(V)$ is closed in $\mathrm{GL}(\overline{V})$ and $\overline{D}$ is closed in $\mathrm{GL}(\overline{V})$ we conclude that A is a closed subgroup of $\overline{D}$ always with respect to the locally compact topology.

The group $\overline{D} = \left(\overline{K}^*\right)^d$, where $d = \dim V$, is the Cartesian product of the multiplicative group $\overline{K}^* = \overline{K}\backslash\{0\}$ of the field. We shall now restrict the field K to be non-Archimedean; then $H = \{x \in \overline{K}^*;\ |x| = 1\}$ is an open compact subgroup and we have $\overline{K}^*/H \approxeq \mathbb{Z}$ (see Weil, 1995, §I.4 and Cassels, 1986). The subgroup H^d is therefore open and compact in $\overline{D}$ and $\overline{D}/H^d \approxeq \mathbb{Z}^d$. With the identification $A \subset \overline{D}$, the subgroup $A \cap H^d = A_0$ is open and compact, and $A/A_0 \approxeq \mathbb{Z}^m$ for some $0 \leqslant m \leqslant d$. It follows in particular that A is a compactly generated group of volume growth $\gamma(n) = \mathrm{O}(n^m)$; see §1.1.

6.4.3 The commutator subgroup

See Chevalley (1951, §II.14, Th. 15) for a proof of the following result.

Let G be some irreducible algebraic group and $\mathfrak{g}$ be its Lie algebra. Then the commutator subalgebra $[\mathfrak{g},\mathfrak{g}]$ corresponds to an irreducible algebraic group $G_2 \subset G$ and for all $x, y \in G$, $[x,y] = xyx^{-1}y^{-1} \in G_2$.

6.4.4 Nilpotent groups and the exponential mapping

See Chevalley (1955, §V.3.4) for a proof of the following result.

Let $x \in \mathfrak{gl}(V)$ be some nilpotent transformation. Then $\exp x = I + x + \sum \frac{x^p}{p!}$ *is a polynomial. Furthermore, if $\mathfrak{g} \subset \mathfrak{gl}(V)$ is a subalgebra consisting entirely of nilpotent transformations, the above mapping establishes a one-to-one correspondence* $\exp : \mathfrak{g} \to G$*, where G is an algebraic irreducible group whose Lie algebra is $\mathfrak{g}$.*

Going back to the solvable group $G = NA$ of §6.4.2 we have the analogue of the formula of (3.22) (see Varadarajan, 1974, §2.13.7):

$$\exp((\mathrm{Ad}\,a)\,\xi) = a\,(\exp \xi)\,a^{-1} = (\exp \xi)^a; \quad \xi \in \mathfrak{n},\ a \in A. \tag{6.1}$$

One final property of the bijection between $\mathfrak{n}$ and N given by the exponential mapping is that it takes the Haar measure on $\mathfrak{n}$ which is the Haar measure of the vector space K^d, with $d = \dim \mathfrak{n}$, to the Haar measure of N. This fact is important but no immediate use of it will be made; nonetheless, in Exercise 6.1 below we outline a proof.

The Baker–Campbell–Hausdorff formula, BCH (see Jacobson, 1962, V.4; Serre, 1965; Varadarajan, 1974, §2.15), gives additional information on the above exponential mapping $\exp\colon \mathfrak{n} \to \mathrm{GL}(V)$ with $\mathfrak{n}$ as in (6.1). To wit, assume that $s \geqslant 2$ is such that $\mathfrak{n}^s = 0$; then the product $\exp \xi_1 \exp \xi_2$, for $\xi_1, \xi_2 \in \mathfrak{n}$, can be expressed as the exp of a finite sum $\sum_{j<s} c_\alpha [\zeta_{\alpha_1}, [\zeta_{\alpha_2}, \ldots, \zeta_{\alpha_j}], \ldots]$ where $\alpha = (\alpha_1, \ldots, \alpha_j)$, with $\zeta_{\alpha_k} = \xi_1$ or ξ_2, and c_α are *rational* coefficients that can be computed explicitly (e.g. for $s = 3$, $c_1 = c_2 = 1$, $c_{1,2} = -c_{2,1} = 1/4$).

We recall that the characteristic of the field K is zero and we shall assume that the norm of the field satisfies the ultrametric inequality ($|x+y| \leqslant \max(|x|,|y|)$, e.g. the fields $\mathbb{Q}_p$). In that case the BCH formula implies not only that the above locally compact groups N are not compactly generated, but even that they are countable unions of compact open subgroups. More explicitly, let us write $N(r) = \exp[\xi;\ |\xi|_{\mathfrak{n}} \leqslant r]$ for the image of the r-ball in $\mathfrak{n}$ for some norm $|\ |_{\mathfrak{n}}$ on $\mathfrak{n}$ that has been fixed. These neighbourhoods of the identity in N have the following property. For all $r \geqslant 1$ there exists $C = C(r)$ such that $N^q(r) = N(r)\cdots N(r) \subset N(C)$ for any $q = 1,\ldots$. The point here of course is that C is independent of q. The dependence on r is in fact polynomial: $C(r) \leqslant C_1 r^c$, and where the constants also depend on s.

Exercise To see this, make repeated use of the BCH formula and use $\mathfrak{n}^s = \{0\}$ to deduce that any element in $N^q(r)$ can be written as $\exp(\xi)$ where $\xi \in \mathfrak{n}$ is a sum of elements of the form

$$c_\alpha c_\beta \cdots c_\gamma \left[\zeta_{k_1}\left[\cdots \zeta_{k_j}\right]\cdots\right] = \eta \quad \text{with } |\zeta_k| \leqslant r.$$

For these brackets, we clearly have $\left|[\zeta_{k_1}[\cdots \zeta_{k_j}]\cdots]\right| \leqslant Cr^s$. Now the coefficients c_α, c_β are just the coefficients of the BCH formula; also, clearly the length of the product $c_\alpha c_\beta \cdots c_\gamma$ is at most s and therefore the coefficient $c_\alpha c_\beta \cdots c_\gamma$ is a rational number a/b with $|b| \leqslant B$; that is, the denominator is bounded by some constant B that is given by the product of at most s among the denominators of the BCH coefficients c_α.

The bottom line is that $|\eta| \leqslant C_1 r^c$ and the ultrametric inequality does the rest.

Exercise 6.1 We can use the BCH formula to prove that the exp mapping preserves the Haar measure as explained above.

For the case $K = \mathbb{R}$ this is an immediate consequence of Varadarajan (1974, §2.14). It is also possible that when K is non-Archimedean the same argument based on calculus works also. Here we propose a different proof. The exp identifies $K^d \ni (t_1,\ldots,t_d) \to \exp(t_1\xi_1 + \cdots + t_d\xi_d)$ where $\xi_1,\ldots,\xi_d$ is a basis of $\mathfrak{n}$. We can choose that basis so that with $V_j = \mathrm{Vect}(\xi_1,\ldots,\xi_j)$ the space V_j is an ideal in V_{j+1}. When this is done a second exponential mapping can be defined (or rather exponential coordinates of the second kind can be defined (see Chevalley, 1955, §V.3.4, Prop. 17)): $K^d \ni (s_1,\ldots,s_d) \xrightarrow{E} \exp(s_1\xi_1)\cdots\exp(s_d\xi_d)$, which also is a bijection, and here it is evident that E preserves the Haar measures. To see this let $N_j = E(V_j)$; then $N_j \lhd N_{j+1}$ are closed subgroups and we can use the standard facts about the Haar measures of G, H and G/H (Bourbaki, 1963, Chapter 7); at the end we use induction on d.

To prove our original assertion on the 'coordinates of the first kind' it suffices therefore to show that the basis $\xi_1,\dots,\xi_d$ can be chosen so that $t_j = s_j + F_j(s_{j+1},\dots,s_d)$ and because of the shape of the coordinate transformation $(s_1,\dots) \to (t_1,\dots)$ we can make repeated use of Fubini and conclude. To see that this is possible one uses a central series $\mathfrak{n}_1 \subset \mathfrak{n}_2 \subset \cdots \subset \mathfrak{n}$ with $[\mathfrak{n}_q,\mathfrak{n}] \subset \mathfrak{n}_{q-1}$. Then let $\zeta = [\xi_{j_1}[\xi_{j_2}\cdots[\xi_{j_k}\cdots]\cdots]]$ be one of the brackets of length $k \geqslant 2$ that occur in the repeated use of the BCH formula that is used to work out $\exp(s_1\xi_1)\exp(s_2\xi_2)\cdots$. If one basis element, say ξ_{j_2}, belongs to $\mathfrak{n}_q$ then clearly $\zeta \in \mathfrak{n}_{q-1}$. All it takes therefore is to adapt the basis $\xi_1,\dots$ with the central series in the obvious way so that if one of the above brackets, $\zeta = u_1\xi_1 + u_2\xi_2 + \cdots$, is such that $u_q \neq 0$ then in that bracket the ξ_{j_r}, with $r = 1,2,\dots$, involved have to have $j_r > q$.

6.5 The C–NC Classification for Solvable Algebraic Groups

6.5.1 The roots

Here $G = NA$ is as in §6.4.2 and the notation $\mathfrak{n}$ for the Lie algebra of N is preserved. By the semisimplicity of $\mathrm{Ad}\,a, a \in A$, in §6.4.4 we can find $\overline{K} \supset K$, a Galois finite extension of the field K such that for an appropriate basis of $\mathfrak{n} \otimes_K \overline{K}$ over $\overline{K}$, all the matrices become diagonal: $\mathrm{Ad}\,a = \mathrm{diag}(\omega_1(a),\dots,\omega_n(a))$, where $\omega_j : A \to \overline{K}^*$ are group homomorphisms. In this notation, $\overline{K}^*$ is the multiplicative group of the field on $\overline{K}\backslash\{0\}$. (For the finiteness of the field extension observe this: let $a_0 \in A$, $\lambda_0 \in \overline{K}$ (which is some field extension), and define $I_0 = [\xi \in \mathfrak{n} \otimes_K \overline{K};\ \mathrm{Ad}\,a_0\xi = \lambda_0\xi]$; then $\mathrm{Ad}\,aI_0 \subset I_0$, for $a \in A$.) Galois extension means as usual that the field automorphisms of $\overline{K}$ that stabilise each element of K stabilise nothing else. The group of these automorphisms is denoted by $\mathrm{Aut}[\overline{K} : K]$. When $K = \mathbb{R}$ we can of course take $\overline{K} = \mathbb{C}$, but for the non-Archimedean fields, $\overline{K}$ is not algebraically closed and in general, $\dim[\overline{K} : K] \geqslant 2$.

Similarly, we can diagonalise the induced action by $\mathrm{Ad}\,a$ on $\overline{W} = W \otimes_K \overline{K}$, where $W = \mathfrak{n}/[\mathfrak{n},\mathfrak{n}]$, where we use the same notation as in §2.1.1 but $\mathfrak{n}$ here is not necessarily the nilradical. The action on $\overline{W}$ becomes $\mathrm{diag}(\theta_1(a),\dots,\theta_t(a))$ and by Jordan–Hölder the set of $\theta_j : A \to \overline{K}^*$ is a subset $(\theta_1,\theta_2,\dots,\theta_t) \subset (\omega_1,\omega_2,\dots,\omega_n)$. The θ are called the roots of the action of A on N and the ω are called the composite roots of that action. The situation is entirely analogous to that in Chapter 2 where we compared the roots in §2.1 with the composite roots in §2.3.4. The Jordan–Hölder series that we use for the comparison is of course $\mathfrak{n} \otimes \overline{K} \supset \mathfrak{n}^2 \otimes \overline{K} \supset \cdots$. If we consider the induced action on the factors

$(\mathfrak{n}^p/\mathfrak{n}^{p+1}) \otimes \overline{K}$ we also see that each composite root ω_j can be written as a product of roots: $\theta_{\alpha_1}\theta_{\alpha_2}\cdots\theta_{\alpha_j}$.

6.5.2 The real roots and their classification

We shall define two sets $\mathscr{L}_W = (L_1,\dots,L_p)$ and $\mathscr{L}_\mathfrak{n} = (L_1,\dots,L_k)$ with the identification $\mathscr{L}_W \subset \mathscr{L}_\mathfrak{n}$ by $L_j = \log|\theta_j|$ or $\log|\omega_j|$ as the case might be, where $|\;|$ indicates the unique absolute value on $\overline{K}$ that extends that on K (Weil, 1995, Chapter 1). It is then clear that each $L_j \in \mathscr{L}_\mathfrak{n}$ can be written as a sum $L_j = L_{\alpha_1} + \cdots + L_{\alpha_j}$ with the $L_{\alpha_s} \in \mathscr{L}_W$. From this it follows that in the **classification** below we can take $\mathscr{L} = \mathscr{L}_W$ or $\mathscr{L}_\mathfrak{n}$ indiscriminately.

(C) We say that G is a C-group if there exist $\beta_j \geqslant 0$, $\sum \beta_j = 1$ such that $\sum \beta_j L_j = 0$ summed over $L_j \in \mathscr{L}$ and such that $\beta_{j_0} L_{j_0} \neq 0$ for some $j_0 = 1,2,\dots$.

(NC) We say that G is NC if for any choice of $\beta_j \geqslant 0$ such that $\sum \beta_j L_j = 0$, with $L_j \in \mathscr{L}$, we must have $\beta_j L_j = 0$ for all $j = 1,2,\dots$.

When $A = \{0\}$ the group reduces to N which with this definition is therefore NC. Note, however, that because of what was said in §6.4.4, this group is not in general compactly generated.

6.5.3 On the definition of the real roots

The subgroup $A_0 \subset A$ defined in §6.4.2(ii) is compact; it then follows that $L(A_0) = 0$ for all $L \in \mathscr{L}_\mathfrak{n}$ and that therefore $L = \widetilde{L} \circ p$ for a new homomorphism $\widetilde{L}\colon \mathbb{Z}^m \to \mathbb{R}$, where $p\colon A \to \mathbb{Z}^m$ denotes the canonical projection of §6.4.2(ii). Abusing notation, in what follows we shall drop the tilde (~) and denote indiscriminately $L = \widetilde{L}$.

6.5.4 The real root space decomposition

To fix ideas we shall consider the roots $(\theta_1,\dots,\theta_t)$ and the set of the corresponding 'real roots' $\mathscr{L}_W$. We can then decompose $\overline{W} = W_1 \oplus \cdots \oplus W_t$ where W_j is spanned by the eigenvector that corresponds to the root θ_j where here we admit the possibility that for two distinct indices $\theta_j = \theta_k$. If we sum together $W_L = W_{\alpha_1} \oplus \cdots$ all the subspaces for which $\log|\theta_{\alpha_1}| = \cdots = L$, we obtain the decomposition $\overline{W} = W_{L_1} \oplus \cdots \oplus W_{L_p}$ for the real roots $\mathscr{L}_W = (L_1,\dots,L_p)$. Just as in §2.1 where $K = \mathbb{R}$, each space is of the form $W_{L_j} = \widetilde{W}_j \otimes_K \overline{K}$ for some K-subspace $\widetilde{W}_j \subset W$ and we have the corresponding *real space decomposition*

$W = \widetilde{W}_1 \oplus \cdots \oplus \widetilde{W}_p$. This fact is easy but yet a trifle harder to see than the case $K = \mathbb{R}$ of §2.1. It is a consequence of the following lemma.

Lemma *Let E be some finite-dimensional vector space over K and let us consider $\overline{E} = E \otimes_K \overline{K}$ as a vector space over K. Every $\alpha \in \mathrm{Aut}[\overline{K} : K]$ then induces $\check{\alpha}$, a K-linear automorphism of $\overline{E}$. Let $E' \subset \overline{E}$ be a $\overline{K}$ subspace such that $\check{\alpha}(E') = E'$ for all $\alpha \in \mathrm{Aut}[\overline{K} : K]$; then there exists $\tilde{E} \subset E$ a K-subspace such that $E' = \tilde{E} \otimes_K \overline{K}$.*

Our assertion on the above real root space decomposition follows immediately from the lemma because for the absolute value on $\overline{K}$ and $\alpha \in \mathrm{Aut}[\overline{K} \colon K]$ we have $|\alpha(x)| = |x|$, $x \in \overline{K}$ (see Cassels, 1986, Chapter 7), and therefore the automorphisms in $\mathrm{Aut}[\overline{K} \colon K]$ permute among themselves the roots $\theta_{\alpha_1}, \ldots$ that correspond to the same real root $L \in \mathscr{L}_W$.

Proof of the lemma Unless $E' = \{0\}$ there exists $x = e_1 + \lambda_2 e_2 + \cdots \lambda_n e_n \in E'$ with $\lambda_j \in \overline{K}$ and $(e_1, \ldots, e_n)$, a basis of E over K.

Let $\bar{x} = \left(\sum \check{\alpha}(x) : \alpha \in \mathrm{Aut}\left[\overline{K} : K\right]\right)$; then $0 \neq \bar{x} \in E' \cap E$ because the extension is Galois. It follows that $E' = \bar{x}\overline{K} \oplus E'_1$ where $E'_1 = E' \cap E_1$ with $E_1 = (e_2\overline{K} + e_3\overline{K} + \cdots)$ simply because $\bar{x}\overline{K} \subset E'$. It follows that $\check{\alpha}(E'_1) = E'_1$ and we can use induction over the dimension. This completes the proof. □

6.6 Statement of the Theorems

6.6.1 Conditions on the measures and the groups

For the rest of the chapter the group G will be assumed to be an irreducible solvable algebraic group $G \subset \mathrm{GL}(V)$ where V is a d-dimensional vector space over the field K. The field is assumed to be of characteristic 0 and is locally compact with an absolute value $|\ |$. We have $G = NA$ as in §6.4.2 and all the previous notation will be preserved.

The theorems that will be proved for G are exactly the C- and NC-theorems of Chapters 2 and 3 for the group G when $K = \mathbb{R}$, and the aim here is to extend these results to the non-Archimedean case. For this reason, to fix ideas (unless otherwise stated) we shall assume that K is such a non-Archimedean field.

Conditions (i), (ii) and (iii) of §2.4.1 on the probability measures for non-Archimedean fields have to be modified because the groups G are not in general connected locally compact groups. To see what goes wrong, assume $\mu \in \mathbb{P}(G)$ has its support in $\pi^{-1}(A_0)$ where we recall that $\pi : G \to G/N$ and $A_0 \subset A$ is the compact open subgroup such that $p : A \to A/A_0 = \mathbb{Z}^m$. In that case all the real roots $L \in \mathscr{L}$ are identically zero on $\mathrm{supp}\,\mu^{*n}$ with $n \geqslant 1$, and these roots therefore do not have any influence on the behaviour of μ^{*n}. The conditions that we shall impose on the measures $\mu_1, \ldots$ of the theorems below are in fact exactly the same (i), (ii), (iii) of §2.4.1 but for the set $\Omega \subset G$ in condition (iii) we demand the following result.

(iii)′ Ω is a relatively compact symmetric neighbourhood of e and $\pi(\Omega)$ generates A.

Let us write $\psi = p \circ \pi \colon G \to \mathbb{Z}^m$ and $\check{\mu} = \check{\psi}(\mu)$. By definition, $\check{\mu}(0) > 0$, and for every $x \in \mathbb{Z}^m$ there exists q such that $\check{\mu}^q(x) > 0$.

Condition (iii)′ guarantees therefore that the random walk on $\mathbb{Z}^m$ controlled by $\check{\mu}$ behaves as it should. For instance, when μ is symmetric, we have $\check{\mu}^n(0) \sim n^{-m/2}$.

6.6.2 Statement of the theorems

Theorem 6.2 (C-theorem) *Let G and $\mu_1, \ldots \in \mathbb{P}(G)$ be as in §6.6.1 and let us assume that G is a* C-*group. As before, denote $\mu^n = \mu_1 * \cdots * \mu_n = \phi_n(g)\, d^r g$ and let $P \subset G$ be some compact subset. There exist then constants C, $c > 0$ such that*

$$\phi_n(g) \leqslant C \exp\left(-cn^{1/3}\right); \quad n \geqslant 2,\ g \in P. \tag{6.2}$$

This implies that

$$\mu^n(P) \leqslant C \exp\left(-cn^{1/3}\right); \quad n \geqslant 1 \tag{6.3}$$

and (6.3) is the analogue of (2.15). Conversely, of course, by conditions (i), (ii) and (iii)′ imposed on the measures, condition (6.3) for n implies condition (6.2) for $n+1$.

Exercise Verify this. It is elementary by the definition of convolution. One of the technical problems that we have to face with these groups is that we do not have at our disposal the Harnack estimate of §2.5 because these groups are not connected. Nonetheless, the second formulation (6.3) allows us to use the reduction of §2.5.3. What this amounts to is the following: let $\nu \in \mathbb{P}(G)$ be some compactly supported fixed measure and, with the notation of the C-theorem, let us set $\mu_n = \nu * \mu^n$. Then, if condition (6.3) is satisfied (for an arbitrary P) for the sequence of measures μ_n, then it is also satisfied for the original sequence μ^n. To see this, for any set P, we have $\mu_n(P) = \int \mu^n(x^{-1}P)\, d\nu(x)$ and observe that for any compact set P_1 we can find another compact set P such that $(\operatorname{supp} \nu)P_1 \subset P$. Then use this P in the integral.

Theorem 6.3 (NC-theorem) *Let G and $\mu \in \mathbb{P}(G)$ be as in §6.6.1 and let us assume that μ is symmetric (i.e. that $\mu(x^{-1}) = \mu(x)$) and that G is an* NC-*group. Then there exist constants $C, c > 0$, such that $\mu^{*n} = \phi_n(g)\, d^r g$ satisfies*

$$\phi_n(e) \geqslant Cn^{-c}; \quad n \geqslant 2. \tag{6.4}$$

The proofs will be given in the next two sections. These proofs are direct generalisations of those given in Chapters 2 and 3. In fact they are considerably simpler than those proofs because the group G is already a semidirect product of N with A and therefore the disintegration of the measures in §2.7.2 simplifies; more significantly, in the proof of the NC-theorem in Chapter 3, we do not need to use Cartan subalgebras and the overgroup $\widetilde{G} \to G$ of (3.14). The proof of (6.4) will be given in §6.7 only for $n = 2n_1$, an even integer. The usual passage to the odd integers using the Harnack principle of §2.5 does not work here and something else has to be done. Given, on the other hand, that these two theorems are essentially just illustrations of our methods on a special class of groups, we shall not elaborate further on that 'something else'.

The proofs in the next two sections will be written out for non-Archimedean fields but with only notational changes they also work for $K = \mathbb{R}, \mathbb{C}$; and for the case of algebraic groups they provide easier versions of the arguments of Chapters 2 and 3.

6.7 Proof of the NC-Theorem

In this section G and μ are as in the theorem and familiarity with the arguments and the notation of §3.5 will be essential; we shall use this notation throughout. We shall denote the random walk on G by $z(j) = s_j = x_1 \ldots x_j \in G$, where $x_j \in G$ are random variables with distribution $\mu \in \mathbb{P}(G)$. The proof of the criterion of §3.3.3 is very general and applies to the above Markov chain $z(j)$.

We shall write $x_j = m_j a_j$ with $m_j \in N$, $a_j = \dot{x}_j = \pi(x_j) \in A$ (for the projection $\pi : G \to G/N$) and $\dot{s}_j = \pi(s_j) = a_1 a_2 \ldots a_j \in A$; then we have, as in (2.21),

$$s_j = m_1 m_2^{\dot{s}_1} \cdots m_j^{\dot{s}_{j-1}} \dot{s}_j. \tag{6.5}$$

We shall fix once and for all some norm $|\ |_{\mathfrak{n}}$ on $\mathfrak{n}$ and, as in §6.4.4, define $\mathfrak{n}(r) = [m \in \mathfrak{n};\ |m| \leqslant r]$ and $N(r) = \exp \mathfrak{n}(r)$. Using the group distance of §1.1 on the compactly generated group A of §6.4.2(ii) we can also define $A(r) = [a \in A;\ |a|_A \leqslant r]$. As pointed out in §6.4.2(ii), we have for the corresponding Haar measures of these sets $|A(r)| \leqslant Cr^m$.

The sets on which the criterion of §3.3.3 will be used are $E_n = N(C)A(Cn^c)$, with $n \geqslant 1$, for appropriate constants.

The Haar measure on G is $d^r g = dm\, da$ for $g = ma$, with $m \in N$, $a \in A$, and therefore the condition §3.3.3(i) of the criterion is met because the right Haar measure satisfies $|E_n| = O(n^C)$ for some appropriate constant. To verify that condition (ii) of the criterion is also satisfied we shall use the following observation:

For all $C > 0$ there exists a constant C_1 such that $L(a) \leqslant C$, with $L \in \mathscr{L}_\mathfrak{n}$ and $a \in A$, implies $(\mathrm{Ad}\, a)(\mathfrak{n}(r)) \subset \mathfrak{n}(C_1 r)$ for all $r > 0$ and therefore also $(N(r))^a = aN(r)a^{-1} \subset N(C_1 r)$, by (6.1).

We shall consider now $\check{s}_j = p(\dot{s}_j) = \psi(s_j)$ where the notation is as in §6.6.1; this is a symmetric random walk on $\mathbb{Z}^m$. With the notation of §6.5.2 we shall then define the event

$$\mathscr{E}_n \colon L(\check{s}_j) \leqslant C; \quad 1 \leqslant j \leqslant n,\ L \in \mathscr{L}_\mathfrak{n},$$

for some appropriate constant. Therefore, by the hypothesis on the measure in §6.6.1 and the gambler's ruin estimate in $\mathbb{Z}^m$ of §3A, we deduce that $\mathbb{P}_0(\mathscr{E}_n) \geqslant Cn^{-c}$ for appropriate constants, where we use the starting probability $\mathbb{P}_0[s_0 = e] = 1$.

From this and by the above observation, we have $|m_j^{\dot{s}_{j-1}}| \leqslant C$ on the event $\mathscr{E}_n$ and therefore $m_1 m_2^{\dot{s}_1} \cdots m_n^{\dot{s}_{n-1}} \in N(C_1)$ by §6.4.4. We conclude therefore from (6.5) that condition (ii) of the criterion in Chapter 3 is satisfied. So we are through.

Note Notice how much simpler this proof is compared with the classical case of Chapter 3. The properties that we explained in §6.4.4 simplify the formulas but one has to get over the usual psychological obstacle that is related to the ultrametric property. What really simplifies the proof is the fact that the group is already a semidirect product of N and A and that $\mathrm{Ad}(A)$ acts semisimply on the Lie algebra of N.

6.8 Proof of the C-Theorem

6.8.1 The construction of the exact sequence

The group $G = NA$ is as in the C-theorem (Theorem 6.2). The notation is as in §6.4.2 and $\mathfrak{n}$ is the Lie algebra of N. By Chevalley (1955, §V.3.4), there exists $N_2 \subset N$ an irreducible algebraic normal subgroup that corresponds to $\mathfrak{n}^2 = [\mathfrak{n}, \mathfrak{n}]$. Since by §6.4.3 the commutator subgroup satisfies $[N,N] \subset N_2$, the locally compact group $H = N/N_2$ is Abelian. Similarly, the locally compact group $\widetilde{G} = G/N_2$ can be identified with $\widetilde{G} = H \leftthreetimes A$. No attempt will be made to identify H or $\widetilde{G}$ with algebraic groups, because this will not be necessary, but for these locally compact groups we have the direct analogue of the exact sequence of (2.18):

$$0 \longrightarrow H \longrightarrow \tilde{G} \longrightarrow A \longrightarrow 0. \tag{6.6}$$

Note On the other hand, to assign analytic group structures on these groups over the field K is easy to do. In the ad hoc proofs that we give of the properties below, this additional structure has not been exploited. Well-known structure theorems do exist, however, for the Abelian analytic group H (see Serre, 1965, §V.7, Theorem 2, Corollary 4). Using the action of A on H it might therefore be possible to obtain more direct proofs in the spirit of what we did in §2.6. This, however, is an opening to another subject: that of p-adic Lie groups.

For the proof of our theorem our main task is to verify that this exact sequence has the same properties as in Chapter 2 and that we can define the roots as in §2.6.3. More explicitly the following facts will be proved.

(a) $H = H_1 \oplus \cdots \oplus H_p$ can be decomposed as a direct sum of closed subgroups and each H_j is isomorphic as a locally compact group with K^{d_j}, for $1 \leqslant j \leqslant p$, with the direct product topology.
(b) The subgroups H_j are normal in $\widetilde{G}$ and the inner automorphisms $\tau_a : x \longmapsto axa^{-1}$, $a \in A$ can be identified with elements of $\mathrm{GL}_K(K^{d_j})$ and we can define what we shall call the roots of the action of A Λ_j, by $\check{\tau}_a(\mathrm{Haar}\, H_j) = \exp(\Lambda_j(a))(\mathrm{Haar}\, H_j)$ where $\check{\tau}_a$ is the induced mapping on measures.
(c) The roots in (b) are group homomorphisms $\Lambda_j : A \to \mathbb{R}$ and we have $\Lambda_j = d_j L_j$ for the real roots in $\mathscr{L}_W$ defined in §6.5.1 that are now counted without repetition.

In particular, it follows from (c) that the roots Λ_j satisfy the property of §6.5.3 and can be identified as $\Lambda_j : \mathbb{Z}^m \to \mathbb{R}$. Furthermore, if $\varphi : G \to G/N_2$ is the canonical projection we can define $\check{\mu}_j = \check{\varphi}(\mu_j) \in \mathbb{P}(\widetilde{G})$ new measures that satisfy conditions (i), (ii) and (iii)$'$ of §6.6.1. As in Chapter 2 the main task for the proof of Theorem 6.2 is to prove that the measure $\check{\mu}^n = \check{\mu}_1 * \cdots * \check{\mu}_n$ on $\widetilde{G}$ satisfies the analogues of (6.2) and (6.3), or in other words, that the C-theorem holds for the group $\widetilde{G}$. Once this is done, estimate (6.3) holds for the original measures because for any $P \subset \widetilde{G}$ we have $\mu^n\left(\varphi^{-1}(P)\right) = \check{\mu}^n(P)$.

For the proof of (6.3) for the group $\widetilde{G}$ we use properties (a), (b), (c) of the exact sequence and the arguments of §2.7–2.10. To see that these arguments are applicable we disintegrate the measures $\check{\mu}_j$ as in §2.7 and because of the semi-direct product structure $\tilde{G} = H \leftthreetimes A$ this disintegration is considerably more transparent than in Chapter 3. Then the arguments of §§2.8–2.10 can be repeated verbatim but for the following two modifications.

Note that in the sampling of §2.9 we cannot use the reduction §2.5.3 as it stands because of what was said in the exercise of §6.6.2. Note also that for formulation (6.3) of the C-theorem we use again the Borel section and the argument at the end of §2.16. This, as in §2.16, is necessary because of the fact that we do not have Harnack at our disposal.

The second modification comes at the very end of §2.10 where we again use the properties in §6.6.1 of the measures $\check{\mu}_1, \ldots$ on $\widetilde{G}$. What needs to be done is to adapt the results of §§2A.1–2A.2 for lattice distributions on $\mathbb{Z}^n$. This is straightforward because the machinery of the central limit theorem (see Feller, 1968 or Gnedenko and Kolmogorov, 1954). Once this is done we proceed exactly as in §2.10, now using the above roots Λ_j on $\mathbb{Z}^n$. The verification of both of these steps is left to the reader.

The rest of the argument in §§2.7–2.10 goes through verbatim and this completes the proof of the C-theorem. It now remains to give the proof of properties (a), (b), (c) of the exact sequence.

Proof of properties (a), (b), (c) The fact that H is Abelian follows from §6.4.3. We shall now choose $\beta_1, \ldots, \beta_p$ finite subsets of $\mathfrak{n}$ such that $d\varphi(\beta_j)$ is a basis of $\widetilde{W}_j$ for $1 \leqslant j \leqslant p$ in the real root space decomposition of W in §6.5.4. We can then complete by a finite set $\beta_0 \subset \mathfrak{n}^2$ in such a way that $\beta = \beta_0 \cup \beta_1 \cup \cdots \cup \beta_p$ is a basis of $\mathfrak{n}$ and on that basis we can use the result in Chevalley (1955, §V.3.4, Prop. 17), which gives what is often referred to as exponential coordinates of the second kind for N (see Varadarajan, 1974, §3.18.11 for real Lie groups). Using these coordinates every $x \in N$ can be written as a product

$$\begin{aligned} x &= y_0 y_1 \cdots y_p, \\ y_i &= \exp\left(z_i^{(1)}\right) \cdots \exp\left(z_i^{(d_i)}\right), \end{aligned} \tag{6.7}$$

where $z_i^{(k)} \in \mathfrak{h}_i$ = (linear combination of elements of β_i) for $0 \leqslant i \leqslant p$, and where $d_i = \dim \tilde{W}_i$ for $i \geqslant 1$.

Exercise Use §6.4.3 and the BCH formula of §6.4.4, as well as the above reference in Chevalley (1955, §V.3.4), and (6.8) below to verify this.

We shall now need the following converse of the BCH (Baker–Campbell–Hausdorff) formula. This is called the Zassenhaus formula (see Magnus et al., 1965, §5.41) and it asserts that

$$\exp(\xi + \eta) = \exp(\xi)\exp(\eta)E_1 E_2 \cdots E_n; \quad \xi, \eta \in \mathfrak{n}, \tag{6.8}$$

where the finitely many cofactors are of the form $E_j = \exp(Z_j(\xi, \eta))$ and where the Z_j are, as in the BCH formula, linear combinations of $[\theta[\cdots]\cdots]$ for $\theta = \xi, \eta$, of length *at least* 2. A direct proof of this based on the BCH formula can be found in Hörmander (1967).

By repeated use of the BCH formula and of (6.8) and §6.4.3 on (6.7), we finally conclude that each $x \in N$ can be written uniquely as

$$x = \exp\big(\zeta_0\big)\exp\big(\zeta_1\big)\cdots\exp(\zeta_p); \quad \zeta_j \in \mathfrak{h}_j,\ 0 \leqslant j \leqslant p. \tag{6.9}$$

Furthermore, the bijection $N \ni x \leftrightarrow (\zeta_0, \zeta_1, \ldots, \zeta_p) \in \mathfrak{n}^2 \times \mathfrak{h}_1 \times \cdots \times \mathfrak{h}_p$ is given by polynomial functions both ways and is bicontinuous for the locally compact topologies induced by the matrices with coefficients in K. To see this we can use the inverse of the exp mapping of §6.4.4 given by the 'logarithm' as explained in Chevalley (1955, §V.3.4, top of p. 124). Now observe that for $j = 1, \ldots, p$ we have $\zeta_j(xy) = \zeta_j(x) + \zeta_j(y)$ (but not necessarily for $j = 0$). It follows that the above bijection induces an isomorphism of $H = N/N_2$ with $H_1 \oplus \cdots \oplus H_p$ and that the subgroups $H_j \subset H$ can be identified with the $\widetilde{W}_j \cong K^{d_j}$ and this preserves of course the Haar measures of the groups. This proves (a).

Part (b) is a consequence of the fact that $(\mathrm{Ad}\,a)\mathfrak{h}_j \subset \mathfrak{h}_j + \mathfrak{n}^2$ for $a \in A$ (this holds by the definition of these spaces) and the fact that the exponential mapping intertwines $\mathrm{Ad}\,a$ and τ_a (see (6.1)). Part (c) follows from the $\mathrm{Ad}\,a$ action on $\widetilde{W}_j$ in §6.5.1. Notice that for this we only need to use the above remarks on Haar measures (Exercise 6.1 is not used here). □

6.9 Final Remarks

In the spirit of what was done in Part 5.3, it can easily be seen that with our notation and for groups and measures as in §6.6.1, we have

$$\phi_n(e) \geqslant C\exp\big(-cn^{1/3}\big); \quad n \geqslant 2,$$

for appropriate constants provided that μ is symmetric as in the NC-theorem. The point here is that this should hold irrespective of whether G is C or NC.

The proof in §6.7 (modified as in §5.13 to use the 'scale' of §5.3.2) gives this at once but with one additional difficulty. One has to prove that the measure $|N(r)|$ grows polynomially in r. This non-trivial fact is a consequence of Exercise 6.1. The other question that has to be clarified here is of course that of the amenability for the groups that we have considered in this chapter. Soluble groups are amenable by general considerations (Reiter, 1968) but how about the classification of §3.1.5? Can this be adapted here?

In the same spirit, a much more difficult exercise consists in adapting Varopoulos (1999b) to give the local central limit theorem $\phi_n(e) \sim n^{-\nu}$ for (6.3) in the NC-theorem for some ν that depends on G and μ. This would be the analogue of (3.3) but this has never been written out.

Finally, it is very likely that with the existing structure theorems on reductive algebraic groups (Borel and Tits, 1965; Bruhat and Tits, 1972) we could make the classification of §1.3 go through for all algebraic groups over a locally compact field of characteristic 0. The Levi decomposition certainly exists for these groups at least up to a point (this goes back to Chevalley, 1955, §V.4, No. 2 and the end of No. 1 on page 143 of Tome III, where one sees what goes wrong for fields that are not algebraically closed, or the more recent extensive literature on the subject, e.g. Humphreys, 1975, §§30.2, 34.5 and the references therein). However, no effort has been made to write a proof in that generality. A similar project for positive characteristics would certainly be much more difficult.

At any rate, all this is outside the scope of this book and it would require a reasonable understanding of 'rationality questions' in algebraic group theory. This specialised and technical subject deals with algebraic groups over fields that are not algebraically closed, and goes beyond the results of Chevalley (1955) on fields of characteristic 0.

Appendix A
Semisimple Groups and the Iwasawa Decomposition

Those readers not familiar with the structure theory of semisimple Lie algebras will find Appendices A, B and C difficult. In these appendices we give a global 'overall' picture of the algebras involved in our theory and also, and this is their main purpose, we give the algebraic ingredient that is needed for the homotopic classification (see §1.6). For more on 'what's involved here' and 'what implies what', see Appendix F and also §12.6 in homology theory.

Throughout the appendix, we shall denote by $\mathfrak{g}$ some finite-dimensional Lie algebra. As before unless otherwise stated, all these algebras will be over the real field but here we shall also make extensive use of the theory of complex semisimple Lie algebras. We shall denote by $\mathfrak{q}$ (resp. $\mathfrak{n}$) the radical (resp. the nilradical) of $\mathfrak{g}$. Our main source of references will tacitly be Jacobson (1962); Bourbaki (1972); Varadarajan (1974); Helgason (1978); additional more specialised references will be given as we go along in these appendices. I should also add that the style of the presentation is informal and sketchy, and as a consequence not as precise as it should be. Readers not familiar with the theory, who wish to take all this seriously, will therefore have to surf the above references quite a lot.

A.1 The Levi Decomposition

With the notation as above there exists $\mathfrak{s} \subset \mathfrak{g}$ some semisimple subalgebra (i.e. the radical of $\mathfrak{s}$ reduces to $\{0\}$) such that $\mathfrak{g} = \mathfrak{q} + \mathfrak{s}$ as vector spaces and $\mathfrak{q} \cap \mathfrak{s} = \{0\}$. This is to say that we can write $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$, a semidirect product. This is called the Levi decomposition of $\mathfrak{g}$. The subalgebra $\mathfrak{s}$ is called a Levi subalgebra.

The above decomposition is 'essentially unique' in the sense that if $\mathfrak{s}_1, \mathfrak{s}_2$ are two Levi subalgebras, there exists $\alpha \in \mathrm{Aut}(\mathfrak{g})$, an automorphism of the algebra such that $\alpha(\mathfrak{s}_1) = \mathfrak{s}_2$.

Inner automorphisms

In fact, this automorphism can be chosen to be an inner automorphism $\alpha \in \mathrm{Int}(\mathfrak{g})$. This means that there exists $x \in G$, the simply connected Lie group that corresponds to $\mathfrak{g}$ such that $\alpha = \mathrm{Ad}(x)$. For our immediate purposes this point is not essential. Later on, however, we shall make use of inner automorphisms and we shall also use the following observation.

Let $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ *be some Levi decomposition and let* $\alpha \in \mathrm{Int}(\mathfrak{s})$ *some inner automorphism of* $\mathfrak{s}$. *Then* α *can be extended to* $\overline{\alpha} \in \mathrm{Int}(\mathfrak{g})$*, that is,* $\overline{\alpha}|_{\mathfrak{s}} = \alpha$. *This is clear by considering* $G = Q \leftthreetimes S$ *the corresponding Levi decomposition of the group* G (Jacobson, 1962; Bourbaki, 1972, Chapter 1; Varadarajan, 1974).

A.2 Compact Lie Groups

Let $\mathfrak{g}$ be some semisimple real Lie algebra and let G, G_1 be two connected Lie groups that admit $\mathfrak{g}$ as a Lie algebra. If we assume that G is compact then so is G_1. We then say that $\mathfrak{g}$ is of compact type. We also say that $\mathfrak{g}$ is of non-compact type (or abusing terminology, just non-compact) if it is the direct sum of simple Lie algebras that are not compact. Recall that a non-Abelian Lie algebra is called simple if its only ideals are $\{0\}$ and the whole algebra. No confusion should arise in the terminology ‘non-compact’ that we shall be using. This is not the same as saying that the corresponding group is not compact. A more precise but also more cumbersome terminology would have been to call them ‘semisimple Lie algebras without non-trivial compact factors’.

Without semisimplicity the above breaks down, as the two groups $\mathbb{R}^n$ and its ‘compact model’ $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ show.

Nonetheless, the above two examples supply all the ‘compact Lie algebras’. More precisely, let G be some compact Lie group. Then the Levi decomposition of the Lie algebra $\mathfrak{g}$ of G is $\mathfrak{g} = \mathfrak{z} \oplus \mathfrak{s}$ where $\mathfrak{z}$ is the centre and $\mathfrak{s}$ is semisimple. Here the Levi subalgebra is unique $\mathfrak{s} = [\mathfrak{g}, \mathfrak{g}]$ (see Helgason, 1978, Chapter II, Proposition 6.6). This means in particular that a simply connected compact Lie group has to be of the form $\mathbb{R}^d \times G$ with G semisimple and compact. But then $d = 0$ and therefore such a group is semisimple.

Some notation

If $\mathfrak{s}$ is some semisimple Lie algebra then $\mathfrak{s}$ can be decomposed uniquely $\mathfrak{s} = \mathfrak{s}_n + \mathfrak{s}_c$ where $\mathfrak{s}_n = \mathfrak{s}_{\text{non-compact}}$ (resp. $\mathfrak{s}_c = \mathfrak{s}_{\text{compact}}$) is the non-compact (resp. compact) component. This notation will be adopted throughout.

Here of course $\mathfrak{s}_n$ is a direct sum of simple algebras of non-compact type in the above sense. It should, however, be noted that this is not the correct way of presenting things. The Cartan–Killing form is what has to be used. The reader who is not familiar with this beautiful theory is strongly urged to take the required time to study it (Jacobson, 1962; Bourbaki, 1972; Varadarajan, 1974; Helgason, 1978).

A.3 Non-compact Lie Algebras and the Iwasawa Decomposition

In this section $\mathfrak{s}$ will denote throughout some non-compact real Lie algebra, that is, $\mathfrak{s}_c = 0$ in the above terminology. By an Iwasawa decomposition of $\mathfrak{s}$ we mean the direct vector space sum

$$\mathfrak{s} = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k} \tag{A.1}$$

that has the properties below. This is sometimes called 'the' Iwasawa decomposition. This abuse of terminology will be justified later, in §A.4.

(i) All three components in (A.1) are Lie subalgebras: $\mathfrak{n}_s \neq 0$ is nilpotent, $\mathfrak{a} \neq 0$ is Abelian and $\mathfrak{k} \neq 0$ is the Lie algebra of a compact group as in §A.2. Furthermore, $[\mathfrak{a}, \mathfrak{n}_s] \subset \mathfrak{n}_s$ and therefore $\mathfrak{n}_s + \mathfrak{a}$ is a soluble algebra.

(ii) More precisely, in the soluble algebra $\mathfrak{n}_s + \mathfrak{a}$ we have $[\mathfrak{n}_s, \mathfrak{a}] = \mathfrak{n}_s$ and the ad-action of $\mathfrak{a}$ on $\mathfrak{n}_s$ is semisimple with real non-zero roots $\mu_1, \ldots, \mu_n \in \mathfrak{a}^*$ (which is the real dual of $\mathfrak{a}$).

An immediate consequence of this is that the nilradical of $\mathfrak{n}_s + \mathfrak{a}$ is $\mathfrak{n}_s$ (see also Knapp, 1986, Prop. 5.10). For our purposes, what is also important is that these roots satisfy the NC-condition; that is, if $\alpha_j \geqslant 0$, for $1 \leqslant j \leqslant n$, are such that $\sum \alpha_j \mu_j = 0$ then all the α_j are zero. Anticipating the definition that will be given later (in Chapter 8), the above conditions say that $\mathfrak{n}_s + \mathfrak{a}$ is an SSA (special soluble algebra) of NC type.

(iii) *The Iwasawa decomposition of the group*. Let S be some connected, not necessarily simply connected, semisimple (real) Lie group and let $\mathfrak{s}$ be its Lie algebra, which is assumed to be of non-compact type and admit the Iwasawa decomposition (A.1). The following properties then hold good.

(a) The analytic subgroups N_s, A, K that correspond respectively to $\mathfrak{n}_s$, $\mathfrak{a}$, $\mathfrak{k}$ are closed.

(b) N_s and A are simply connected and the analytic mapping

$$N_s \times A \times K \longrightarrow S; \quad (n,a,k) \longrightarrow nak$$

is an analytic diffeomorphism of the (analytic) C^∞ manifolds.

(c) K is simply connected if and only if S is.

This lets us write $S = N_s AK$, called the Iwasawa decomposition of S.

(iv) The centre $Z = Z(S)$ of S, which by the semisimplicity is necessarily discrete, lies in K: $Z \subset K$. Furthermore, S/Z has trivial centre (because $\pi\colon S \to S/Z$ is a covering mapping and since also the centre $Z_1 = Z(S/Z)$ is discrete, it follows that $\pi^{-1}(Z_1)$ is a discrete normal subgroup of S and thus it is central). Furthermore, K/Z is compact (see Helgason, 1978, V1, Theorem 1.1).

(v) From the above properties it follows that $K \cong \mathbb{R}^d \times K_1$ for some compact group K_1 (see Hochschild, 1965, XIII, §2.1). The subgroup K_1 is uniquely determined in K because any compact subgroup of $\mathbb{R}^d \times K_1$ has to lie in K_1 (to see this, project on $\mathbb{R}^d$). Furthermore, when S and therefore K are simply connected then K_1 is a simply connected compact group and therefore it has to be semisimple (see §A.2). See Helgason (1978); Gangoli and Varadarajan (1980); Knapp (1986).

A.4 Uniqueness

One issue that arises is whether the decomposition (A.1) is essentially uniquely determined (i.e. up to an automorphism of $\mathfrak{s}$ as in §A.1) by the properties that we have enumerated, possibly together with additional similar properties if necessary. The author of this book does not feel very comfortable in this area, and the only explicit reference he is aware of where this problem of uniqueness is addressed is Onischik and Vinberg (1988, §5.4.5).

Nonetheless, what will be done is to describe an explicit construction that we shall call the Iwasawa construction. This will be done in the three steps below and they will lead to a decomposition as in (A.1). We do not obtain just one single decomposition with this construction but several, depending on the choices that are made on the way in the three different steps. Nonetheless, here we pick up a uniqueness property that will allow us to call (A.1) 'the Iwasawa decomposition'. More precisely, if the two decompositions

$$\mathfrak{s} = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k} = \mathfrak{n}_s' + \mathfrak{a}' + \mathfrak{k}'$$

are obtained from two different sets of choices, then there exists $\alpha \in \mathrm{Int}(\mathfrak{s})$ such that

$$\alpha(\mathfrak{n}_s) = \mathfrak{n}'_s, \qquad \alpha(\mathfrak{a}) = \mathfrak{a}', \qquad \alpha(\mathfrak{k}) = \mathfrak{k}'. \tag{A.2}$$

Notice that if we pass to the complex field, this uniqueness is related to the theory of parabolic subgroups (see Humphreys, 1975; Gangoli and Varadarajan, 1980 and also §A.8 below),

A.5 First Step: The Cartan Decomposition and the Choice of $\mathfrak{k}$

This is a fundamental step and it lies at the heart of the theory of real semisimple Lie algebras. To conform with the notation of Helgason (1978), which we shall follow very closely, we write $\mathfrak{g}_0$ rather than $\mathfrak{s}$ for such an algebra. The Cartan decomposition of the algebra is then a decomposition $\mathfrak{g}_0 = \mathfrak{k} + \mathfrak{p}$ where $\mathfrak{k}$ is a compact algebra, that is, the algebra of some compact group, not necessarily semisimple, as in §A.2 (see Helgason, 1978, Chapter II, §5 for a more intrinsic definition; the two definitions are easily seen to be equivalent but the fact will not be needed). This will supply the component $\mathfrak{k}$ in the Iwasawa decomposition (A.1). The subspace $\mathfrak{p}$ is not a subalgebra but we have instead $[\mathfrak{p},\mathfrak{p}] \subset \mathfrak{k}$, $[\mathfrak{k},\mathfrak{p}] \subset \mathfrak{p}$.

The construction of the Cartan decomposition is done as follows. We complexify $\mathfrak{g} = \mathfrak{g}_0 \otimes \mathbb{C}$ to obtain a complex semisimple Lie algebra. Using the fundamental root space decomposition and the structure of $\mathfrak{g}$, we can then find $\mathfrak{g}_k \subset \mathfrak{g}$ some *compact real form*. Several such compact real forms exist. This means that $\mathfrak{g}_k \subset \mathfrak{g}$ is a subalgebra of $\mathfrak{g}$ viewed as a real Lie algebra. Here $\mathfrak{g}$ over $\mathbb{R}$ is a real Lie algebra of $\dim_{\mathbb{R}} \mathfrak{g} = 2\dim \mathfrak{g}_0$ and both $\mathfrak{g}_0$ and $\mathfrak{g}_k$ can be identified with real subalgebras of $\mathfrak{g}$. Two things play together in the choice of $\mathfrak{g}_k$.

First, $\mathfrak{g}$ is a complexification of $\mathfrak{g}_k$, meaning $\mathfrak{g} = \mathfrak{g}_k + i\mathfrak{g}_k$ and $\mathfrak{g}_k \cap i\mathfrak{g}_k = \{0\}$. This is exactly the same way that $\mathfrak{g}$ is a complexification of $\mathfrak{g}_0$. This forces $\mathfrak{g}_k$ to be a *semisimple* real Lie algebra. Hence the terminology ‘real form’. Furthermore, a choice of $\mathfrak{g}_k$ can be made so that $\mathfrak{g}_k$ is a compact semisimple real algebra, hence the term ‘compact real form’. None of this is evident (see Helgason, 1978). One more condition is imposed on the choice of $\mathfrak{g}_k$, namely that it ‘fits’ well with the original algebra $\mathfrak{g}_0$ in the sense that

$$\mathfrak{g}_0 = \mathfrak{g}_0 \cap \mathfrak{g}_k + \mathfrak{g}_0 \cap i\mathfrak{g}_k; \quad \mathfrak{k} = \mathfrak{g}_0 \cap \mathfrak{g}_k; \quad \mathfrak{p} = \mathfrak{g}_0 \cap i\mathfrak{g}_k.$$

This is the Cartan decomposition. Several such decompositions can be given depending on the choice of $\mathfrak{g}_k$. But if

$$\mathfrak{g}_0 = \mathfrak{k}_1 + \mathfrak{p}_1 = \mathfrak{k}_2 + \mathfrak{p}_2$$

are two different ones there exists $\psi \in \mathrm{Int}(\mathfrak{g}_0)$ such that

$$\psi(\mathfrak{k}_1) = \mathfrak{k}_2, \quad \psi(\mathfrak{p}_1) = \mathfrak{p}_2. \tag{A.3}$$

A.6 Second Step: The Choice of $\mathfrak{a}$

We shall now fix a Cartan decomposition as in §A.5. The choice of $\mathfrak{a}$, or rather 'a choice' of $\mathfrak{a}$, in the Iwasawa decomposition (A.1) is as follows. We choose $\mathfrak{a} \subset \mathfrak{p}$, some subspace that is Abelian in the sense that $[X,Y] = 0$ for $X,Y \in \mathfrak{a}$ and such that $\mathfrak{a}$ is maximal under that condition. One can then prove that if $\mathfrak{a}_1, \mathfrak{a}_2$ are two different choices then there exists $\alpha \in \mathrm{Int}(\mathfrak{s})$, with $\mathfrak{s} = \mathfrak{g}_0$ in the notation of (A.1) and §A.5, such that (see Knapp, 1986, §5.13)

$$\alpha(\mathfrak{k}) = \mathfrak{k}, \quad \alpha(\mathfrak{p}) = \mathfrak{p}, \quad \alpha(\mathfrak{a}_1) = \mathfrak{a}_2. \tag{A.4}$$

A.7 Third Step: The Choice of $\mathfrak{n}$

Some choice of the Cartan decomposition and of $\mathfrak{k}$ and $\mathfrak{a}$ will be fixed as in §§A.5–A.6. From this we shall construct $\mathfrak{n}$, some subalgebra that will satisfy the conditions of the Iwasawa decomposition (A.1). Finitely many possibilities will be given for this and they will have the property that if $\mathfrak{n}_1, \mathfrak{n}_2$ are two such choices then there exists $\alpha \in \mathrm{Int}(\mathfrak{g}_0)$ such that (see Knapp, 1986, §5.18 or Gangoli and Varadarajan, 1980, equation (2.2.12))

$$\alpha(\mathfrak{k}) = \mathfrak{k}, \quad \alpha(\mathfrak{p}) = \mathfrak{p}, \quad \alpha(\mathfrak{a}) = \mathfrak{a}, \quad \alpha(\mathfrak{n}_1) = \mathfrak{n}_2.$$

This, in combination with (A.3) and (A.4), gives the required uniqueness in (A.2). The construction of $\mathfrak{n}$ is done as follows.

First one proves that the (Abelian) action of $\mathrm{ad}\,\mathfrak{a}$ on $\mathfrak{g}_0$ is diagonalisable with real roots for some appropriate basis of $\mathfrak{g}_0$. This allows us to construct the root space decomposition of $\mathfrak{g}_0$ where the roots are 0 and the non-zero roots $\lambda \in \Sigma \subset \mathfrak{a}^*$ so that we have

$$\mathfrak{g}_0 = \mathfrak{g}_{0,0} + \sum_{\lambda \in \Sigma} \mathfrak{g}_{0,\lambda}; \quad \left[\mathfrak{g}_{0,\lambda}, \mathfrak{g}_{0,\mu}\right] \subset \mathfrak{g}_{0,\lambda+\mu}, \tag{A.5}$$

with $\mathfrak{g}_{0,\cdot}$ the corresponding root spaces.

Now an order relation can be given on $\mathfrak{a}^*$ and we shall denote by $\Sigma^+ \subset$

Σ the set of positive roots for that order relation. For instance, we could fix $a_1, \ldots, a_r \in \mathfrak{a}$ some basis and define the order relation by saying that $\varphi > 0$ in $\mathfrak{a}^*$ if the first among the coordinates $\varphi(a_1), \ldots, \varphi(a_r)$ that is not zero is positive (cf. Helgason, 1978, Chapter 3, §5).

We shall then define $\mathfrak{n}_s = \sum_{\lambda \in \Sigma^+} \mathfrak{g}_{0,\lambda}$; then from (A.5) it follows automatically that $\mathfrak{n}$ is a nilpotent algebra and that properties (i), (ii) of §A.3 hold good. What needs proving is that we indeed have an Iwasawa decomposition and that the properties in §A.3 hold.

Remark on the terminology

The roots λ in Σ are called the restricted roots. These are not what one usually refers to as the roots, which are those obtained in a *complex* Lie algebra $\mathfrak{g}$ by the action of a Cartan subalgebra $\mathfrak{h}$.

For instance, it is possible that both $\lambda \in \Sigma$ and $2\lambda \in \Sigma$. Also the multiplicity of these restricted roots is not necessarily 1.

The construction and the geometric configuration of the restricted roots is of course closely related to the roots of the complex Lie algebra $\mathfrak{g} = \mathfrak{g}_0 \otimes \mathbb{C}$ but we do not have to go into this here. On the other hand, this become crucial when we wish to compute the exact exponent ν in the local central limit theorem of (3.3) (see also Bougerol, 1981). As we said, none of this will be considered in this book (Helgason, 1978 is the reference that we have followed, but all the above can also be found in Gangoli and Varadarajan, 1980 and Knapp, 1986 and other books on representation theory).

A.8 Uniqueness up to Automorphism of the Iwasawa Radical. Borel Subgroups

With the notation that we have introduced and the Levi decomposition $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ we shall fix $\mathfrak{n}_s + \mathfrak{a} + \mathfrak{k}$, some Iwasawa decomposition of the non-compact component of $\mathfrak{s}$ if that component is $\neq 0$. Otherwise we set $\mathfrak{n}_s = \mathfrak{a} = 0$. With this notation that we defined in §4.1 the Iwasawa radical $\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_s + \mathfrak{a}$.

The main purpose of this appendix is to show that $\mathfrak{r}$ is unique up to an automorphism $\alpha \in \mathrm{Aut}(\mathfrak{g})$ (in fact $\alpha \in \mathrm{Int}(\mathfrak{g})$, but this is not important now). As a result the classification that says that $\mathfrak{g}$ is NB if and only if $\mathfrak{r}$ is NC is a genuine classification. This uniqueness is now a consequence of the uniqueness of the Levi decomposition in §A.1 together with (A.2) and the final remark in §A.1. Recall also from §4.1 that, without proving this uniqueness of $\mathfrak{r}$, the fact that we obtain a classification nonetheless follows by its equivalence with the analytic and geometric characterisations of the B–NB conditions (see §4.1). On the issue of the above uniqueness, the considerations that follow are relevant but outside the scope of the book, so we shall be brief.

If we pass to the Lie group G and if we assume that G is a complex algebraic group, that is, an algebraic subgroup of $\mathrm{GL}_n(\mathbb{C})$ (see Humphreys, 1975; Chevalley, 1951 and also Chapter 6), then it is a fundamental fact that we can find $B \subset G$ some closed subgroup that is soluble and cocompact; that is, the homogeneous space G/B is compact for the Hausdorff topology. We could have chosen B to be a Borel subgroup. Such subgroups are characterised by the following two conditions. First B is soluble and closed. And second, no larger subgroup $\widetilde{B} \supsetneqq B$ is soluble. What is important here and all round, is that if B, B_1 are two Borel subgroups there exists an automorphism $\alpha \in \mathrm{Aut}(G)$ such that $\alpha(B) = B_1$. Now it is clear from both the analytic and the geometric theory that G is B or NB if and only if the Borel subgroups B are C or NC, respectively.

This gives yet another characterisation for this special class of groups of the B–NB condition. This characterisation is algebraic in nature. Furthermore, we can relate this with the Iwasawa radical because when G is semisimple the subgroup N_sA coming from the Iwasawa components $\mathfrak{n}_s + \mathfrak{a}$ is closely connected to the Borel subgroups of the complexified group G_c (see Gangoli and Varadarajan, 1980).

The disadvantage of using Borel subgroups to characterise the B–NB condition is that the method can only be used for *complex algebraic groups*. The restriction that we have to work with in an algebraically closed field does not pose serious problems because we can easily complexify and pass from a real Lie group to a complex one. The restriction that the Lie group G, and hence $\mathfrak{g}$ its Lie algebra, has to be algebraic is more serious. Indeed, the following question arises. Let $\mathfrak{g} \subset \mathfrak{gl}(V)$ be some Lie subalgebra of the full Lie algebra of linear transformations on the vector space V (over $\mathbb{R}$ or $\mathbb{C}$) and let $\overline{\mathfrak{g}} \subset \mathfrak{gl}(V)$ be its algebraic closure (see Chevalley, 1951, §II.14; Varadarajan, 1974, §3.1.15). Is it then true that $\mathfrak{g}$ is B (resp. NB) if and only if $\overline{\mathfrak{g}}$ is? We shall not address this question here because, although interesting, it lies outside the scope of this book.

A closing remark about the algebraic B–NB classification Unfortunately, there is no way of proving the uniqueness (e.g. (A.2), (A.3) or the corresponding fact on Borel subgroups) without going deeply into the details of a different subject. A glance at the references I have given will also show that such a digression is really out of the question and pointless in this book. But the hope is that what we have done in this appendix demystifies the ‘magic words’ Iwasawa decomposition. With regard to the uniqueness, as we have already said, one can always fall back on the analytic geometric classifications and forget all about it.

Finally, for readers familiar with the subject, the following is an equivalent and more intrinsic version of the algebraic classification of §4.1. The notation is the same as before, and let $\mathfrak{b}$ be some minimal parabolic subalgebra of $\mathfrak{s}$ (cf. Gangoli and Varadarajan, 1980, §2.3). Then $\mathfrak{g}$ is B- (resp. NB-) if $\mathfrak{q}+\mathfrak{b}$ is C- (resp. NC-). Furthermore, all these subalgebras $\mathfrak{q}+\mathfrak{b}$ are conjugate by inner automorphisms and they coincide with the set of all maximal amenable subalgebras of $\mathfrak{g}$ (cf. Warner, 1970, §1.2.3).

A.9 The Nilradical of the Iwasawa Radical $\mathfrak{r}$

The notation is as before. We shall denote by $\mathfrak{n}_r \subset \mathfrak{r}$ the nilradical of the soluble algebra $\mathfrak{r}$. What is certain is that $\mathfrak{n}$, being a nilpotent ideal of $\mathfrak{r}$, lies in $\mathfrak{n}_r$; but it is also true that $\overline{\mathfrak{n}} = \mathfrak{n}+\mathfrak{n}_s$ is a nilpotent ideal of $\mathfrak{r}$ and $\overline{\mathfrak{n}} \subset \mathfrak{n}_r$.

To see this observe that $\mathfrak{n}_s = [\mathfrak{n}_s,\mathfrak{a}] \subset [\mathfrak{r},\mathfrak{r}] \subset \mathfrak{n}_r$, where the last inclusion is a standard fact on soluble algebras (see Varadarajan, 1974, §3.8.3). On the other hand, $\overline{\mathfrak{n}}$ is an ideal in $\mathfrak{r}$ because $\mathfrak{n}$ is an ideal in $\mathfrak{g}$, and because $[\mathfrak{n}_s,\mathfrak{a}] \subset \mathfrak{n}_s$, $[\mathfrak{q},\mathfrak{g}] \subset \mathfrak{n}$ by the same fact as before.

We shall make essential use of the above in what follows. To complete the picture, note that we have the more precise fact $\mathfrak{n}_r = \mathfrak{n}+\overline{\mathfrak{n}}$, but only marginal use will be made of this.

Exercise Prove the above. Observe that $\mathfrak{n}_r \cap \mathfrak{q} \subset \mathfrak{n}$ and therefore $\mathfrak{n}_r \cap \mathfrak{q} = \mathfrak{n}$ because $\mathfrak{n} \subset \mathfrak{n}_r$. Now let $\pi\colon \mathfrak{r} \to \mathfrak{r}/\mathfrak{q} = \mathfrak{n}_s+\mathfrak{a}$, the canonical mapping. For our assertion we need to see that $\pi(\mathfrak{n}_r) \subset \mathfrak{n}_s$. But this is obvious since $\pi(\mathfrak{n}_r)$ is a nilpotent ideal in $\mathfrak{n}_s+\mathfrak{a}$ and thus it has to be contained in the nilradical of this algebra. The nilradical of $\mathfrak{n}_s+\mathfrak{a}$ is $\mathfrak{n}_s$ by §A.3(ii) and we are done.

A.10 A Lemma in the Representations of a Semisimple Lie Algebra

Let $\mathfrak{g}$ be some real non-compact semisimple algebra in the sense of §A.2. We have for the Cartan and the Iwasawa decompositions,

$$\mathfrak{g} = \mathfrak{k}+\mathfrak{p} = \mathfrak{n}_s+\mathfrak{a}+\mathfrak{k}, \quad \mathfrak{a} \subset \mathfrak{p},$$
$$\mathfrak{g}_c = \mathfrak{g}\otimes_{\mathbb{R}}\mathbb{C}, \quad \mathfrak{g}_k = \mathfrak{k}+i\mathfrak{p} \subset \mathfrak{g}_c.$$

The notation is similar but not absolutely identical to what we had before;

for example, here $\mathfrak{g}$ is the $\mathfrak{g}_0$ or the $\mathfrak{s}$ of §A.4 or §A.3. Now let $\rho: \mathfrak{g} \to \mathfrak{gl}(V)$ be some (real) Lie algebra representation in the Lie algebra of linear automorphisms of the real vector space V. We then have the following lemma.

Lemma A.1

(i) *There exists a basis of V with respect to which all the linear transformations $(\rho(a);\ a \in \mathfrak{a})$ become simultaneously diagonal with real eigenvalues $\lambda_1(a), \ldots, \lambda_n(a) \in \mathbb{R}$. Here $n = \dim V$ and multiplicity is admitted.*

(ii) *We have $\lambda_1(a) + \cdots + \lambda_n(a) = 0$ for all $a \in \mathfrak{a}$.*

(iii) *If $\lambda_j(a) = 0$, $a \in \mathfrak{a}$, $j = 1, \ldots, n$ then $\rho = 0$ is the trivial identically zero representation.*

Part (iii) follows immediately from (i) and the fact that $\operatorname{Ker}\rho = [x \in \mathfrak{g};\ \rho(x) = 0]$ is an ideal. That ideal has to be the whole algebra as soon as it contains $\mathfrak{a}$.

Part (ii) is automatic from the fact that $\sum \lambda_j(a) = \operatorname{trace} \rho(a) = 0$. This holds because $x \to \operatorname{trace} \rho(x)$ gives a Lie algebra homomorphism which has to be 0 by the semisimplicity of $\mathfrak{g}$. (Alternatively, this trace clearly vanishes on $[\mathfrak{g}, \mathfrak{g}]$ which is the whole algebra by the semisimplicity.)

The proof of (i) depends on the construction of the Iwasawa decomposition. To see (i) we proceed as follows. We first complexify and extend ρ to a representation of $\mathfrak{g} \otimes \mathbb{C} = \mathfrak{g}_c$ on $V_c = V \otimes \mathbb{C}$: $\rho_c = \rho: \mathfrak{g}_c \to \mathfrak{gl}_{\mathbb{C}}(V_c)$.

It suffices now to show that we can find $\langle \cdot, \cdot \rangle$, some Hermitian scalar product on V_c, for which all the transformations $\rho(a)$ are Hermitian:

$$\langle \rho(a)u, v \rangle = \langle u, \rho(a)v \rangle; \quad a \in \mathfrak{a},\ u, v \in V_c. \tag{A.6}$$

Indeed, once (A.6) has been seen, part (i) of the lemma follows.

The reason why (i) follows is that all the eigenvalues of $\rho(a)$, with $a \in \mathfrak{a}$, have to be real. Since on the other hand $\rho(a) \in \mathfrak{gl}(V_c)$ are real transformations (since $\rho(a)V \subset V$ for $a \in \mathfrak{a}$), all the corresponding eigenvectors which are a basis of V_c over $\mathbb{C}$ actually lie in V. This diagonalises each $\rho(a)$ individually. The fact that the diagonalisation can be done simultaneously for all $\rho(a)$, with $a \in \mathfrak{a}$, follows by the commutativity of $\mathfrak{a}$ using an easy standard argument that will be left as an exercise for the reader (for example, use induction over the dimension of $\mathfrak{a}$, or the existence of a common eigenvector).

The construction of the Hermitian scalar product that satisfies (A.6) is not evident. More explicitly we start from, say, the standard Hermitian scalar product $\langle \cdot, \cdot \rangle_\circ$ induced by some fixed bases $u_1, \ldots, u_n \in V$ (which is to be orthonormal!). We then restrict the representation ρ_c to $\rho_k: \mathfrak{g}_k \to \mathfrak{gl}_{\mathbb{C}}(V_c)$ and consider the induced representation $\hat{\rho}: G_k \to \mathrm{GL}_{\mathbb{C}}(V_c)$ of the simply connected compact group G_k that corresponds to $\mathfrak{g}_k$. Here we use only the fact that $\mathfrak{g}_k$, considered as a real Lie algebra, is a compact semisimple algebra and G_k is a compact (real) Lie group. We then define

$$\langle u,v\rangle = \int_{G_k} \langle\hat{\rho}(g)u,\hat{\rho}(g)v\rangle_\circ\, dg; \quad u,v\in V_c,\ dg = \text{Haar measure.}$$

This gives a Hermitian product (since it is sesquilinear with $\langle u,u\rangle \gtrsim \langle u,u\rangle_0$) for which all the transformations $\hat{\rho}(g)$, with $g\in G_k$, are unitary. By taking the differential of $\hat{\rho}$ we obtain ρ_k and it follows that all the transformations $\rho_k(x)$, $x\in\mathfrak{g}_k$ are skew-Hermitian.

Exercise Verify that

$$0 = \frac{d}{dt}\left\langle\hat{\rho}\left(e^{tx}\right)u,\ \hat{\rho}\left(e^{tx}\right)v\right\rangle\Big|_{t=0} = \langle\rho_k(x)u,v\rangle + \langle u,\rho_k(x)v\rangle\,; \quad x\in\mathfrak{g}_k. \quad \text{(A.7)}$$

But this gives (A.6) since $\mathfrak{a}\subset\mathfrak{p}\subset i\mathfrak{g}_k$ and by (A.7) all the transformations in $i\mathfrak{g}_k$ are Hermitian. This completes the proof of the lemma.

This lemma is bound to exist explicitly somewhere in the literature. But we have given a proof of this here for the convenience of the reader.

Appendix B
The Characterisation of NB-Algebras

B.1 Notation

All the notation and the results of the previous appendix will be used again. We write $\mathfrak{g} = \mathfrak{q} \rtimes \mathfrak{s}$ for the Levi decomposition of the algebra $\mathfrak{g}$ which will be assumed to be an NB-algebra. We shall denote

$$\begin{aligned} &\mathfrak{s} = \mathfrak{s}_c + \mathfrak{s}_n; \quad \mathfrak{s}_c = \text{compact}, \ \mathfrak{s}_n = \text{non-compact}, \\ &\mathfrak{s}_n = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k} \quad \text{if } \mathfrak{s}_n \neq 0, \\ &\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_s + \mathfrak{a} \subset \mathfrak{g} \quad \text{where } \mathfrak{n}_s = \mathfrak{a} = 0 \text{ if } \mathfrak{s}_n = 0. \end{aligned} \tag{B.1}$$

These denote the decomposition of $\mathfrak{s}$ into its compact and non-compact components, the corresponding Iwasawa decomposition of $\mathfrak{s}_n$ when this is not zero and the corresponding Iwasawa radical. When $\mathfrak{q} = \{0\}$ and $\mathfrak{g}$ is semisimple then it is an NB-algebra and there is not much else to say. So we shall assume throughout that the radical $\mathfrak{q}$ and the nilradical $\mathfrak{n}$ are both not zero.

B.2 Further Notation

As in §3.4 we shall now fix $\mathfrak{h} \subset \mathfrak{q}$ (possibly $\{0\}$) some nilpotent subalgebra such that

$$\mathfrak{n} + \mathfrak{h} = \mathfrak{q}, \quad [\mathfrak{h}, \mathfrak{s}] = 0 \tag{B.2}$$

and write

$$\mathfrak{r} = \bar{\mathfrak{n}} + \bar{\mathfrak{h}}, \quad \bar{\mathfrak{n}} = \mathfrak{n} + \mathfrak{n}_s, \quad \bar{\mathfrak{h}} = \mathfrak{h} \oplus \mathfrak{a}; \tag{B.3}$$

here and throughout the notation $\oplus$ will be reserved to indicate a direct *Lie algebra* sum (i.e. $\mathfrak{h} \cap \mathfrak{a} = \{0\}$, $[\mathfrak{h}, \mathfrak{a}] = 0$ in (B.3)). This is in contrast to the sign

$+$ that indicates the sum of vector spaces. This is often, but not always (for instance not in (B.2)), a direct sum of vector spaces.

We saw in §A.9 that

$$\bar{\mathfrak{n}} = \mathfrak{n} + \mathfrak{n}_s \subset \mathfrak{n}_r = \text{nilradical of } \mathfrak{r}. \tag{B.4}$$

In fact, equality holds in (B.4) but this is not essential to know at this point. What counts is that $\bar{\mathfrak{n}}$ is an ideal in $\mathfrak{r}$ (cf. §A.9) and that the NB-condition on $\mathfrak{g}$ means that $\mathfrak{r}$ is NC. Since $\mathfrak{n}_r + \overline{\mathfrak{h}} = \mathfrak{r}$, because of (B.3)–(B.4), we have that the real roots of the ad-action of $\overline{\mathfrak{h}}$ on $\mathfrak{n}_r$, and also on every ideal of $\mathfrak{r}$, satisfy the NC-condition in the sense of the following imprecise but convenient terminology: since $\overline{\mathfrak{h}}$ acts by ad on every ideal $I \subset \mathfrak{r}$, with $I = \mathfrak{n}_r, \mathfrak{n}, \mathfrak{n} + \mathfrak{n}_s, \ldots$, its action on the complexification $I \otimes \mathbb{C}$ admits a root space decomposition and the real parts of the roots, counted with multiplicity, $R_1, \ldots, R_k$, with $k = \dim I$, will be called, as in Chapter 2, the real roots of that action. Furthermore, if it happens that the relation $\sum \alpha_j R_j = 0$, for $\alpha_j \geqslant 0$, implies $\alpha_j R_j = 0$, for $j = 1, \ldots, k$, then we shall say that the NC-condition is satisfied (or simply that 'I is NC'!). From the definition, it is clear that for the two ideals $I_2 \subset I_1 \subset \mathfrak{r}$, the real roots of I_2 are a subset of the real roots of I_1. As a result, I_2 is NC as soon as I_1 is.

This real root space decomposition is exactly what we did in Chapters 2 and 3 (see §2.3.4) for the action of $\operatorname{ad}\mathfrak{h}$ on $\mathfrak{n}$, and if we preserve the notation of §3.8.2 we have

$$\mathfrak{n} = \mathfrak{n}_0 + \mathfrak{n}_1 + \cdots + \mathfrak{n}_r, \quad \mathfrak{n}_R = \mathfrak{n}_1 + \cdots + \mathfrak{n}_r \tag{B.5}$$

for the real root spaces with distinct real roots $L_0 = 0, L_1, \ldots, L_r$ (with L_j non-zero for $j \neq 0$). Here we recall that $\mathfrak{n}_0$, by abuse of notation, could be the zero space if there is no root with real part equal to zero. All the other spaces $\mathfrak{n}_j$ for $1 \leqslant j \leqslant r$ are non-zero. In (B.5) we have

$$[\mathfrak{n}_j, \mathfrak{s}] \subset \mathfrak{n}_j, \quad [\mathfrak{n}_j, \mathfrak{n}_0] \subset \mathfrak{n}_j; \quad 0 \leqslant j \leqslant r. \tag{B.6}$$

This holds because of the commutation relation (B.2). Also we recall here that unless $[\mathfrak{n}_i, \mathfrak{n}_j] = 0$ we have $L_i + L_j = L_k$. For all this, go back to §3.8.2.

B.3 Two Lemmas

With this notation we can now state the following two lemmas.

Lemma B.1 *The real roots L_j, with $0 \leqslant j \leqslant r$, satisfy the* NC*-condition. Explicitly, if we assume that $r \geqslant 1$ and that $\alpha_j \geqslant 0$, for $1 \leqslant j \leqslant r$, are such that $\sum \alpha_j L_j = 0$ then $\alpha_j = 0$, with $1 \leqslant j \leqslant r$.*

What this lemma says is that the radical $\mathfrak{q}$ of an NB-algebra is NC, and gives the proof of Proposition 4.3.

As a consequence of the lemma we can assert that

$$\mathfrak{n}_R = \mathfrak{n}_1 + \cdots + \mathfrak{n}_r \subset \mathfrak{g} \tag{B.7}$$

is a subalgebra (see §3.8.2) and because of (B.6), $\mathfrak{n}_R$ is in fact an ideal. As in Part 3.3, we conclude from the above that when $\mathfrak{g}$ is an NB-algebra then

$$\mathfrak{g} = \mathfrak{n}_R \leftthreetimes [\mathfrak{n}_0 + \mathfrak{h} + \mathfrak{s}]. \tag{B.8}$$

This is exactly what was exploited in Part 3.3 in the case of amenable algebras, that is, algebras for which $\mathfrak{s}$ is of compact type.

As we shall see presently, the converse of the lemma fails and the radical $\mathfrak{q}$ could well be an NC-algebra and yet $\mathfrak{g}$ fail to be NB. (This is an exercise after studying §B.5 below (cf. §4.1). An explicit counterexample is $\mathbb{R}^2 \leftthreetimes \mathrm{SL}(\mathbb{R}^2)$ for the natural action in the semidirect product).

The second lemma (below) refers explicitly to non-amenable Lie algebras.

Lemma B.2 *Let the notation be as before and* $\mathfrak{g}$ *be assumed* NB. *Then we have*

$$[\mathfrak{s}_n, \mathfrak{n}_0] = \{0\}. \tag{B.9}$$

The proofs of Lemmas B.1 and B.2 will be given below in §B.4 but before that we shall use the two lemmas together with what we saw in §3.8.3 to recapitulate and reorganise our notation so that we do not have to come back later. We have

$$\begin{gathered} \mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}, \quad \mathfrak{q} = \mathfrak{n}_R \leftthreetimes (\mathfrak{n}_0 + \mathfrak{h}), \quad \mathfrak{s} = \mathfrak{s}_c \oplus \mathfrak{s}_n, \\ [\mathfrak{n}_R, \mathfrak{s}] \subset \mathfrak{n}_R, \quad [\mathfrak{n}_0, \mathfrak{s}] \subset \mathfrak{n}_0, \\ [\mathfrak{h}, \mathfrak{s}] = 0, \quad [\mathfrak{n}_0, \mathfrak{s}_n] = 0. \end{gathered} \tag{B.10}$$

We shall then define

$$\mathfrak{g}_R = (\mathfrak{n}_0 + \mathfrak{h}) \leftthreetimes \mathfrak{s}_c, \quad \mathfrak{m} = \mathfrak{g}_R \oplus \mathfrak{s}_n, \tag{B.11}$$

where the radical of $\mathfrak{g}_R$ is $\mathfrak{q}_R = \mathfrak{n}_0 + \mathfrak{h}$ is an R-algebra (by the definition of $\mathfrak{n}_0$; see §§2.2.2, 3.8.2). This implies that

$$\mathfrak{g}_R = \mathfrak{q}_R \leftthreetimes \mathfrak{s}_c \text{ is an R-algebra.} \tag{B.12}$$

With this notation (B.8) becomes

$$\mathfrak{g} = \mathfrak{n}_R \leftthreetimes (\mathfrak{g}_R \oplus \mathfrak{s}_n) = \mathfrak{n}_R \leftthreetimes \mathfrak{m}. \tag{B.13}$$

This final splitting, we shall presently show, may be used to give a complete characterisation of NB-algebras. Note also that, as we observed in §3.8, $\mathfrak{n}_R$ is uniquely determined and is independent of the choice of $\mathfrak{h}$.

B.4 Proof of the Lemmas

Since $[\mathfrak{n}_j, \mathfrak{s}] \subset \mathfrak{n}_j$ $(0 \leqslant j \leqslant r)$ by §A.10, the Abelian algebra $\mathfrak{a}$ acts (by 'acts' we refer, of course, both here and throughout to the adjoint action) diagonally on each $\mathfrak{n}_j$ and the roots of that action $\lambda_1^j, \ldots, \lambda_{p_j}^j$ are real $\lambda_i^j \in \mathfrak{a}^*$, where $p_j = \dim \mathfrak{n}_j$. Furthermore, the trace vanishes, as in Lemma A.1:

$$\sum_i \lambda_i^j = 0; \quad 0 \leqslant j \leqslant r. \tag{B.14}$$

Here, to account for multiple roots, repetition is allowed in the enumeration of these roots.

Using Jordan–Hölder we see that the real roots (i.e. the real parts of the roots as in §B.2) of the action of the nilpotent algebra $\overline{\mathfrak{h}} = \mathfrak{h} + \mathfrak{a}$ on $\mathfrak{n}$ are then

$$\mu_i^j = L_j + \lambda_i^j \in \overline{\mathfrak{h}}^* = \mathfrak{h}^* + \mathfrak{a}^*; \quad 0 \leqslant j \leqslant r,\ 1 \leqslant i \leqslant p_j. \tag{B.15}$$

Assume that $\alpha_j \geqslant 0$, with $1 \leqslant j \leqslant r$ (we omit here the case $j = 0$), are such that $\sum_j \alpha_j p_j L_j = 0$; then

$$\sum_j \sum_{i=1}^{p_j} \alpha_j \mu_i^j = \sum_j \alpha_j p_j L_j + \sum_j \alpha_j \sum_{i=1}^{p_j} \lambda_i^j = 0. \tag{B.16}$$

The vanishing in (B.16) is guaranteed by the hypothesis on the α_j and (B.14). Now, as explained in §B.2, the NB-condition on $\mathfrak{g}$ implies the NC-condition on the μ_i^j and this, together with (B.16), gives $\alpha_j = 0$, $1 \leqslant j \leqslant r$, because by definition $L_1, \ldots, L_r \neq 0$. This is what is stated in the conclusion of Lemma B.1.

The same argument applies to the ad-action of $\mathfrak{a}$ on $\mathfrak{n}_0$ when this space is not zero. The roots of that action are then $\lambda_1^0, \ldots, \lambda_{p_0}^0 \in \mathfrak{a}^*$ and the NB-condition on $\mathfrak{g}$ and (B.14) imply that all the λ_i^0 are zero. This in turn, by §A.10, again implies that the ad-action of $\mathfrak{a}$ on $\mathfrak{n}_0$ and also the action of the whole of $\operatorname{ad}\mathfrak{s}_n$ is trivial. This proves the required (B.9).

Notice that only the trace was used in the above argument, and therefore the explicit use of Jordan–Hölder that we made to identify the roots in (B.15) is not essential.

B.5 The Unimodular Case

We recall that a Lie algebra $\mathfrak{g}$ is called unimodular if

$$\operatorname{trace}(\operatorname{ad} x) = 0; \quad x \in \mathfrak{g}. \tag{B.17}$$

Exercise (i) Prove that an algebra $\mathfrak{g}$ is unimodular if and only if any Lie group that corresponds to $\mathfrak{g}$ is unimodular, that is, left and right Haar measures coincide (see for example Bourbaki, 1963, Chapter 7 or Helgason, 1984, Chapter 1, Lemma 1.4).

Exercise (ii) R-algebras (see §2.2.2) are unimodular because

$$\begin{aligned}\operatorname{trace}\operatorname{ad} x &= \sum (\text{characteristic roots of } \operatorname{ad} x)\\ &= \sum (\text{real parts of characteristic roots of } \operatorname{ad} x),\end{aligned}$$

since $\operatorname{ad} x$ is a real transformation.

Exercise (iii) Semisimple algebras $\mathfrak{s}$ are unimodular and $\operatorname{trace}\operatorname{ad} x = 0$ with $x \in \mathfrak{s}$ because we have $[\mathfrak{s}, \mathfrak{s}] = \mathfrak{s}$.

Exercise (iv) If $\mathfrak{g}$ is unimodular its radical $\mathfrak{q}$ has to be unimodular (see Bourbaki, 1963, Chapter 7, §2.6; in fact, it is 'if and only if', but we do not immediately need this).

Exercise (v) The soluble algebra $\mathfrak{q}$ is unimodular if and only if in some (or equivalently all) triangulation of the $\operatorname{ad}\xi$ action on $\mathfrak{q} \otimes \mathbb{C}$, with $\xi \in \mathfrak{q}$, the sum of the roots (which are the diagonal elements of the triangular matrices) vanishes (cf. §2.3.3). In particular, if $\mathfrak{q}$ is NC it has to be an R-algebra.

Conclusion An algebra $\mathfrak{g}$ is NB and unimodular if and only if $\mathfrak{g} = \mathfrak{g}_R \oplus \mathfrak{s}$, where $\mathfrak{g}_R$ is an R-algebra and $\mathfrak{s}$ is semisimple.

The 'if' part follows from Exercises (ii), (iii).

The 'only if' follows from (B.8) because by the unimodularity we can conclude that $\mathfrak{n}_R = \{0\}$ because, by Exercises (iv), (v) and Lemma B.1, the radical $\mathfrak{q}$ in (B.10) is an R-algebra.

B.6 Characterisation of Non-unimodular NB-Algebras. The Necessary Condition

Let us now assume that $\mathfrak{g}$ is an NB-algebra that is not unimodular and let the notation be as before. Then by Exercise (iv), $\mathfrak{q}$ is not unimodular and $\mathfrak{n}_R \neq 0$.

Putting together our previous considerations and using the Hahn–Banach argument of §2.3.1 we can say the following. *In the decomposition* (B.8), (B.13) *we saw that we could find* $\xi \in \mathfrak{q}_R$ *such that the characteristic roots of* $\operatorname{ad}\xi$ *acting on* $\mathfrak{n}_R$*, namely* $\lambda_1, \dots, \lambda_n$*, with* $n = \dim \mathfrak{n}_R \geqslant 1$*, counted with multiplicity, satisfy*

$$\operatorname{Re}\lambda_j > 0; \quad 1 \leqslant j \leqslant n. \tag{B.18}$$

B.7 The Sufficiency of the Condition

Property (B.18) will be used in Appendix F. In the rest of this appendix we shall examine a 'converse' of that property.

In the proofs that follow we shall be brief because this is something that is neither important nor used again.

Hypothesis (H) Let $\mathfrak{g}$ be some arbitrary Lie algebra that can be written as

$$\mathfrak{g} = \mathfrak{n}_R \leftthreetimes (\mathfrak{g}_R \oplus \mathfrak{s}), \tag{B.19}$$

where $\mathfrak{n}_R$ is nilpotent, $\mathfrak{g}_R$ is an R-algebra and $\mathfrak{s}$ is semisimple or 0. Clearly, by splitting $\mathfrak{s} = \mathfrak{s}_c + \mathfrak{s}_n$ of (B.10) we can absorb $\mathfrak{s}_c$ with $\mathfrak{g}_R$ and assume in (B.19) that $\mathfrak{s} = \mathfrak{s}_n = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k}$ as in (B.1). We shall further assume that the following condition is satisfied.

There exists $\xi \in \mathfrak{q}_R =$ the radical of $\mathfrak{g}_R$ that satisfies one and thus both, of the following equivalent conditions.

(H) The roots of $\operatorname{ad}\xi|_{\mathfrak{n}_R}$ $\lambda_1, \dots, \lambda_n$ (counted with multiplicity $n = \dim \mathfrak{n}_R \geqslant 1$) satisfy $\operatorname{Re}\lambda_j > 0$, with $j = 1, \dots, n$.

(H′) $\|\operatorname{Ad}(e^{-t\xi})|_{\mathfrak{n}_R}\| \leqslant Ce^{-ct}$, with $t > 0$, for some $C, c > 0$.

Under these conditions $\mathfrak{g} \neq 0$ and $\mathfrak{g}$ is not unimodular. For the equivalence (H) $\Leftrightarrow$ (H′) we use the fact that $\mathfrak{q}_R$ is soluble. This will be left as an exercise because it is essentially contained in §3.9.1.

The sufficiency We shall show that under (H), *the algebra* $\mathfrak{g}$ *is an* NB*-algebra*.

There are two reasons for the second formulation (H′).

First, (H′) can be weakened in a natural way and we could impose instead the following condition.

(WH) There exists $\xi \in \mathfrak{g}_R$ (and not necessarily in the radical) for which (H′) holds.

The *conclusion* is the same: (WH) $\Rightarrow$ (NB).

The point is that since in general $\mathfrak{g}_R$ is not soluble, any argument that involves characteristic roots has to be done with care because we cannot use Lie's theorem. As a result, the norm of $\mathrm{Ad}(e^{-t\xi})$ is easier to handle.

The second reason is more speculative in nature but it is worth recording. Condition (WH) says that the 'flow induced by the 1-parameter group $e^{t\xi}$ is contracting'. This suggests connections with the theory of dynamical systems. This point of view, however, will not be pursued.

At any rate, putting together §§B.5–B.7 we see that we have a complete characterisation of NB-algebras. This will turn out to be useful in Appendix F on geometric theory.

B.8 Proof of the Sufficiency

The notation is as in (B.19) and we assume that condition (H) is satisfied. We denote by $\mathfrak{q}_R$ the radical of $\mathfrak{g}_R$ and, as before, by $\mathfrak{q}$ and $\mathfrak{n}$, respectively, the radical and nilradical of $\mathfrak{g}$. The following then holds:

$$\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R, \quad \mathfrak{n}_R \subset \mathfrak{n} \subset \mathfrak{q},$$
$$\mathfrak{n} = \mathfrak{n}_R \leftthreetimes (\mathfrak{n} \cap \mathfrak{q}_R) = \mathfrak{n}_R \leftthreetimes \hat{\mathfrak{n}}, \quad \hat{\mathfrak{n}} = \mathfrak{n} \cap \mathfrak{q}_R,$$
$$\mathfrak{q}_R/\hat{\mathfrak{n}} = \mathfrak{q}/\mathfrak{n} \text{ is Abelian.}$$

The easy verification will be left to the reader.

In the proof below, as we shall see, $\hat{\mathfrak{n}}$ will play a role analogous to that of $\mathfrak{n}_0$ in (B.5).

A nilpotent subalgebra $\mathfrak{h} \subset \mathfrak{q}_R$, say a Cartan subalgebra, can now be found so that $\mathfrak{q}_R = \hat{\mathfrak{n}} + \mathfrak{h}$ and therefore also $\mathfrak{q} = \mathfrak{n} + \mathfrak{h}$. This is possible because of Chevalley (1955, §VI.4.5), as in §3.4. To avoid confusion with this notation, we stress the point that this nilpotent algebra has nothing to do with the $\mathfrak{h}$ that we fixed in §B.2 and which was used to define $\mathfrak{q}_R$.

What we can say about these subalgebras is this:

(i) $[\mathfrak{h}, \hat{\mathfrak{n}}] \subset \hat{\mathfrak{n}}$ and the real roots (in the sense of §B.2) of the action of the nilpotent algebra $\mathfrak{h}$ on $\hat{\mathfrak{n}}$ are zero. This holds because $\mathfrak{q}_R$ is an R-algebra.

(ii) Similarly, $[\mathfrak{h}, \mathfrak{n}_R] \subset \mathfrak{n}_R$ and the real roots $L_1, \ldots, L_n$, with $n = \dim \mathfrak{n}_R$, of the action of $\mathfrak{h}$ on $\mathfrak{n}_R$ counted with multiplicity, satisfy the following NC-condition: $\sum \alpha_j L_j = 0$; $\alpha_j \geqslant 0$, with $1 \leqslant j \leqslant n$, imply that $\alpha_j = 0$, with $1 \leqslant j \leqslant n$. In particular none of the L_j is zero.

We shall show at the end of this section that (ii) follows from condition (H). But before this we shall complete the proof of the sufficiency.

Proof that $\mathfrak{g}$ is NB under condition (H) In the argument that follows, $\mathfrak{s}_n$ could be $\{0\}$ but as already said, $\mathfrak{g}_R$, $\mathfrak{n}_R$ do not vanish. The Iwasawa radical that corresponds then to some decomposition of $\mathfrak{s}_n = \mathfrak{n}_s + \mathfrak{a} + \mathfrak{k}$ in $\mathfrak{g} = \mathfrak{n}_R \rightthreetimes (\mathfrak{g}_R \oplus \mathfrak{s}_n)$ is

$$\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_s + \mathfrak{a} = \overline{\mathfrak{n}} + \overline{\mathfrak{h}}; \quad \overline{\mathfrak{n}} = \mathfrak{n} + \mathfrak{n}_s; \quad \overline{\mathfrak{h}} = \mathfrak{h} \oplus \mathfrak{a}.$$

Now we shall need to use the stronger fact in §A.9 that $\overline{\mathfrak{n}} = \mathfrak{n}_r \subset \mathfrak{r}$ is the nilradical of $\mathfrak{r}$. We need to show that the real roots of the action of $\overline{\mathfrak{h}}$ on $\overline{\mathfrak{n}}$ satisfy the NC-condition (in the sense of §B.2). This will complete the proof that $\mathfrak{g}$ is an NB-algebra.

By the obvious Jordan–Hölder argument applied to the ad-action of $\overline{\mathfrak{h}}$ on $\mathfrak{n}$ and $\overline{\mathfrak{n}}/\mathfrak{n}$, the real roots $R_1, R_2, \ldots \in (\overline{\mathfrak{h}})^*$ of the $\operatorname{ad}\overline{\mathfrak{h}}$-action on $\overline{\mathfrak{n}}$ are obtained by putting together the following two sets of real functionals on $\overline{\mathfrak{h}}$:

(a) The real roots of the action of $\operatorname{ad}\overline{\mathfrak{h}}$ on $\mathfrak{n}$. These roots will be written $R_1, \ldots, R_m$, counted with multiplicity, with $m = \dim \mathfrak{n}$.
(b) The real roots of the induced action $\operatorname{ad}\overline{\mathfrak{h}}$ on $\overline{\mathfrak{n}}/\mathfrak{n} = \mathfrak{n}_s$. The roots will be written $R_{m+1}, \ldots, R_p$, with $p = \dim \overline{\mathfrak{n}}$.

Recall that $\overline{\mathfrak{h}} = \mathfrak{h} \oplus \mathfrak{a}$ and for the restrictions of these roots (remember that they are elements of $(\overline{\mathfrak{h}})^*$) to the two subspaces $\mathfrak{h}$ and $\mathfrak{a}$ we can assert the following.

For the first class in (a) (after possible reordering) we have

$$\begin{gathered} R_j\big|_{\mathfrak{h}} = L_j; \quad j = 1, \ldots, n \ (n = \dim \mathfrak{n}_R), \\ R_{n+1} = \cdots = R_m = 0, \end{gathered} \tag{B.20}$$

where $L_1, \ldots$ are the real roots of the $\operatorname{ad}\mathfrak{h}$-action on $\mathfrak{n}_R$ as in property (ii). To see the vanishing of the remaining roots in (B.20), observe that in the direct vector space decomposition $\mathfrak{n} = \mathfrak{n}_R + \hat{\mathfrak{n}}$ we have $[\overline{\mathfrak{h}}, \mathfrak{n}_R] \subset \mathfrak{n}_R$ and $[\overline{\mathfrak{h}}, \hat{\mathfrak{n}}] = [\mathfrak{h}, \hat{\mathfrak{n}}] + [\mathfrak{a}, \hat{\mathfrak{n}}]$. Now $[\mathfrak{a}, \hat{\mathfrak{n}}] \subset [\mathfrak{s}_n, \mathfrak{q}_R] = 0$ by (B.19). As for the ad-action of $\mathfrak{h}$ on $\hat{\mathfrak{n}}$, it has pure imaginary roots since $\mathfrak{q}_R$ is an R-algebra and $\mathfrak{h}, \hat{\mathfrak{n}} \subset \mathfrak{q}_R$.

For the roots in class (b), using $[\mathfrak{h}, \overline{\mathfrak{n}}] \subset [\mathfrak{q}, \mathfrak{g}] \subset \mathfrak{n}$ we deduce that

$$R_j\big|_{\mathfrak{h}} = 0; \quad j = m+1, \ldots, p. \tag{B.21}$$

Furthermore, the $R_j\big|_{\mathfrak{a}}$, with $j = m+1, \ldots, p$, can be identified with $\mu_1, \ldots$ the roots of the ad-action of $\mathfrak{a}$ on $\mathfrak{n}_s$ of §A.3(ii) – in the case $\mathfrak{s}_n = 0$, (B.21) says that all the R_j are zero.

Now let $\alpha_j \geqslant 0$ be such that

$$\sum_{j=1}^{m} \alpha_j R_j + \sum_{j=m+1}^{p} \alpha_j R_j = 0. \tag{B.22}$$

By restricting this relation to $\mathfrak{h}$ we deduce from property (ii), (B.20), (B.21) that $\alpha_1 = \cdots = \alpha_m = 0$ and therefore also $\alpha_j R_j = 0$, with $1 \leqslant j \leqslant m$. This new information once inserted back in (B.22), together with the NC-property of the roots $\mu_1, \ldots$ of §A.3(ii), imply that $\alpha_j = 0$, with $j = m+1, \ldots, p$. So from (B.22) we have concluded that all the $\alpha_j R_j = 0$. The bottom line is that we have proved that the Iwasawa radical $\mathfrak{r}$ is an NC-algebra. In other words, we have completed the proof that $\mathfrak{g}$ is an NB-algebra. Finally, we give the following proof.

Proof of property (ii) This is a consequence of Lie's theorem and the triangulation of the action of $\operatorname{ad}\mathfrak{q}$ on $\mathfrak{n}_R \otimes \mathbb{C}$ (see Varadarajan, 1974, §3.7.3). Let $\varphi_1, \ldots, \varphi_n \in \operatorname{Hom}_{\mathbb{R}}[\mathfrak{q}; \mathbb{C}]$, with $n = \dim \mathfrak{n}_R$, denote the diagonal elements in that triangulation. Then for any $\xi \in \mathfrak{q}$ the complex roots of $\operatorname{ad}\xi$ for that action are $\varphi_1(\xi), \ldots, \varphi_n(\xi)$.

Now let ξ be as in condition (H). We can then write $\xi = \chi + \psi$, with $\chi \in \mathfrak{n}$, and $\psi \in \mathfrak{h}$ (not necessarily uniquely) and since $\operatorname{ad}\chi$ is a nilpotent transformation on $\mathfrak{n}_R$, $\varphi_j(\chi) = 0$ $(1 \leqslant j \leqslant n)$ the conclusion is that $\varphi_j(\psi) = \lambda_j$ $(1 \leqslant j \leqslant n)$ for the characteristic roots λ_j of $\operatorname{ad}\xi$ as in (H) and therefore by condition (H) on ξ, we have $\operatorname{Re}\varphi_j(\psi) > 0$. But the existence of such a $\psi \in \mathfrak{h}$ for which this holds clearly implies condition (ii).

B.9 Additional Remarks on the Sufficiency of the Condition

In the geometric theory in Appendix F of Part II, we shall show again by different means that condition (H′), in fact the even weaker condition (WH), suffices to guarantee that the simply connected group $\widetilde{G}$ that corresponds to $\mathfrak{g}$ is NB.

This geometric proof, in accordance with the speculations of §B.7, has a strong dynamical system flavour. This geometric proof, a posteriori, will be seen to be the correct way of going about things.

We shall see, in particular there, that it is not the fact that $\xi \in \mathfrak{g}_R$, as stipulated in condition (H), that really matters. What counts is that $\xi \in \mathfrak{m}$ (of (B.11)) and that ξ has the exponential contraction property (H′) for the flow $g \to ge^{t\xi}$, with $g \in \widetilde{G}$, together with the fact that $|(\operatorname{Ad} n)\xi| \leqslant C(1 + |n|_M)^C$, with $n \in M$. Both these conditions are of a dynamical nature.

Despite this, if we insist on giving an algebraic proof that (WH) $\Rightarrow$ (NB), with the notation of Appendix C below, we can go about it as follows. (Only a sketch of the idea will be given here and the reader is invited to treat this as a – difficult but instructive – exercise and write down a proof.)

When $\xi \in \mathfrak{g}$ and $g = e^{\xi} \in \widetilde{G}_R$ then $g = qs$, $q \in \widetilde{Q}_R \widetilde{S}_c$ (these are the simply connected groups that correspond to $\mathfrak{q}_R$ and $\mathfrak{s}_c$ as in §C.1 below). Then as in the random walk argument of §2.7.1 we can write

$$(qs)^n = (q_1 s_1)(q_2 s_2) \cdots (q_n s_n) = q_1 q_2^{s_1} \cdots q_n^{s_1 \cdots s_{n-1}} s_1 s_2 \cdots s_n = \tilde{q}_1 \cdots \tilde{q}_n s^n = q_n^* s^n,$$

with $q_1 = q_2 = \cdots = q$, $s_1 = s_2 = \cdots = s$. The compactness of $\tilde{S}_c$ is then used together with the fact that $\tilde{q}_1, \ldots, \tilde{q}_n \in \widetilde{Q}_R$, which is a soluble group. Lie's theorem can thus be used to triangulate $\operatorname{Ad} \tilde{q}_j \big|_{\mathfrak{n}_R}$ and they all have the same eigenvalues as those of $\operatorname{Ad} q$. It follows that the eigenvalues of $\operatorname{Ad} q_n^*$ on $\mathfrak{n}_R$ are $r_1^n, \ldots$ (where $r_1, \ldots$ are the eigenvalues of $\operatorname{Ad} q$); see (3.62).

This together with the (WH) condition can be used as a substitute for condition (H) in the proof of property §B.8(ii). Indeed, condition (WH) together with the above considerations shows that $|r_j| < 1$ for the eigenvalues.

Appendix C
The Structure of NB-Groups

All the notation from Appendices A and B will be preserved and, throughout here, $\mathfrak{g}$ will be some NB-algebra.

C.1 Simply Connected Groups and Their Centres

Here $\widetilde{G}, N_R, \widetilde{G}_R, \widetilde{S}_n, \widetilde{M}$ will denote the simply connected groups that correspond to the algebras $\mathfrak{g}, \mathfrak{n}_R, \mathfrak{g}_R, \mathfrak{s}_n, \mathfrak{m}$ of §B.3 respectively. Clearly (B.13) induces

$$\widetilde{G} \cong N_R \leftthreetimes \widetilde{M}; \quad \widetilde{M} \cong \widetilde{G}_R \oplus \widetilde{S}_n,$$

and the groups N_R and $\widetilde{M}$ can be identified to subgroups of $\widetilde{G}$.

To obtain the global structure of a general connected group G with Lie algebra $\mathfrak{g}$ we shall use the following result.

Lemma *The centre $Z = Z(\widetilde{G})$ of $\widetilde{G}$ is contained in $\widetilde{M}$: $Z \subset \widetilde{M}$.*

To prove this observe that since N_R is a simply connected nilpotent group, every element $x \in N_R$ can be expressed uniquely by $x = \exp \xi$, where $\xi \in \mathfrak{n}_R$ will be denoted by $\xi = \log x$ (see Varadarajan, 1974, §§2.1.3, 3.6). Furthermore, by the definitions

$$\begin{aligned} &\log x^a = (\operatorname{Ad} a)(\log x) = e^{\operatorname{ad}\zeta} \log x; \\ &a = \exp(\zeta),\ a \in \widetilde{G},\ \zeta \in \mathfrak{g},\ x^a = axa^{-1}. \end{aligned} \tag{C.1}$$

Now let $z \in Z$ and let us write (uniquely) $z = xm$, with $x \in N_R, m \in \widetilde{M}$. We shall show that $x = e$, the identity. This will complete the proof.

To see this let $a = \exp \zeta \in \widetilde{M}$ for some $\zeta \in \mathfrak{m}$. Then since $z = z^a = x^a m^a = xm$ we have $x^a = x$ (and $m^a = m$ because $m, a \in \widetilde{M}$). This together with (C.1) implies, by the uniqueness of the logarithm, that $e^{\operatorname{ad}\zeta}(\log x) = (\log x)$. If $x \neq e$

this means that $\log x \in n_R$ is an eigenvector with eigenvalue 1 for all $e^{\mathrm{ad}\,\zeta}$, where $\zeta \in \mathfrak{m}$. But this contradicts the condition (B.18) unless $x = e$. This completes the proof.

C.2 A General NB-Group

In this section, the notation $\oplus$ denotes the direct product of groups. The conclusion from the previous section is that if G is an arbitrary connected NB-group then there exists $Z \subset \widetilde{M}$ a discrete subgroup in $\widetilde{G}$ that is central in $\widetilde{G}$ such that

$$G = \widetilde{G}/Z = N_R \leftthreetimes M; \quad M = \widetilde{M}/Z. \tag{C.2}$$

If we use Iwasawa decomposition in §A.3 to write $\widetilde{M} = \widetilde{G}_R \oplus KAN_s$, where $KAN_s = \widetilde{S}_n$, it follows that Z lies in the subgroup $\widetilde{G}_R \oplus K$. (To see this we project Z on the component KAN_s; since that projection is central, it lies in K by §A.3(iii).)

This means that M can be written as a product

$$M = \widetilde{M}/Z = LAN_s = AN_sL; \quad L = \frac{\widetilde{G}_R \oplus K}{Z}, \tag{C.3}$$

where the representation $M \ni m = anl$, with $a \in A$, $n \in N_s$ and $l \in L$, is one-to-one, and it induces a diffeomorphism. The group L is clearly an R-group. The subgroups L, A, N (we drop the index s because no confusion can arise) are not normal in M and this makes the representation (C.3) awkward to work with. The one conclusion that can be drawn from (C.3), however, is the following.

Maximal compact subgroups *Let $U \subset L$ be some maximal compact subgroup (necessarily connected) and let V be some maximal compact subgroup of G. Then* $\dim U = \dim V$.

To see this we invoke the general theory of maximal compact subgroups (see Hochschild, 1965, §XV.3; Helgason, 1978, §VI.2, and also Chapter 12 below) and the fact that $AN \simeq \mathbb{R}^s$ (a diffeomorphism) to deduce that we have, diffeomorphically,

$$G = U \times \mathbb{R}^p \simeq V \times \mathbb{R}^q; \quad p, q \geqslant 0. \tag{C.4}$$

Expression (C.4) implies that $\dim U = \dim V = n$ and $p = q$ because n is the largest index for which the real cohomology $H^n(G, \mathbb{R})$ does not vanish.

An immediate consequence of this and the general theory (see Hochschild, 1965, Chapter XV, Theorem 3.1(iii)) is that some conjugate of $V \subset L$ and that U is a maximal compact subgroup of G.

C.3 The Quasi-Isometric Modification

Apart from (C.4), the group decomposition (C.3) is not very useful in itself. To get a more useful decomposition of M, at least geometrically, we can resort to a device that will be used systematically later on in the geometric theory.

The idea consists in modifying the group M and constructing a new 'replica' group M_1, together with some diffeomorphism $\varphi\colon M \to M_1$ which, although not a group isomorphism, preserves important features of the structures. Here it is the left-invariant Riemannian structure that will be preserved, in the sense that $|d\varphi|, |d\varphi^{-1}| \leqslant C$, that is, φ is a quasi-isometry in the sense of Chapter 11. Once we set our mind to doing this the actual construction of M_1 is not very difficult.

Starting from $\widetilde{S}_n = N_s AK$ we first construct $S^* = N_s A \oplus K$ and then define $M^* = \widetilde{G}_R \oplus S^*$, the direct product of the two groups. This is a new group that is in a natural bijective correspondence $\tilde{\varphi}\colon \widetilde{M} \to M^*$ with M and $\tilde{\varphi}$ has the property $|d\tilde{\varphi}|$, $|d\tilde{\varphi}^{-1}| \leqslant C$ for the two left-invariant Riemannian structures. This is the pivot of course. The reader is invited to try to prove this for themselves before checking it out in Example 11.9 below, where the proof is seen to be a consequence of a more general principle. What is used there is the fact that $\widetilde{G}_R \times N_s AZ(K)$ can be identified with a cocompact subgroup of both $\widetilde{M}$ and M^*.

Now we observe that $\tilde{\varphi}$ identifies the subgroup Z with $Z^* = \tilde{\varphi}(Z) \subset M^*$, which is also a central discrete subgroup in the new group M^*. This holds because $Z(\widetilde{M}) = Z(\widetilde{G}_R) \times Z(\widetilde{S}_n)$ and $Z(\widetilde{S}_n) = Z(K)$ (here we use the notation $Z(\cdot)$ for the centre of a group and we also use the fact §A.3(iv)). As a consequence, the identification $\tilde{\varphi}$ intertwines the left action of Z on $\widetilde{M}$ with the left action of Z^* on M^*. As a consequence, we deduce that the induced mapping

$$\varphi\colon M = LNA = \widetilde{M}/Z \xrightarrow{\cong} \left[\frac{\widetilde{G}_R \oplus K}{Z}\right] \oplus NA = L \oplus NA \tag{C.5}$$

is also a diffeomorphism and satisfies $|d\varphi|, |d\varphi^{-1}| \leqslant C$ for left-invariant Riemannian metrics.

Note that when the group G is unimodular (and NB) then $N_R = \{0\}$ and then (C.5) applies to $G = M$.

Appendix D

Invariant Differential Operators and Their Diffusion Kernels

D.1 Definitions and Notation

Let $X = R \times K$ be some principal bundle as in Chapter 4, where R is some connected Lie group and K some compact C^∞ manifold which, to fix ideas, we shall assume to be a Lie group. We shall use all the notation of Chapters 4 and 5 and denote by dk the Haar measure on K and dr, $d^r r$ the left and right Haar measures on R.

We shall consider on X second-order subelliptic operators D. This will mean here that locally, modulo lower-order terms, these operators can be represented by $-\sum Y_j^2$, the sum of squares of vector fields that satisfy the Hörmander condition (i.e. they, together with their successive brackets, span the tangent space). The operators that we shall consider in this appendix will be invariant by the left action of R on the bundle X.

When $G = RK$ is a Lie group that is identified with the bundle $X = R \times K$, then the invariant subelliptic Laplacians $\Delta = -\sum Y_j^2$ on G that we have considered in the previous chapters give rise to invariant differential operators on X.

For an appropriate closure of D we shall consider the semigroup $T_t = e^{-tD}$ and the heat diffusion kernel $T_t f(x) = \int f(y) p_t(x, dy)$; see Yosida (1970, Chapter IX). In the situation that we shall examine in this appendix, we shall always have $p_t(x, dy) = p_t(x, y)\, dy$ for dy some appropriate measure on X and some smooth kernel $p_t(x, y)$, with $t > 0, x, y \in X$.

The most natural measure to consider on X is $dr\, dk$ because among other things it is left invariant by R-action and therefore the kernel $p_t^{(l)}(x, y)$ of T_t with respect to that measure is left invariant. This also is the measure that is identified with left Haar measure dy in the identification with $G = RK$ (cf. §4.4.6).

The main thing that will be done in this appendix is to prove the following upper Gaussian estimate.

For all $0 < a < b$ *there exist* $C, c > 0$ *such that*

$$p_t^{(l)}(x,y) \leqslant C\exp(-cd^2(x,y)); \quad x,y \in X,\ t \in [a,b]. \tag{D.1}$$

Here and throughout $d(\cdot,\cdot)$ denotes some fixed R-invariant distance on X which is locally given by Euclidean distance on the compact neighbourhoods of the manifold.

The above measure is not the most convenient to use on X. This we already saw in (4.23) where, always with the identification $G \simeq RK$, we had $d^r g = \chi^2 d^r r\, dk$. To be systematic we shall fix notation, write $dx = d^r r\, dk$, with $x = (r,k) \in R \times K$, and consider a whole family of measures of the form $\alpha^2(x)\, dx$ for strictly positive smooth functions of the form $\alpha(x) = \theta(r)\varphi(k)$, where θ is multiplicative, that is, $\theta(r_1 r_2) = \theta(r_1)\theta(r_2)$. The reason why we confine ourselves to such functions is simple. Indeed, let L^* be the formal adjoint of an arbitrary operator L with respect to an arbitrary measure $d\sigma$. Then the formal adjoint of the operator $\alpha^{-1}L\alpha$ (the conjugation by pointwise multiplication by $f \to \alpha f$) with respect to $\alpha^2\, d\sigma$ is $\alpha^{-1}L^*\alpha$. Therefore if the above conjugation is to preserve R-left invariance, α has to be of the above form. Note also the following. Let α be as above and let Y be some R-invariant vector field on X. Also let Y^* be the formal adjoint of the first-order differential operator Y with respect to dx. Then the formal adjoint with respect to $\alpha^2\, dx$ is $Y^* + a(k)$ for some smooth $a \in C^\infty(K)$.

The Markovian property In the above definitions D is a Markovian generator if $D1 = 0$, that is, if the differential operator D has no constant term, and then the semigroup generated $T_t = e^{-tD}$ is Markovian. This property is lost by the conjugation $\alpha^{-1}D\alpha$ by one of the functions that we considered above. But we can recover that property if α is an eigenfunction and $T_t\alpha = e^{-\lambda t}\alpha$. Then $e^{-\lambda t}T_t$ is Markovian. This is what we did in §5.3.

Since we shall only be considering here a finite time interval, $t \in [c,b]$, this particular choice of α will not be essential. Observe, however, that by the R-invariance $D1$ is always a bounded smooth function and therefore $D1 + a \geqslant 0$ for some $a > 0$ and the semigroup $e^{-ta}e^{-tD}$ is sub-Markovian. This guarantees the positivity of the kernel $p_t(x,dy)$. This kernel, as we shall see in the next section, will be smooth.

D.2 The Harnack and the Gaussian Estimates

Since we are not aiming for generality, we shall make throughout the assumption that, globally, $D = -\sum Y_j^2$ for R-invariant vector fields that satisfy the

Hörmander condition. We shall also assume that for one of the measures considered in the previous section, namely $\check{d}x = \alpha^2\, dx$, the fields are skew adjoined, that is, $Y_j^* = -Y_j$. This assumption is certainly satisfied in the group case, $G = R \cdot K$ with $\Delta = -\sum Y_j^2$ as in §D.1 for the measure $d^r g = \chi^2\, dx$ (cf. (4.23)). The smoothness of the kernel $u = p_t(x,y)$ with respect to $\check{d}x$ is then clear because it is symmetric and therefore satisfies the subelliptic equation $[2(\partial/\partial t) - D_x - D_y]u = 0$.

Using the general theory (see Varopoulos et al., 1992 and the references given there) we also see that the kernel with respect to some measure $\alpha^2\, dx$ as above will satisfy the following Harnack estimate:

Let $0 < a_1 < b_1 < a_2 < b_2$ *and* $c > 0$ *be given. Then there exists* $C > 0$ *such that*

$$\begin{aligned} &p_{t_1}(x_1,y_1) \leqslant C p_{t_2}(x_2,y_2); \quad x_1,x_2,y_1,y_2 \in X, \\ &a_1 < t_1 < b_1,\ a_2 < t_2 < b_2,\ d(x_1,x_2) < c,\ d(y_1,y_2) < c. \end{aligned} \tag{D.2}$$

This holds for any other choice of the measure $\beta^2\, dx$ (for any other function β as in §D.1) since we can clearly pass from one to another by multiplying this kernel by $\alpha^2(y)/\beta^2(y)$ and use again the special nature of the weight functions α and β.

The uniformity in x_1, y_1 is the only new point that needs checking. But this is clear by R-invariance of D and the special nature of the weights α. Indeed, these imply that there exists a multiplicative positive character χ on R such that $p(rx,ry) = \chi(r)p(x,y)$, with $r \in R$, because this is what happens to the measures that we are considering when we translate them by $r \in R$. This generalises the remark that we already used in §5.4.

Now we shall stick to the previous measure $\check{d}x = \alpha^2\, dx = \alpha^2\, d^r r\, dk$ involved in our assumption on D and consider $p_t(x,y)$, the kernel of the operator $\alpha e^{-tD} \alpha^{-1}$ with respect to dx. Using this special measure and standard arguments we shall prove the following Gaussian estimate:

Let $t \in [a,b]$ *be some bounded fixed interval. Then there exist* $C, c > 0$ *such that*

$$p_t(x,y) \leqslant C \exp\left(-cd^2(x,y)\right) m_R^{-1}(y); \quad x,y \in X,\ t \in [a,b], \tag{D.3}$$

where m_R *denotes the modular function on* R*, that is,* $d^r r = m_R(r)\, dr$.

This estimate is uniform in x, y but not in t. By passing to the left measure dr we obtain the original estimate (D.1) that we set ourselves the task of proving.

Observe finally that conjugating the semigroup e^{-tD} by α, as we did, makes no difference to the Gaussian estimate. The reason is that the effect of this conjugation is to multiply the kernel by $\alpha(xy^{-1})$. This factor is in turn absorbed by the Gaussian factor because of the multiplicative nature of θ in the definition $\alpha = \theta\varphi$ of §D.1.

D.3 Proof of the Gaussian Estimate

In this section we shall denote by $\| \ \|_2$ and $\langle \cdot, \cdot \rangle$ the corresponding scalar product in the Hilbert space $L^2(X; dx)$. We recall that $dx = d^r r\, dk$, and that $\check{d}x = \alpha^2\, dx$ is as in §D.2. The proof of (D.3) that we shall give below follows well-known lines (see Varopoulos et al., 1992, Chapter IX) and we shall be brief.

The first step is to fix φ, some real Lipschitz function on X that satisfies $|\nabla\varphi| = \sum |Y_j\varphi| \leqslant 1$ for the vector fields Y_j that give the operator D as in §D.2. We then have for any $\rho \in \mathbb{R}$,

$$-\langle e^{\rho\varphi}\alpha D\alpha^{-1}e^{-\rho\varphi}f, f\rangle = -\int Y_j\left(\alpha^{-1}e^{-\rho\varphi}f\right)Y_j\left(\alpha^{-1}e^{\rho\varphi}f\right)\check{d}x; \qquad f \in C_0^\infty(X), \tag{D.4}$$

since $dx = \alpha^{-2}\check{d}x$ and where we consider only real functions f. If we compute

$$e^{\mp\rho\varphi}Y_j\left(\alpha^{-1}e^{\pm\rho\varphi}f\right) = Y_j\left(\alpha^{-1}f\right) \pm \rho\,(Y_j\varphi)\,\alpha^{-1}f$$

and use the $|\nabla\varphi| \leqslant 1$ we conclude that the left-hand side of (D.4) is bounded from above by $\rho^2\|f\|_2^2$ and this implies that we can bound the L^2-operator norm by

$$\left\|e^{\rho\varphi}e^{-t\alpha D\alpha^{-1}}e^{-\rho\varphi}\right\|_{2\to 2} \leqslant e^{\rho^2 t}.$$

In terms of the kernel $p_t(x,y)$ this gives

$$\left|\iint p_t(x,y)e^{\rho(\varphi(x)-\varphi(y))}f(x)g(y)\,dx\,dy\right| \leqslant e^{\rho^2 t}\|f\|_2\,\|g\|_2; \qquad f, g \in C_0^\infty(X),\ \rho \in \mathbb{R}. \tag{D.5}$$

Now we fix $h(x) \in C_0^\infty(X)$ non-negative with compact support in, say, the unit ball of X. We shall then set $f = h_{r_1}$, $g = h_{r_2}$ for the translated functions $h_r(x) = h(rx)$ for the R-action on X and for which we have $\|h_r\|_p^p = m_R(r)^{-1}\|h\|_p^p$ for the modular function and any L^p-norm. With this choice of f, g and $t = 2$ the left-hand side of (D.5) can be bounded from below if $\rho > 0$ by

$$cp_1(r_1, r_2)e^{\rho(\varphi(r_1)-\varphi(r_2))-c\rho}\|h_{r_2}\|_1\,\|h_{r_1}\|_1$$

for some $c > 0$ because of the Harnack estimate. This combined with (D.5) and the above remark will finally give that there exists $C > 0$ that depends on h but is independent of ρ and φ such that

$$p_1(r_1,r_2) \leqslant Cm^{-1/2}(r_1)m^{-1/2}(r_2)\exp\left(2\rho^2 + \rho(\varphi(r_2) - \varphi(r_1)) + c\rho\right).$$

The 'trick' at this point is this. For $d(r_1,r_2)$ large we choose φ such that $\varphi(r_1) - \varphi(r_2) \sim d(r_1,r_2)$ and then $\rho = \varepsilon^{-1}d(r_1,r_2)$ for some small $\varepsilon > 0$. This will give the required estimate because, modulo the Gaussian factor $\exp(-cd^2(r_1,r_2))$, the cofactor $m_R(r_1)^{-1/2}m_R(r_2)^{-1/2}$ can be replaced by $m_R(r_1)^{-1}$ or $m_R(r_2)^{-1}$.

D.4 Applications to a Special Class of Operators

We shall go back to the situation that we examined in §5.11. We had there a projection $\pi\colon \widetilde{R} \to R$ between two connected Lie groups, that is, $R = \widetilde{R}/P$ where $P = \operatorname{Ker}\pi$. We were also given the usual identification of $G \simeq RK$ some other connected Lie group with the principal bundle $X = R \times K$. For the new principal bundle $\widetilde{X} = \widetilde{R} \times K$ we consider then the induced projection $\pi\colon \widetilde{X} \to X$.

The issue in §5.11 is the following. We start with a left-invariant Laplacian Δ on G and this is identified with an R-invariant operator $D = -\sum Y_j^2$ on X that satisfies the conditions of §D.2 with the measure $\alpha^2\, d^r r\, dk$ where here $\alpha = \chi$ and $\check{d}x = d^r g$ (with the above identification) as in (5.13). Now a new measure $\check{d}\tilde{x} = \tilde{\alpha}^2\, d^r \tilde{r}\, dk$, for $\tilde{x} = (\tilde{r},k) \in \widetilde{X}$, can be defined on $\widetilde{X}$ with $\tilde{\alpha}(\tilde{r}) = \alpha(\pi(r))$. And the issue is to define an operator $\widetilde{D}$ on $\widetilde{X}$ that satisfies all the conditions of §D.2 with the measure $\check{d}\tilde{x}$ and for which

$$\widetilde{D}(f \circ \pi) = (Df) \circ \pi; \quad f \in C_0^\infty(X). \tag{D.6}$$

This is what we call a lifting of the operator D to $\widetilde{X}$ in §5.6.4.

Here we shall give the construction but leave the easy verifications to the reader.

Each Y_j can be written (uniquely) as $Y_j = a_j(k)Z_j + A_j$ where Z_j is a left-invariant field in R, and A_j is some vector field on K, with $a_j \in C^\infty(K)$.

We can then pick up some (arbitrary) left-invariant field $\widetilde{Z}_j$ on $\widetilde{R}$ such that $d\pi(\widetilde{Z}_j) = Z_j$. We can also choose $\widetilde{U}_1,\ldots,\widetilde{U}_r$ left-invariant fields on $\widetilde{R}$ that are tangent to the subgroup $P = \ker\pi$ at the identity e and which are chosen to span at e the tangent space of P. Notice that P is not connected in general, but in defining these fields this makes no difference (we consider the component of the identity of P). We can then define

$$\widetilde{D} = -\sum \widetilde{Y}_j - \sum \widetilde{U}_j^2; \quad \widetilde{Y}_j = a_j(k)\widetilde{Z}_j + A_j(k),\ j = 1,2,\ldots.$$

This operator satisfies all the required conditions.

Exercise (The verification) The fact that $\widetilde{D}$ is subelliptic and that (D.6) holds follows from the subellipticity of D and the definition of the fields $\widetilde{Z}_j$ and $\widetilde{U}_j$. The fact that $\widetilde{Z}_j^* = -\widetilde{Z}_j$ (with respect to $\check{d}\tilde{x}$) follows from the same fact for the Z_j (with respect to $\check{d}x$). Finally, the fact that $\widetilde{U}_j^* = -\widetilde{U}_j$ follows from the definition. For these verifications we use the fact that $\tilde{\alpha}(gx) = \tilde{\alpha}(xg) = \tilde{\alpha}(x)$, $x \in \widetilde{R}$, $g \in P$ and that $d^r_{\widetilde{R}} = d^r_P \otimes d^r_{\widetilde{R}/P} = d^r_P \otimes d^r_R$ (see Bourbaki, 1963, Chapter 7: this notation has to be adapted when P is not connected).

A reformulation in terms of R-invariant operators All the operators $T = e^{-t\alpha D\alpha^{-1}}$ that we have considered are R-invariant operators on $X = R \times K$ with normal representation $T = L \otimes \{*\mu_{k_1,k_2}\}$ as in §4.4. Furthermore, $d\mu_{k_1,k_2}(r) = \varphi_{k_1,k_2}(r)\,dr$. We shall denote by $\phi(x_1,x_2)$ the kernel of T with respect to dx as in §5.3.3. Then if we test formula (5.7) on f with $f\,dx \sim \delta_e$ with $e = (e_R, k_2)$ for e_R, the neutral element of R, and $k_2 \in K$, we obtain $\phi(x_1, e) = \varphi_{k_1,k_2}(r_1)$ when $x_1 = (r_1, k_1) \in X$.

But the Gaussian estimate (D.3) on the kernel ϕ now gives

$$\phi(x_1, e) = O\left(\exp\left(-c|r_1|^2\right)\right)$$

for some $c > 0$. This shows that the measures μ_{k_1,k_2} satisfy uniformly the upper Gaussian estimate of §2.12.2 and that therefore T is a Gaussian operator, as needed in §5.10.

D.5 Questions Related to the Lower Gaussian Estimate

There exists what seems to be a discrepancy between the lower Gaussian estimate in the condition (5.60) and the definition of Gaussian operators in §5.10, where no mention of such an estimate is made. This discrepancy is only apparent because of the following two independent reasons.

The first is by a well-known elementary device that combines the Harnack estimate and the upper Gaussian; as a consequence the lower Gaussian comes for free! (see Varopoulos, 1990 or Varopoulos et al., 1992). We feel confident that the interested reader can fill in the details.

There is, however, another much better reason why we do not need to worry about this lower Gaussian. It is that for our applications it is simply not needed. Indeed, in the appendices of Chapter 5, where we give these applications, it is clear that we can in fact make do with far less!

Appendix E
Additional Results. Alternative Proofs and Prospects

As indicated in the title, we shall give in this appendix, with precise references but no proofs, additional information on $p_t(x,y)$, the kernel of the heat diffusion semigroup $e^{-t\Delta}$ on a connected Lie group G. The style of the presentation will be informal to say the least.

E.1 Small Time Estimates

Here we shall consider operators of the form

$$\Delta = -\sum_{j=1}^{n} Y_j^2 + Y_0$$

for left-invariant vector fields $Y_0, Y_1, \ldots$ such that the $Y_1, \ldots, Y_n$ already satisfy the Hörmander condition. We shall denote by $p_t(x,y)$ the kernel of $e^{-t\Delta}$ with respect to the left Haar measure dy on G, and by $d(x,y)$, for $x,y \in G$, the left-invariant distance induced by the fields $Y_1, \ldots, Y_n$ the usual way: that is, $|g| = d(e,g) = \inf(t_1 + t_2 + \cdots + t_p)$ when $g = \exp(t_1 Y_{i_1}) \exp(t_2 Y_{i_2}) \cdots \exp(t_p Y_{i_p})$ for some choice of $p \geqslant 0$ and of $i_1, \ldots$. The fundamental estimate then is this:

For all $0 < a < b$ *and* $\varepsilon > 0$ *there exists* $C > 0$ *such that*

$$C^{-1} \exp\left(-\frac{d^2(x,y)}{(4-\varepsilon)t}\right) \leqslant p_t(x,y) \leqslant C \exp\left(-\frac{d^2(x,y)}{(4+\varepsilon)t}\right); \quad x,y \in G,\ t \in [a,b]. \tag{E.1}$$

The upper estimate, especially in the case $Y_0 = 0$, is not too difficult to prove (see Varopoulos et al., 1992; Varopoulos, 1990, 1996b). Note, however, that an interesting and much more general and difficult result can be found in Hebisch (1992).

The lower estimate in (E.1) is considerably more difficult (see Varopoulos, 1990; Varopoulos, 1996b).

The sharpness of the Gaussian reflected by the $4 \pm \varepsilon$ of the exponent is something that was used crucially in one of the early proofs of the C-theorems of Chapter 2. Indeed, that sharpness of the exponent allows us to simplify considerably the use of the appendix in §2A (see the few lines that follow (2A.23)). This was done in Varopoulos (1994a).

Estimate (E.1) extends easily to operators on a principal bundle $X = R \times K$ where the fields $Y_0, \dots, Y_n$ now are assumed to be R-invariant. It is here that the need to incorporate a drift term in the operator Δ becomes clear. Indeed, even if we start from a driftless Laplacian Δ (i.e. with Y_0), then by taking the adjoint with respect to one of the measures $\alpha^2\,dx$ of §D.1, an additional drift field Y_0 automatically appears.

E.2 General Estimates for the Heat Diffusion Kernel

In §3.3 for amenable groups (see also §4.2) we have come across the local central limit theorem. A proof of this was given in Varopoulos (1999b). From another perspective, a number of 'off-diagonal' (i.e. with both t and x tending to ∞) can be found in Varopoulos (1996a). In that paper in particular we break the cofactor $m_R(y)^{-1}$ of (D.3) into interesting and relevant factors. (Both the above two papers are very difficult to read!)

It is plausible that if one works really hard one can put everything together and obtain optimal estimates of the form

$$\phi_t(x) \sim e^{-\lambda t} t^{-\nu} \exp\left(-\frac{|x|^2}{4t}\right) M(x),$$

for any NB-group, where λ is as usual the spectral gap and the factor $M(x)$ is what corresponds to $m_R(y)^{-1}$ in (D.3), or better still as in Varopoulos (1996a). The problem here is that the existing proofs, although they are basically just elaborations of the techniques that we have developed in this book, are extremely long and complicated to write down. One already gets a taste of this in the two above-mentioned papers that have appeared in print. As a consequence one has to overcome a strong psychological obstacle before embarking on this subject which, despite this handicap, is interesting and rewarding.

E.3 Bi-invariant Operators and Symmetric Spaces

As already seen on a number of occasions, the most natural example of the principal bundles of §4.2 arise in the context of a semisimple group S with a finite centre, which in the non-compact case, as explained in Appendix A, can be written $S = NAK$ and identified to $R \times K = X$ with $R = NA$. The corresponding *symmetric space* $\Sigma = S/K$ can then be identified with R and we can consider the R-invariant differential operators on X that correspond to left-invariant operators on G, which are K-bi-invariant (i.e. invariant by both left- and right-action by elements of K). These can be identified to G-invariant operators on the symmetric space Σ.

This gives a special class of operators that have been subject to deep and extensive study and on which a vast and comprehensive literature exists, for example Helgason (1984). Obviously, we do not intend to say anything in that direction and at any rate it would be very difficult to say anything new here.

In Varopoulos (1996b) we have, however, explained that in this special case the character χ and the eigenfunction φ_0, together with eigenvalue λ that played such a crucial role in §5.3, are old friends from the classical theory of spherical functions on symmetric spaces (see Helgason, 1984).

E.4 A Fundamentally Different Approach to the B–NB Classification

Unlike the previous sections of this appendix, here we shall describe something which bears explicit incidence in the presentation of this book.

We shall briefly describe the advantages and disadvantages of formulating the main B–NB classification in terms of the heat diffusion semigroup $e^{-t\Delta}$ rather than the convolution powers of measures that we actually used. In other words, §1.3.2 versus §1.3.1.

In this different approach we would have to start by constructing the heat diffusion semigroup, and also, and this is more difficult, the space of continuous paths $z(t) \in G$, with $t > 0$, of that diffusion. Even without too much knowledge of the subject, one knows, for example, that Bernoulli random walks and Brownian motion in $\mathbb{R}^d$ go together and that more often than not (and certainly for $d \geqslant 2$) it is easier to get exact exponents and exact formulas for Brownian motion.

The same thing happens here and this is why, for instance, the exact exponent ν of the local central limit theorem in (3.3) is easier to calculate for diffusion than for a random walk (Varopoulos, 1999b).

But even if we stick to things that we have developed in this book, like the gambler's ruin estimate in §5A, the corresponding proofs are definitely more satisfactory for continuous time diffusion. This is the point of view taken in Varopoulos (1996b) where a natural proof for these results was directly inspired by homogenisation theory (this is a subject that lies between PDEs and potential theory; see Jikov et al., 1991). This particular proof works only for elliptic Laplacians but still there is no doubt that this is the 'correct proof'.

So where are the disadvantages that made us adopt the different approach and work with random walks instead? The answer is simple.

We wanted to formulate the main B–NB classification in self-contained terms and not to have to construct the heat diffusion semigroup, let alone the corresponding path space, first. But even if we were prepared to accept this more sophisticated formulation, the actual potential-theoretic work in the appendices of Chapter 5 would perhaps have been simpler (we would have avoided then the use of error terms in the Taylor development of (5A.5) and (5A.8)) but the proofs would have been less accessible to readers and certainly far less self-contained.

PART II

GEOMETRIC THEORY

7

Geometric Theory. An Introduction

Overview of Chapter 7

This chapter serves several purposes.

A number of definitions will be given and notation will be fixed. These will be used throughout the geometric theory in the rest of the book.

Some of the basic theorems will be stated in a precise manner. The proofs will be given later.

Above all we shall describe some of the basic ideas in the theory. An effort has been made to make this chapter reader-friendly in order not to discourage those who are unfamiliar with geometric ideas. For the same reason, the differential forms in §7.6 will be treated here in an informal style and the more formal definitions will be deferred until Part III.

Guide for the reader To get a global picture of what this geometric theory of Part II is all about, the following route is recommended. First peruse the easy Chapter 7. Then go to the *overviews* of Chapters 8, 9 and 10 where we explain the general strategy of the proofs of the theorems of Chapter 7. Then the reader can decide which aspects of the proofs to explore first. It takes all those three chapters to give the proofs but they are compartmentalised and the components are independent.

Then comes something that is important for a global understanding of the theory: the reader should look at the *overview* of Chapter 11 as soon as possible and certainly before diving into the technicalities of Chapters 8–10. The reason is that in Chapter 11 we show that the special groups of Theorems 7.10 and 7.11 can be used as the building blocks that allow us to give the general B–NB classification of the geometric theory.

7.1 Basic Definitions and Notation

7.1.1 Manifolds

All the manifolds that we shall be considering will be C^∞ and connected, without further mention. Let $f : M \to M_1$ be some continuous mapping between two such manifolds, each assigned with a Riemannian distance. We say that f is Lipschitz, or more precisely $f \in \operatorname{Lip} R$, with $R > 0$, if for the corresponding distances we have

$$d_1 (f(x), f(y)) \leqslant R\, d(x,y); \quad x, y \in M. \tag{7.1}$$

This definition extends of course to general metric spaces, but in the case of Riemannian manifolds the differential

$$\mathrm{d}f : TM \longrightarrow TM_1 \tag{7.2}$$

exists almost everywhere and definition (7.1) is equivalent to the fact that

$$|\mathrm{d}f| \leqslant R \tag{7.3}$$

almost everywhere (see Federer, 1969, §3.1.9), where $|\mathrm{d}f|$ denotes the norm of $\mathrm{d}f\colon TM \to TM_1$ for the Riemannian norms $|\ |_m$, $|\ |_{f(m)}$ on T_mM, $T_{f(m)}M$, with $m \in M$, respectively.

Caution Some care is needed before we can actually assert that (7.1) and (7.3) are equivalent. The standard counterexample is the increasing non-zero function with zero derivative almost everywhere. To avoid this type of pathology, which is totally irrelevant to us here, we could start by assuming that f is locally Lipschitz, that is, for each $m \in M$ there exists $m \in \Omega \subset M$ some neighbourhood of m on which f satisfies (7.1) for some constant (depending on Ω). Then (7.1), (7.3) are equivalent, with the same $R > 0$.

Exercise Use elementary real analysis to prove the above statement. For example, since the problem is local, we can assume that $M = \mathbb{R}^d$ and that f is compactly supported and satisfies (7.3) a.e. For a fixed mollifier $\varphi \in C_0^\infty$ with $\int \varphi = 1$, it suffices to show that for the smooth function $f * \varphi$ we have $|\mathrm{d}(f * \varphi)| \leqslant R$. (This gives $f * \varphi \in \operatorname{Lip} R$ and we also have $f * \varphi \to f$ uniformly as $\varphi \to$ Dirac δ-mass at 0 and proves (7.1)). To see this we write the difference

$$\mathrm{D}_\varepsilon = f * \varphi(x+\varepsilon) - f * \varphi(x) = \int (f(x+\varepsilon-y) - f(x-y))\varphi(y)\, dy$$

and use Lebesgue dominated convergence to see that $\lim \left|\frac{1}{\varepsilon}\mathrm{D}_\varepsilon\right| \leqslant R$. Alternatively, we can argue as follows. The restriction of f on each smooth curve is absolutely continuous with a derivative with respect to arc length that is bounded by R and, since on that curve f is the integral of its derivative, we are done. This second proof is slightly incomplete (why is that?) and takes Fubini to make it proper.

The two manifolds M, M_1 will be called *quasi-isometric* if there exist

$$M \xrightarrow{f} M_1 \xrightarrow{f^{-1}} M, \tag{7.4}$$

inverses of each other, such that both are Lip$R$ for some fixed $R > 0$.

We shall consider pointed Riemannian manifolds (M, O), (M_1, O_1); that is, we shall fix points $O \in M$, $O_1 \in M_1$. We shall also write $|m| = d(O, m)$ and similarly $|m_1|_1 = d_1(O_1, m_1)$ for the distance from the origin.

Now let $f : M \to M_1$ be such a map that is smooth, or more generally locally Lipschitz. We shall say that f is a *polynomial map* if there exists $C > 0$ such that

$$|\mathrm{d}f| \leqslant C(1 + |m|)^C; \quad m \in M. \tag{7.5}$$

By $|\mathrm{d}f|$ here and in what follows we denote the operator norm of $\mathrm{d}f\colon T_mM \to T_{m_1}M_1$ for the Riemannian norms where $m_1 = f(m)$. An easy consequence of the definition of distances in a Riemannian manifold is that there exist constants such that

$$|f(m)|_1 \leqslant C(1 + |m|)^C. \tag{7.6}$$

Exercise 7.1 Prove this.

We say that (M, O) and (M_1, O_1) are *polynomially equivalent* if there exist locally Lipschitz mappings $f : M \to M_1$ and $f^{-1} : M_1 \to M$ that are polynomial mappings and inverses of each other.

It is finally evident that the composition of two Lipschitz maps is Lipschitz and the composition of two polynomial maps is polynomial. For the latter assertion we need to use (7.6).

Note finally that all the general Riemannian manifolds that we consider henceforth will tacitly be assumed to be complete (see Cheeger and Ebin, 1975; Helgason, 1978). But anyway, this completeness plays no role in our considerations.

Warning In many statements, for example (7.5), we omit the qualification ‘a.e.’ (i.e. ‘almost everywhere’) when it is either obvious or irrelevant.

7.2 Riemannian Structures on Lie Groups

7.2.1 Definitions

Let G be some connected Lie group and let us denote by

$$L_g : x \longrightarrow gx, \quad \mathrm{d}L_g : T_eG \longrightarrow T_gG; \quad x, g \in G \tag{7.7}$$

the left translation on G. We shall fix $|\cdot|^2 = \langle \cdot, \cdot \rangle$, a scalar product and a norm on T_eG, the tangent space at $e \in G$ the neutral element of G. We shall then use the left translation to identify T_eG with T_gG and thus obtain a Euclidean norm on T_gG for all $g \in G$. This determines a left-invariant Riemannian structure on G.

Two different norms, $|\cdot|_1$ and $|\cdot|_2$, on T_e determine two different quasi-isometric Riemannian structures on G. This means that the identity mapping is a quasi-isometry from one to the other. All the Lie groups that we shall be considering will be assigned with such a left-invariant Riemannian structure and the properties that we shall study will be invariant under quasi-isometric changes and therefore independent of the particular choice of the left-invariant structure chosen.

Recalling the notation of §1.1, we denote by $|g|_\Omega$ the distance of $g \in G$ from e associated to some compact neighbourhood of the identity $\Omega \subset G$. With the notation of the previous subsection we denote $|g| = d(g, e)$ for some left-invariant Riemannian structure. For any $P \subset G$ a compact neighbourhood of e then there exist constants C, C_1 such that

$$C_1|g| \leqslant |g|_P \leqslant C|g|; \quad g \in G, \ |g| > C. \tag{7.8}$$

Exercise 7.2 Prove this. Use the fact that for every compact set $K \subset G$ and P as above there exist constants such that (7.8) holds for $g \in K \backslash P$. Then iterate 'along a path'.

It is, furthermore, clear that any group homomorphism $f : G \to G_1$ is Lipschitz for the corresponding Riemannian structures. The more interesting examples of quasi-isometries between Lie groups from our point of view are not induced by group isomorphisms.

Example 7.3 The group of rigid motions on the complex plane is the semidirect product of the group of translations on $\mathbb{C} \approxeq \mathbb{R}^2$ and the group of rotations $\mathbb{T} = \mathbb{R} \bmod (2\pi)$; see §2.3.2. The action that defines the semidirect product is then given by the rotation $z \to e^{i\theta}z$, for $z \in \mathbb{C}$, $\theta \in \mathbb{R}$. With our previous notation this group is $\mathbb{R}^2 \leftthreetimes \mathbb{T}$ and the coordinates $(z, \theta (\bmod 2\pi))$ can be used to identify $\mathbb{R}^2 \leftthreetimes \mathbb{T}$ with the product group $\mathbb{R}^2 \times \mathbb{T}$ but this identification is not a group isomorphism. Later, in §8.3.5, we shall examine in detail general semidirect products and identifications like these. But as a useful exercise the reader may wish to prove directly here that the above identification is a quasi-isometry.

Example 7.4 We have already defined, in Chapter 1, the volume growth $\gamma(R) = \mathrm{Vol}\,[g \in G;\ |g| \leqslant R]$ where Vol in Chapter 1 was the left Haar measure of the group. Clearly the Riemannian volume of the left-invariant Riemannian structure on G is just left Haar measure (up to a multiplicative constant). We have already pointed out the fact (see §2.5.4) that $\gamma(R)$ grows polynomially if and only if G is an R-group (see Guivarc'h, 1973; Jenkins, 1973) and otherwise $\gamma(R)$ grows exponentially. From this and our definitions it follows in particular that if G_1 and G_2 are two groups that are polynomially equivalent as Riemannian manifolds then they both are R-groups or not R-groups simultaneously. To illustrate this issue further we may anticipate results that will be proved in Appendix F. Namely, a connected Lie group G is an R-group if and only if it is polynomially equivalent to a Lie group of the form $\mathbb{R}^d \times K$, where K is some compact group.

Example 7.5 The exponential mapping for a simply connected nilpotent group N $\exp\colon \mathfrak{n} \to N$ from the Lie algebra to the group has been used systematically in the analytic theory and we have already used the fact that it is a polynomial equivalence between the Euclidean structure on $\mathfrak{n}$ and the group Riemannian structure on N. It is (7.6) that was used, for example, in Chapter 3. That the more precise property (7.5) holds follows from the exact formula that gives the differential of the exponential mapping (see also Appendix F and Varadarajan, 1974, §2.14).

When $H \subset G$ is a closed subgroup of the Lie group the induced Riemannian structure on H from the Riemannian structure of G is of course the left-invariant Riemannian structure. As explained in detail in §2.14, however, one should not confuse the intrinsic distance in H and the distance induced by the ambient space G. This situation holds for general Riemannian manifolds and submanifolds.

7.3 Simply Connected Soluble Groups

Let Q be some simply connected soluble Lie group and let $\mathfrak{q}$ denote its Lie algebra. We have already pointed out on several occasions that Q is diffeomorphic with $\mathbb{R}^d$, that is, it is diffeomorphic with the Lie algebra $\mathfrak{q}$ assigned with its Euclidean structure. Unfortunately, the exponential mapping $\exp : \mathfrak{q} \to Q$, which can always be defined in some small neighbourhood of 0, cannot be used to define this diffeomorphism. This is the case when $\mathfrak{q}$ is nilpotent but not in general (see Dixmier, 1957). To obtain a convenient way to identify $\mathfrak{q}$ with Q one has to use the exponential coordinates of the second kind. Let us recall the definition briefly, but for more details the reader will have to refer to Varadarajan (1974, §3.18.11) or Chevalley (1955, §V.3.4).

7.3.1 Exponential coordinates of the second kind

We can find $\{e\} = \mathfrak{q}_0 \subset \mathfrak{q}_1 \subset \cdots \subset \mathfrak{q}_d = \mathfrak{q}$ a chain of subalgebras where d is the dimension of Q and where $\mathfrak{q}_j$ as a vector space is generated by $\mathfrak{q}_{j-1}$ and an additional vector $\xi_j \in \mathfrak{q}_j$, for $j = 1, \dots, d$, and where $(\operatorname{ad}\xi_j)\mathfrak{q}_{j-1} \subset q_{j-1}$, that is, $\mathfrak{q}_{j-1}$ is an ideal in $\mathfrak{q}_j$. That this is possible can be found in Varadarajan (1974, §3.7.5). We obtain thus an ordered basis of $\mathfrak{q}$, namely, $(\xi_1, \xi_2, \dots, \xi_d)$. For every $1 \leqslant j \leqslant d$ we shall use the abbreviation $e(t\xi_j) = \exp(t\xi_j) \in Q$, or even the shorter $e_j(t)$, with $t \in \mathbb{R}$, for the one-parameter subgroup generated by ξ_j. We obtain therefore a mapping

$$\mathbb{R}^d \ni t = (t_1, \dots, t_d) \longrightarrow e_1(t_1\xi_1) \cdots e(t_d\xi_d) = g \in Q. \tag{7.9}$$

The result that is proved in Varadarajan (1974, §3.18.11) is that (7.9) is a bijective diffeomorphism and t are the exponential coordinates of the second kind of Q.

The mapping (7.9) is a polynomial equivalence between $\mathbb{R}^d$ and Q when Q is nilpotent. To verify this one can use the Baker–Campbell–Hausdorff (BCH) formula (Varadarajan, 1974, §2.15) and its converse the Zassenhaus formula (Magnus et al., 1965; see also Chapter 6 for more details). One can also find an explicit proof of this in the literature (e.g. Varadarajan, 1974, §3.6.6, where it is seen that simply connected nilpotent groups are analytic subgroups of $\mathrm{GL}(V)$ for some vector space, i.e. that they are linear groups, and Chevalley, 1955, §V.3.4, where these facts are proved for such linear groups).

The above is in accordance with a general principle (a 'rule of thumb' is a more accurate description) that 'everything involving nilpotent groups is polynomial'.

Other examples where (7.9) is a polynomial equivalence can be given for R-groups. We shall prove this later but in fact we shall not make immediate use of this fact in what follows (see §8.4.1 and Appendix F).

Apart from these examples, the correspondence (7.9) is not polynomial in any sense whatsoever. To estimate the behaviour of this mapping in the general situation the exponential function e^t, rather than polynomials, has to be used in (7.5) and (7.6); see Appendix F.

Despite this, general simply connected soluble Lie groups can be classified by properties involving polynomials. To a large extent this is the subject matter of the geometric theory that we shall develop in this second half of the book. The key theorems will be presented in the next subsection but before that we shall end this section by giving a first illustration that explains why soluble simply connected Lie groups play such a central role in the theory. General connected Lie groups will not be examined until Chapter 11. Details will be given there. One of the facts that will emerge is the following:

Let G be some simply connected Lie group, for example, the universal cover of an arbitrary connected Lie group. Then there exists U some simply connected soluble Lie group and K some compact connected Lie group such that

$$G \simeq U \times K, \tag{7.10}$$

where $\simeq$ stands for a quasi-isometry between the corresponding Riemannian manifolds.

What this shows is that if we are only interested in 'large-distance geometry', that is, the behaviour far out at infinity for Lie groups, we can restrict our attention to the soluble simply connected groups that we described in this section. Assertion (7.10) applies only to simply connected groups, but in Chapter 11 we shall see that the same philosophy applies to all connected groups.

7.4 Polynomial Homotopy and the Geometric Theorems of Soluble Groups

7.4.1 Definitions

Let $f, f_1 : (M,O) \to (M_1,O_1)$ be two locally Lipschitz polynomial mappings between two pointed Riemannian manifolds as in §7.1. We then say that $f \sim f_1$ are *polynomially homotopically equivalent* if there exists F a polynomial homotopy between them. To wit,

there exists

$$F : M \times [0,1] \longrightarrow M_1, \tag{7.11}$$

a locally Lipschitz mapping for the natural Euclidean structure of $[0,1]$. (To be able to define dF almost everywhere without any difficulty we may as well assume that F extends to some locally Lipschitz mapping on $M \times [-\varepsilon, 1+\varepsilon]$ for some $\varepsilon > 0$. This clearly is always possible.) We shall assume that F is a homotopy between f and f_1 and satisfies:

$$\begin{aligned} &F(m,0) = f(m), \quad F(m,1) = f_1(m); \quad m \in M, \\ &|\mathrm{d}F(m,\lambda)| \leqslant C(1+|m|)^C; \quad m \in M,\ 0 \leqslant \lambda \leqslant 1, \end{aligned} \tag{7.12}$$

for some constants C where $|\ |$ denotes the norm of dF from $T_mM \times T_\lambda[0,1]$ to TM_1, for any $0 \leqslant \lambda \leqslant 1$, and where the product Riemannian structure is used.

The above obviously presupposes that f, f_1 are polynomial mappings in the sense of (7.5) and it is clear that $f \sim f_1$ is an equivalence relation between polynomial mappings.

We say that the two pointed manifolds (M,O), (M_1,O_1) are *polynomially homotopic* to each other if there exist polynomial mappings $M \xrightarrow{f} M_1 \xrightarrow{h} M$ such that $f \circ h$ and $h \circ f$ are polynomially homotopic to the identity mapping. We say that (M,O) is *polynomially retractable* if the identity mapping is polynomially homotopic to the constant mapping $k(m) = O$. To wit, in that case in (7.12),

$$F(m,1) = m, \quad F(m,0) = O; \quad m \in M \tag{7.13}$$

and $F(m,\lambda)$ shrinks the whole space M to O while respecting the polynomial property

$$|dF| \leqslant C_0(1+|m|)^C \tag{7.14}$$

for some constants $C, C_0 > 0$.

7.4.2 Lie groups

In what follows we shall be considering left-invariant Riemannian structures on connected Lie groups G. In such cases the base point will always be assumed to be the identity $e \in G$. But of course, since the Riemannian structures involved are homogeneous (i.e. invariant by left translation), any other point $g_0 \in G$ could have been taken as the base point and the mention of the base point in what follows can be omitted without creating any ambiguity.

Example 7.6 Every simply connected nilpotent Lie group is polynomially retractable. This is a consequence of Example 7.5. Indeed, it suffices to intertwine the radial retract $F(\xi,\lambda) = \lambda\xi \in \mathfrak{g}$, for $\xi \in \mathfrak{g}, 0 \leqslant \lambda \leqslant 1$, on the vector space that is the Lie algebra of G, with $\exp : \mathfrak{g} \to G$; then $\exp \circ F \circ \exp^{-1}$ gives the retract of G.

Example 7.7 More generally, when (7.9) is a polynomial equivalence, the soluble group Q in §7.3 is polynomially retractable. As already pointed out, this is the case for R-groups.

Example 7.8 The interest of the above definition is that it goes beyond R-groups for which (7.9) is polynomial. Already in Chapter 1 we indicated that the soluble group of affine motions on the real line $x \to ax + b$, for $a > 0$, $b \in \mathbb{R}$, is polynomially retractable and we used geodesics to obtain the retract.

This example falls under a more general scheme that is neither in the spirit nor the scope of the book but which should be described (or at least mentioned) because, together with the nilpotent groups of the previous example, it provides the basic model for the general definition. This is done in the next example (the reader unfamiliar with the terminology can ignore this).

Example 7.9 (Rank 1 symmetric spaces and Cartan–Hadamard manifolds) The group of affine motions that we considered in the previous example can be identified isometrically with two-dimensional hyperbolic space, that is, with the complex unit disc $D = [z \in \mathbb{C};\ |z| < 1]$ assigned with the Poincaré metric $|dz|^2/(1-|z|^2)^2$ (see Helgason, 1978, Exercise 1.G). In the introduction we used instead the upper-half complex plane obtained by conformal mapping. Other soluble groups of type NA of §3.2.3 with $\dim A = 1$ can be used as isometric models for the symmetric spaces of non-compact type and rank 1. This, in some sense, is what the Iwasawa decomposition (see Appendix A) is all about. See §8.2.1 for more on this.

Be that as it may, what counts is that these simply connected soluble groups with their group Riemannian structure are Cartan–Hadamard manifolds. These Riemannian manifolds, M, are by definition simply connected, complete and have strictly negative sectional curvature everywhere: $K(\cdot) \leqslant c_0 < 0$. These properties imply that M is diffeomorphic to $\mathbb{R}^d$ some Euclidean space. For a Cartan–Hadamard manifold M we can fix a base point $O \in M$ and for every $m \in M$ we shall consider the unique geodesic parametrised by length $\gamma_m(t) \in M$, with $0 \leqslant t \leqslant |m|$ and $\gamma(0) = O$, $\gamma(|m|) = m$. The fact that such a geodesic exists is a consequence of the completeness of M. The negative curvature guarantees that this geodesic is unique. The homotopy that retracts M to O as in (7.13), (7.14) can then be given: $F(m,\lambda) = \gamma_m(\lambda/|m|)$. The basic theory of Jacobi fields then implies the polynomial property (7.14) – in fact linear, because here C in (7.14) can be taken to be 1. For all this see Cheeger and Ebin (1975) or Helgason (1978).

Example 7.9 is important because it shows clearly that the property of being polynomially retractable for a manifold does not depend on the volume growth at infinity. The affine group $x \to ax + b$ is not unimodular and not an R-group and therefore is of exponential volume growth (see Guivarc'h, 1973; Jenkins, 1973). This last fact is of course very easy to verify directly.

If one is to describe the notion in informal terms, one should rather say that for a manifold to be polynomially retractable it must possess some smoothness (i.e. regularity) at infinity.

We now state the main geometric theorems for soluble, simply connected groups.

Theorem 7.10 (Geometric NC) *Let Q be some simply connected soluble Lie group that is assumed to be an* NC*-group in the sense of* §2.2*. Then Q is polynomially retractable.*

Theorem 7.11 (Geometric C) *Let Q be some simply connected soluble Lie group that is assumed to be a* C*-group. Then Q is not polynomially retractable.*

7.5 A Polynomial Filling Property

The proof of Theorem 7.10 will be done directly, essentially by constructing more or less explicitly the polynomial retract. For Theorem 7.11 however, the situation is more involved, even when we examine simple examples. It is here that homotopy and homology theory comes to our rescue. The reader does not need to know anything about these two subjects to follow this aspect of things. However, the spirit of these two subjects is very much present in both this and the next section, where we shall give the general description of the situation.

7.5.1 Notation

We shall denote throughout by $\square^n = [0,1]^n$ the unit cube of dimension n. We can embed $\square^n \subset \mathbb{R}^n$ in Euclidean space, and denote by $\partial \square^n$ its topological boundary. Both $\square^n$ and $\partial \square^n$ will be assigned with the induced Euclidean distance and this will allow us to define Lipschitz mappings into some Riemannian manifold.

In what follows we shall restrict our attention to manifolds $M = Q$ as in §7.3, that is, simply connected soluble Lie groups assigned with their left-invariant Riemannian structure. Let us fix $n = 1, 2, \ldots$ and let $R > 0$ (this is a 'free parameter'). We shall consider Lipschitz mappings

$$f\colon \partial \square^n \longrightarrow M; \quad f \in \operatorname{Lip} R. \tag{7.15}$$

For every such f it is possible, clearly, to find a Lipschitz extension (in several ways because locally M is some Euclidean space and then one can use the argument in the exercise below)

$$\hat{f} : \square^n \longrightarrow M, \quad \hat{f}\big|_{\partial \square^n} = f; \quad \hat{f} \text{ is Lipschitz.} \tag{7.16}$$

We shall denote by $\operatorname{Lip} \hat{f}$ the Lipschitz constant of $\hat{f}$, that is, the optimal A for which

$$d\left(\hat{f}(x), \hat{f}(y)\right) \leqslant A\, d(x,y); \quad x, y \in \square^n. \tag{7.17}$$

We shall then optimise over $\hat{f}$ and consider

$$A(f) = \inf\left[\operatorname{Lip}\hat{f};\ \hat{f} \text{ satisfies } (7.16)\right]. \tag{7.18}$$

This gives the most economical way of 'filling in' the boundary. A trivial example of this situation occurs when $n = 1$ and $\partial\,\square^n$ consists of just two points. Then the quantity in (7.18) is just the distance of these two points. In fact, in that case $A \sim R$ in (7.17) and for some functions $\hat{f}$ in (7.16), (7.17) we have $\operatorname{Lip}\hat{f} \sim R$. This of course is the point, and in the general situation for any $n \geqslant 1$ we set

$$\phi_n(R) = \sup\left[A(f);\ f \text{ satisfies } (7.15)\right]. \tag{7.19}$$

Like every 'min–max' definition it is tricky to write it out and to digest it but (7.19) simply gives a measure of how much we lose in the Lipschitz constant to 'fill the cube $\square^n$ in M'.

Definition 7.12 We say that M (which is a soluble simply connected group Q as in §7.3) has the *$\mathscr{F}_n$ property* (in other words the *polynomial filling property in dimension n*) if there exist constants C_0, C such that

$$\phi_n(R) \leqslant C_0 R^C; \quad R \geqslant 1. \tag{7.20}$$

The reason why we restricted attention to $R \geqslant 1$ is because we are only interested in the large-scale behaviour of M, that is, 'far out' at infinity.

It is evident of course that when M is polynomially retractable then M satisfies $\mathscr{F}_n$ for all $n = 1, 2, \ldots$ because we can use the retract of (7.13), (7.14) to fill in the cube. This is in direct analogy with the fact that for a retractable topological space, all the homotopy groups $\pi_n = 0$ for $n \geqslant 1$.

Exercise Verify this (see Bott and Tu, 1982, §13, 'The Extension Principle'). Assume, as we may, that the centre of $\square^n$ is the origin. Then we cut off a small ball B of size $\sim 10^{-10}$ centred at the origin. Then we use the radial mapping in $\mathbb{R}^n$ and the polynomial retract on M to extend the f of (7.15) to $\hat{f}\colon (\square^n \setminus B) \to M$ with ∂B going to 0. Then set $\hat{f}(B) = 0$.

The importance of the notion lies in the following result.

Theorem 7.13 (Theorem C) *Let us assume that Q is a simply connected soluble Lie group that is a* C*-group. Then there exists $n \geqslant 2$ such that property $\mathscr{F}_n$ fails.*

As we have already pointed out, $\mathscr{F}_1$ always holds, hence $n \geqslant 2$. To specify the index further we denote by $N \lhd Q$ the nilradical of Q and write $Q/N = A$ which is a Euclidean space. We may also denote by $\dim A = r(Q)$ the rank of Q, but this terminology is not standard. We then have the following theorem.

Theorem 7.14 (Refinement of Theorem C) *With the notation of Theorem* C *there exists* $2 \leqslant n \leqslant r(Q)+1$ *for which* $\mathscr{F}_n$ *fails*.

This refinement comes out automatically from the proof that will be given in Chapters 9 and 10. For sure, more effort is necessary to determine the indices $n \geqslant 2$ for which $\mathscr{F}_n$ holds or fails. This of course will depend on the geometry of the roots in the C-condition in §2.2. This problem has not been addressed and presents some independent interest and is no doubt difficult (cf. the epilogue at the end of the book). However, what is much more interesting, and more intriguing, is to relate the indices n for which $\mathscr{F}_n$ fails to the potential-theoretic properties of the group as in Part I.

As already pointed out in §1.4, the above notion is closely related with notions in Gromov (1991).

7.6 Differential Forms

7.6.1 Definitions and notation

In the same way that ideas from homotopy theory were used in the previous section to capture the geometric aspect of the C-condition, ideas from homology theory can also be used. This will be informally described in this section. More precise statements will be given in Part III of the book.

We may as well again restrict our attention to a manifold $M = Q$ which is a simply connected soluble group, as in §7.3, assigned with its left-invariant Riemannian structure. Our previous notation will be preserved. For every $m \in M$ a Euclidean norm can then be assigned on the exterior algebra $\bigwedge T_m^* M$ of the cotangent space (see Warner, 1971). This means that for local coordinates and $\mathrm{d}x_I = \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$, for some multi-index $I = (i_1, \ldots, i_k)$, we can define the Riemannian norm $|\mathrm{d}x_I|$. We could for instance fix, once and for all, the global coordinates $(t_1, \ldots, t_d)$ given by (7.9). The $|\mathrm{d}x_I|$ in general 'explodes' at infinity because the $\mathrm{d}x_1, \ldots, \mathrm{d}x_d$ are in general anything but orthonormal for the induced Riemannian structure. Using these local coordinates we can then define a differential form

$$\omega = \sum_I \omega_I(m)\mathrm{d}x_I; \quad m \in M, \tag{7.21}$$

written in local coordinates. In 'sophisticated language', ω is a section of $\bigwedge T^* M$.

The norm

$$|\omega(m)| = \sum |\omega_I(m)|\, |\mathrm{d}x_I| \tag{7.22}$$

can then be defined by taking the modulus $|\omega_I|$ of the coefficients $\omega_I \in \mathbb{R}$. We then say that ω is of polynomial growth if

$$|\omega(m)| \leqslant C_0 (1+|m|)^C ; \quad m \in M, \tag{7.23}$$

where C, C_0 are fixed positive constants and $|m|$ denotes the distance of m from the base point (i.e. $e \in Q$, the neutral element of the group).

We have said nothing about the nature of the coefficients $\omega_I(m)$. This was deliberate because here we are faced with a dilemma. There are two obvious choices.

7.6.2 The coefficients are smooth: $\omega_I \in C^\infty$

Then we are dealing with smooth forms. This is the more standard definition. Two problems arise however. Estimate (7.23) is then artificial. A more natural estimate would have been to give bounds for all derivatives $\partial/\partial x_i$, $\partial^2/\partial x_i \partial x_j, \dots$ of the coefficients.

The more serious problem is that in what follows we shall want to use $f\colon M \to M_1$, some polynomial map as in §7.1, and pull back $f^*\omega$, the differential form ω from M_1 to M. The smoothness is then lost because f is not a priori assumed smooth.

7.6.3 The alternative definition

The alternative definition is to say that $\omega_I \in L^\infty_{\text{loc}}$; that is, locally bounded (and perhaps demand in addition that in fact the coefficients ω_I are actually continuous). This makes estimate (7.23) natural and guarantees that we can pull back a differential form with L^∞_{loc} coefficients of polynomial growth by a polynomial map $f\colon M \to M_1$. The difficulty here consists in defining $\mathrm{d}\omega$, the differential of the form, and this is something that very much has to be done. This problem is dealt with by the use of the theory of currents (see de Rham, 1960) which in naive terms can be described as differential forms with coefficients that are distributions, for example L^∞_{loc} here.

We shall have to develop a bit of that in Chapter 10 and more extensively in Part III but for the time being note that if ω is a differential form as in (7.21) with coefficients that are L^∞_{loc} or even L^1_{loc} (or even measures) and if

$$\theta \in \bigwedge T^*M, \quad \theta = \sum \theta_J \mathrm{d}x_J \tag{7.24}$$

is some smooth compactly supported form, then the scalar product $\langle \omega, \theta \rangle$ can be defined unambiguously by the integral of

$$\omega \wedge \theta = \sum \pm \omega_I \theta_J \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_d. \tag{7.25}$$

In the summation we take J and I that are complementary, that is, they are disjoint and $I \cup J = [1, 2, \ldots, d]$. Here of course we use for $\pm$ the correct sign that gives

$$\pm \mathrm{d}x_I \wedge \mathrm{d}x_J = \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_d.$$

The scalar product

$$\langle \omega, \theta \rangle = \int_Q \omega \wedge \theta \tag{7.26}$$

can then be defined and $\mathrm{d}\omega$ can be defined weakly by Stokes' theorem (see de Rham, 1960) but even without going through that definition we can define and say the following:

The form ω in (7.21) with L^∞_{loc} coefficients is closed and $\mathrm{d}\omega = 0$ if and only if $\langle \omega, d\theta \rangle = 0$ for all smooth compactly supported $\theta \in \bigwedge T^*M$.

One can easily verify that the notion is stable by the pullback by a map $f\colon M \to M_1$. We shall return to all that in more detail in Part III. But we have explained enough here to be able to state the following result.

Theorem 7.15 (Theorem C for differential forms) *Let Q be a simply connected soluble* C*-group as in the geometric* C *theorem* (*Theorem 7.11*). *Then there exists $\omega \in \bigwedge T^*Q$ some differential form of polynomial growth*

$$\omega = \sum_{I \neq \emptyset} a_I \, \mathrm{d}x_I, \tag{7.27}$$

with continuous coefficients, that is closed (*i.e.* $\mathrm{d}\omega = 0$ *in the above sense*) *and is such that if θ is a differential form with L^∞_{loc} coefficients for which $\mathrm{d}\theta = \omega$ then θ is not of polynomial growth.*

We have not given the 'weak' definition of $\mathrm{d}\theta$ that is used here but this is of course given by $\langle \mathrm{d}\theta, \varphi \rangle = \pm \langle \theta, \mathrm{d}\varphi \rangle$ for all test forms φ (at least this is correct for homogeneous forms and the appropriately chosen $\pm$). All this will be developed in Part III.

The point of the above theorem and also of restricting the multi-indices in (7.27) to $I \neq \emptyset$ is that there exist plenty of solutions of $\mathrm{d}\theta = \omega$ but none of these solutions is a form θ that grows polynomially.

The immediate consequence of the above Theorem 7.15 is Theorem 7.11 which asserts that soluble simply connected C-groups are not retractable. We shall elaborate on this in Part III but what is involved is of course clear. If a group Q is polynomially retractable, one can use in the usual way (see de Rham, 1960; Dubrovin et al., 1990) the polynomial homotopy F of (7.12) and that retracts Q to e and solves the Poincaré equation $\mathrm{d}\theta = \omega$.

Example 7.16 (Non-retractable group) We consider the group $D = \mathbb{R}^2 \rtimes \mathbb{R}$ of §2.3.2 where the action of $\mathbb{R}$ on $\mathbb{R}^2$ is given by $\mathbb{R} \ni y \rightarrowtail (e^y x_1, e^{-y} x_2)$ for $(x_1, x_2) \in \mathbb{R}^2$, that is, a diagonal action and the two roots for the action of the Lie algebra are real with $L_1(y) = y$, $L_2(y) = -y$. We have therefore what is no doubt the simplest example of a C-group. It is interesting to observe that the coordinates (x_1, x_2, y) with $(x_1, x_2) \in \mathbb{R}^2$, $y \in \mathbb{R}$ are simply the exponential coordinates of the second kind defined in §7.3.1.

Using the above global coordinates, we shall consider on D the form $\omega = \mathrm{d}x_1 \wedge \mathrm{d}x_2$, which is clearly closed, $\mathrm{d}\omega = 0$ and ω grows polynomially; in fact $|\omega(g)| = c$, some constant, with $g \in D$.

This holds because the basis $\partial/\partial y$, $e^y(\partial/\partial x_1)$, $e^{-y}(\partial/\partial x_2)$ is an orthonormal basis of the tangent space of the group at the point (x_1, x_2, y). This fact is a consequence of the more general description of Riemannian structures on a semidirect product that we shall give in §8.3.5. But in this special case the reader could treat this as an exercise and prove it directly. The reader could also amuse themselves by trying to prove that no solution $\mathrm{d}\theta = \omega$ can be given with a polynomial form θ. In §9.2.3 we shall spell out the details of one possible way of seeing this.

We can consider a slightly more general example than the group D above, namely the group $D_{\alpha,-\beta} = \mathbb{R}^2 \rtimes \mathbb{R}$, where the only difference is that the action of $\mathbb{R}$ on $\mathbb{R}^2$ is given now by $y \rightarrowtail \left(e^{\alpha y} x_1, e^{-\beta y} x_2\right)$ for $\alpha, \beta > 0$ and the same notation. This is a C-group and we can show again that it is not (polynomially) retractable, by constructing as before some (smooth) differential form ω that is closed, $\mathrm{d}\omega = 0$, in the classical sense and which satisfies the conditions of the above Theorem 7.15. This construction will be treated in the following exercise.

Exercise 7.17 For the same reasons as before, $|\mathrm{d}x_1| = e^{\alpha y}$, $|\mathrm{d}x_2| = e^{-\beta y}$ at (x_1, x_2, y), and $|\mathrm{d}x_1 \wedge \mathrm{d}x_2| = e^{(\alpha-\beta)y}$ and we shall assume, as we may, by making the change of variables $y \to -y$ if necessary, that $\alpha > \beta$. The form

$$\begin{aligned} \omega = {} & f_1(x_1, x_2)\mathrm{d}x_1 \wedge \mathrm{d}F_1(y) + f_2(x_1, x_2)\mathrm{d}x_2 \wedge \mathrm{d}F_2(y) \\ & + \theta(x_1, x_2, y)\mathrm{d}x_1 \wedge \mathrm{d}x_2 \end{aligned} \tag{7.28}$$

is closed if $\partial\theta/\partial y = (\partial f_1/\partial x_2)F_1' - (\partial f_2/\partial x_1)F_2'$ and it can be made of polynomial growth if

$$\mathrm{d}F_2 = 0,\ \ y < C, \qquad F_1 = F_2 = 0,\ \ y > 0,$$

for some $C > 0$. The fact that we cannot solve polynomially $\mathrm{d}\theta = \omega$ is an easy extension of the argument in §9.2.3 (for an appropriate choice of f_i, F_i), but as an exercise the reader may again find amusement in trying to give their own reasons for that.

However, as we shall explain in §9.3.5, this easy approach will not be pursued because it does not lead to a general proof of Theorem 7.10.

Note These examples have existed in the literature for some time (see Gromov, 1991). We shall see more of them later in §9.2.

8
The Geometric NC-Theorem

Overview of Chapter 8

In §8.5 one finds the proof of the NC-theorem (Theorem 7.10). Before that, a fair amount of machinery will be developed. Section 8.1 consists mostly of notation but the substantial algebraic and geometric idea of 'replicas' is introduced in §8.3 and that of 'polynomial sections' in §8.4. These new ideas are interesting in their own right and permeate all the proofs of Part II. With these new tools, the exposition in Chapters 9 and 10 becomes much more readable than the original in Varopoulos (2000b).

Whether the same can be said about the proofs of the NC-theorem in this chapter is debatable. Taking into account the development of these new tools, the proof given here is longer than the original one in Varopoulos (2000b). It is, however, not substantially different though much cleaner than the original, which uses neither replicas nor polynomial sections. A brief outline of this original proof is given in §8.5.2 at the very end of the chapter. More significantly, the entire problem will be re-examined in Appendix F in the context of general NB-groups (whether soluble or not, whether simply connected or not).

Of the two theorems in §7.4, the NC-theorem (Theorem 7.10) is by far the easier to prove. In fact, its proof relies on just one standard idea from hyperbolic geometry (which is complex analysis in the unit disc if the dimension is 2). We explain this idea in some detail, albeit in an informal manner, in §8.2.

What we need to use is one of the central concepts in Riemannian geometry, namely that of geodesic flow when curvature is negative (Cheeger and Ebin, 1975; Helgason, 1978). In §8.2 we show how this can be adapted in our case where the curvature is not ('exactly') negative. The author debated whether to spend more time developing this purely Riemannian notion. But this would have taken us too far afield. The result was a ramble, perhaps unsatisfactory, through these important ideas that are presented mostly in the form of examples (see §8.2).

Unsatisfactory or not, the reader should spend time in §8.2 because once this notion of the generalised geometric flow is understood, then no one can have any doubt about how the proof of the NC-theorem should go. Furthermore, the theorem that we shall prove in Appendix F, which in some sense is the ultimate result in this direction, also hinges on the same notion of generalised geodesic flow.

8.1 Differentiation on Lie Groups

8.1.1 A description of the tangent space

Let G be some connected Lie group. Then $\mathfrak{g}$, the Lie algebra of G, is identified with $T_e(G)$, the tangent space at the identity, and one can use the exponential map $\exp : \mathfrak{g} \to G$ to identify $\mathfrak{g}$ with G near zero, and for every $\xi \in \mathfrak{g}$, if $\varphi(t) = e^{t\xi} = \exp(t\xi)$ is the one-parameter subgroup generated by ξ, then $\dot\varphi(0) \in T_e(G)$ is the vector corresponding to ξ. Here, for any mapping $\varphi : \mathbb{R} \to G$, we adopt the notation $\dot\varphi(t) = \mathrm{d}\varphi(\partial/\partial t)$ at t. The above statements in fact amount to a number of definitions for example $\mathfrak{g}$, exp, $e^{t\xi}$, which in some sense are 'circular'. It depends where we start so to speak, and then the other definitions follow. For all this we shall refer the reader to Varadarajan (1974, §§2.10–2.15) and to Helgason (1978, §II.5).

The left translation $L_g : x \to gx$ in G of §7.2 is then used, and $\xi \in \mathfrak{g}$ is identified with the vector

$$\mathrm{d}L_g\,\xi = \dot\rho(\xi, g) \in T_g(G); \quad g \in G, \tag{8.1}$$

where we use the notation $\dot\rho(\xi, g)$ for $\dot\rho(0, \xi, g)$ with $\rho(t, \xi, g) = ge^{t\xi}$. This is a convenient way of identifying $\mathfrak{g}$ with $T_g(G)$; it allows us to write down in the next two subsections a number of important formulas. (For a more formal approach to these formulas see Greub et al., 1973, §1.1.)

8.1.2 The inverse function

Let $J : G \to G$ be given by the involution $g \to g^{-1}$. Then

$$\mathrm{d}J\,\xi = -(\mathrm{Ad}\,g)\xi; \quad \xi \in \mathfrak{g}. \tag{8.2}$$

If $\xi \in \mathfrak{g}$ then it corresponds to $\dot\rho = \dot\rho(\xi, g) \in T_g(G)$. Therefore $dJ(\dot\rho) = \dot\theta(0)$ where

$$\theta(t) = e^{-t\xi} g^{-1} = g^{-1} g e^{-t\xi} g^{-1} = g^{-1} \exp\bigl(-t(\mathrm{Ad}\,g)\xi\bigr)$$

by Varadarajan (1974, formula (2.13.7)) (a formula we have already systematically used in the analytic theory; see for example (3.22). Hence our assertion.

8.1.3 The product

Let $P : G \times G \to G$ be the product mapping $\mathrm{P}(g_1, g_2) = g_1 g_2$. Then the Lie algebra of $G \times G$ is $\mathfrak{g} \times \mathfrak{g}$ and, for $\xi = (\xi_1, \xi_2) \in \mathfrak{g} \times \mathfrak{g}$ and $g = (g_1, g_2) \in G \times G$, with our previous notation we have

$$\left(g_1 e^{t\xi_1}, g_2 e^{t\xi_2}\right) = \rho\,(t, \xi, g). \tag{8.3}$$

It follows that $\mathrm{d}P(\dot{\rho}) = \dot{\theta}(0)$ with

$$\theta(t) = g_1 e^{t\xi_1} g_2 e^{t\xi_2} = g_1 g_2 e^{t(\mathrm{Ad}\, g_2^{-1})\xi_1} e^{t\xi_2} \tag{8.4}$$

for the same reasons as before. Here we shall use the BCH formula which gives (see Varadarajan, 1974, §2.15)

$$e^{t\zeta} e^{t\eta} = \exp\left(t(\zeta + \eta) + \tfrac{1}{2} t^2 [\zeta, \eta] + \cdots\right).$$

This, together with the fact that $\mathrm{d}\exp = \text{Identity}$ at $t = 0$ gives the formula

$$dP(\xi_1, \xi_2) = (\mathrm{Ad}\, g_2^{-1})\xi_1 + \xi_2 \tag{8.5}$$

for the differential at (g_1, g_1). This formula can be put in a more general form as follows.

Let M be some C^∞ manifold and let $\phi_i \colon M \to G$, with $i = 1, 2$, be two smooth mappings such that $\phi_i(m_0) = g_i$ for some point $m_0 \in M$. For the product mapping $F(m) = \phi_1(m)\phi_2(m)$, with $m \in M$, we then have

$$dF(m_0) = (\mathrm{Ad}\, g_2^{-1})\mathrm{d}\phi_1(m_0) + \mathrm{d}\phi_2(m_0), \tag{8.6}$$

where both sides of (8.6) are *identified with vectors in the Lie algebra* $\mathfrak{g}$ as in (8.1). To see this we could, for instance, compose the mapping $(\phi_1, \phi_2) \colon M \to G \times G$ with P.

One illustration of formula (8.6) that will be used in §8.4 is the following. Let $H \subset G$ be a closed subgroup such that $\mathrm{Ad}(H)$ is polynomial in the sense that the norms of the operators satisfy

$$|\mathrm{Ad}_{\mathfrak{g}}(h)| \leqslant C\,(|h|_H + 1)^C; \quad h \in H, \tag{8.7}$$

for some constants and where $\mathrm{Ad}_{\mathfrak{g}}$ indicates the action on $\mathfrak{g}$, the Lie algebra of G. Let us further assume that $\phi_1 : M \to G$ and $\phi_2 : M \to H$ are polynomial maps from the pointed manifold (M, O) to the corresponding groups. Then $F(m) = \phi_1(m)\phi_2(m)$ is a polynomial map $F \colon M \to G$. This incidentally, combined with

(8.9) below, is an easy way of seeing the ‘rule of thumb’ for nilpotent groups that we explained in §7.3.1.

8.1.4 Applications

All the notation of the previous subsections will be preserved. We shall illustrate the previous formulas with a number of applications. Many of them, especially those that are not directly relevant to the proof of Theorem 7.10, will be treated in the exercises.

The norm of the Ad **mapping** The mapping $\mathrm{Ad}(g) \in \mathrm{GL}(\mathfrak{g})$ is multiplicative in g and, if $|g|_G \leqslant n$, we can write $g = g_1 \cdots g_n$ as a product with $|g_i| \leqslant C$. This implies that there exist constants such that

$$|\mathrm{Ad}\, g| \leqslant C \exp(C|g|); \quad g \in G \tag{8.8}$$

for the Euclidean norm of the operator on $\mathfrak{g}$.

What is more subtle is that if $G = Q_R$ is a soluble simply connected R-group (see §7.3.1) then there exist constants such that

$$|\,\mathrm{Ad}\, g| \leqslant C(1+|g|)^C; \quad g \in G, \tag{8.9}$$

that is, a special case of (8.7). This is an immediate consequence of Lie’s theorem (see §§2.3.3, 3.9.1). This will not be used until much later, but for more details and an explicit proof of (8.9) see Appendix F.

Semidirect products Let $G = N \leftthreetimes K$ be the semidirect product of two connected Lie groups. Every $g \in G$ can be written uniquely $g = nk$, with $n \in N$, $k \in K$. We shall denote the corresponding projections by $\pi_N(g) = n$ and $\pi_K(g) = k$. The projection $\pi_K \colon G \to K$ is a group homomorphism; therefore $|d\pi_K| \leqslant C$. On the other hand, for $\pi_N \colon G \to N$, we can only assert that there are constants such that

$$|d\pi_N| \leqslant C \exp(C|g|); \quad g \in G, \tag{8.10}$$

where $|g|$ denotes as usual the distance in G from e. Inequality (8.10) is an immediate consequence of the fact that $n = gk^{-1}$ and of (8.2), (8.5), (8.8).

Consequences of (8.10) are the facts about distance distortion that we have already encountered in §2.14.2, namely, that for the distances in the groups G and N we have

$$|n|_N \leqslant C \exp\left(C|g|_G\right); \quad g \in G \tag{8.11}$$

for some constants $C > 0$.

Exercise Prove (8.11) and the results in §2.14.2 using the above ideas.

Exponential retract Let G be some connected real Lie group; we shall say that G admits an *exponential retract* (see §7.4) if there exists a locally Lipschitz map

$$F : G \times [0,1] \longrightarrow G, \tag{8.12}$$

for the product Riemannian structure such that for some constants $C > 0$ we have

$$|\mathrm{d}F(g,\lambda)| \leqslant C\exp(C|g|), \quad F(g,0) = e, \quad F(g,1) = g; \quad g \in G. \tag{8.13}$$

Example 8.1 Let $G = N \leftthreetimes K$ be the semidirect product of two simply connected nilpotent groups, and let F_N, F_K be polynomial retracts of N and K respectively. Such retracts exist (by Examples 7.6 and 7.7). We shall also demand that there exist constants such that $|F_K(g,t)| \leqslant C|g| + C$. (That such a retract can always be constructed will not be an issue here. This is, however, correct: see the end of Appendix F.) With the above notation we shall write

$$\begin{aligned}
&\widetilde{F}_N(g,\lambda) = F_N(\pi_N(g),\lambda),\\
&\widetilde{F}_K(g,\lambda) = F_K(\pi_K(g),\lambda); \quad g \in G,\ 0 \leqslant \lambda \leqslant 1.
\end{aligned}$$

Then let

$$F_G(g,\lambda) = \widetilde{F}_N(g,\lambda)\widetilde{F}_K(g,\lambda); \quad g \in G,\ 0 \leqslant \lambda \leqslant 1. \tag{8.14}$$

By formula (8.6) we have

$$\mathrm{d}F_G = (\mathrm{Ad}\,\widetilde{F}_K)^{-1}\mathrm{d}\widetilde{F}_N + \mathrm{d}\widetilde{F}_K. \tag{8.15}$$

By the conditions that we have imposed on F_N, F_K, if we combine (8.15), (8.10), (8.11) and (8.8) we deduce that F_G is an exponential retract of G as in (8.13).

This example will be developed further in §8.2.2 and especially in Appendix F where it will be considered in greater generality. Notice too that in this example we can also demand that $F_G(e,\lambda) = e$, $0 \leqslant \lambda \leqslant 1$, because this can be made to hold for F_N and F_K. In §12.2 on the other hand, we shall see that this additional condition actually comes for free.

In fact (see Appendix F), it can be shown that every soluble simply connected group admits such an exponential retract, but this will not be essential to us. What will be essential for the proof of the NC-theorem (Theorem 7.10) is to specialise the above group, $G = N \leftthreetimes K$, and assume that K is Abelian and that it acts on N by diagonal matrices. More precisely, we shall denote by $\mathfrak{k}$ the Lie algebra of K and its action on $\mathfrak{n}$, the Lie algebra of N, is given for an appropriate basis by

$$\operatorname{ad}\xi = \operatorname{diag}(L_1\xi, L_2\xi, \ldots, L_n\xi); \quad \xi \in \mathfrak{k}, \tag{8.16}$$

where $L_1, \ldots, L_n \in V = \mathfrak{k}^*$ are the roots, which are assumed to be real. The group G will be assumed to be an NC-group and we shall fix some $\zeta \in \mathfrak{k}$ such that

$$L_j\zeta > 1; \quad j = 1, \ldots, n, \tag{8.17}$$

as in §2.3.1. We then have the following result.

Proposition 8.2 *The above group $G = N \leftthreetimes K$ that satisfies* (8.16)*,* (8.17) *admits a polynomial retract.*

This proposition is one of the two components needed for the proof of the NC-theorem (Theorem 7.10). The other component is the fact that (essentially) every NC-group, as in this theorem, is polynomially homotopically equivalent to some special group as in Proposition 8.2. This will be in Proposition 8.3 below and with this we can complete the proof of the NC-theorem. The reason why we say 'essentially' is that condition (8.17) excludes the case that some of the L_i are zero, which can of course occur for NC-groups. We shall, however, see in §§8.4–8.5 below that this causes no serious problems.

R-groups Another non-trivial, but also inessential, application of the formulas of §§8.1.2–8.1.4 is the direct proof that simply connected soluble R-groups are polynomially retractable (as an exercise give the proof!). This is of course a special case of the NC-theorem 7.10. But it can also be used as an intermediate step in an alternative approach to that theorem. We shall examine these groups further in Appendix F and in Example 8.10 because much more is true here and, not surprisingly, an R-group as above is polynomially equivalent to Euclidean space see Example 7.7.

8.2 Strict Exponential Distortion and the Proof of Proposition 8.2

8.2.1 Hyperbolic geometry and 'heuristics'

We shall start with the group of affine motions $x \rightarrowtail ax + b, a = e^{\alpha} > 0, \alpha, b \in \mathbb{R}$. We have already pointed out, without a proof, in §1.4.2 the fact that this group can be identified isometrically for its group Riemannian structure with the upper-half complex plane $H = [z = b + ia \in \mathbb{C};\ a > 0]$ assigned with its Poincaré (hyperbolic) metric $a^{-2}|dz|^2$.

Exercise The reader is invited to use the general facts of §8.3.5 below to give a proof of this. Much more is given in Helgason (1978, Chapter 1, Further Results G) and also Helgason (1984, Introduction, §4).

For our considerations in this and the next chapter it is very important to recall a phenomenon on the distances in H that is basic in hyperbolic geometry. This is illustrated in Figure 8.1 drawn in H where the bottom line denotes the real axis.

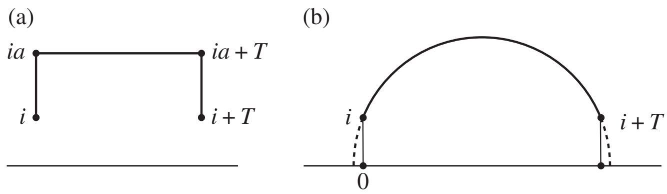


Figure 8.1 (a) A path of length $\approx \log T$, where $a = e^{\alpha}$; $\alpha = 10\log(T+10)$ and T is large; (b) the geodesic joining i to $i+T$, i.e. a circle orthogonal to the boundary.

The path in Figure 8.1(b) is the geodesic for the non-Euclidean (Lobachevski) geometry of H. In Figure 8.1(a), on the segment between ia and $ia+T$, if T is large the Poincaré metric is $\approx T^{-20}\times$ the Euclidean metric. Therefore this segment has bounded length (in fact it tends to 0 as $T \to \infty$) and this proves the assertion that the length of the path $\approx \log T$.

In terms of parametrisations, we could say that the path in Figure 8.1(a) is parametrised by $p(t)$, with $0 < t < 1$, where the segment from i to ia is covered when $t \in [0, 1/3]$. We need a speed $\dot{p}(t) \approx \log T$ to be able to do that because on that segment $p(t) = ie^{3\alpha t}$, and the non-Euclidean speed is measured by the derivative $\alpha'(t)$ (see (8.18) below for the notation). The situation is similar for $t \in [2/3, 1]$, which runs through the segment from $ia+T$ to $i+T$.

To be able to cover the segment from ia to $ia+T$ in Figure 8.1(a) we need a 'huge' speed $\approx T$, but fortunately we are talking here of the Euclidean speed $p'(t)$. The speed in the Poincaré metric stays bounded if T is large enough.

Let us reinterpret this phenomenon in terms of the group $ax+b$, which is the semidirect product

$$G = \mathbb{R} \leftthreetimes \mathbb{R}; \quad G \ni g = (b, e^{\alpha}),\ \alpha, b \in \mathbb{R}$$

(see §2.3.2 and also §8.3 below), concentrating on the heuristics and leaving the precise computations for later sections. The path $p(t)$ that we considered in H has then two components:

$$p(t) = (b(t), \alpha(t)) \in \mathbb{R} \leftthreetimes \mathbb{R}. \tag{8.18}$$

It is in fact the product $p(t) = b(t)\alpha(t)$ in the group G. It follows that if we wish to compute the speed (i.e. $\dot{p}(t) = dp(\partial/\partial t)$) for the Riemannian metric of the group G we can use the basic formula (8.6) to obtain

$$\dot{p}(t) = (\mathrm{Ad}\,\alpha(t))^{-1}\, b'(t) + \alpha'(t). \tag{8.19}$$

The reader should have in mind here that b, α lie in the two subgroups $\cong\mathbb{R}$ that give the semidirect product and therefore the ordinary derivatives b' and α' have been used. A much more precise description of Riemannian metrics in a semidirect product will in fact be given in §8.3.

For the time being, the heuristic interpretation of formula (8.19) that we have exploited here is as follows.

The path that joins the two points First we go vertically up in Figure 8.1(a); that is, we move with $\alpha(t)$ while keeping $b(t) = 0$. We do that until $\mathrm{Ad}\,\alpha(t)$ in (8.19) is so large that it will essentially 'kill' the speed coming from the necessary movement in the horizontal direction in Figure 8.1(a). This says that the factor Ad^{-1} compensates for $b'(t)$ in (8.19), until we reach $ia + T$ and then we reverse our first move in the vertical direction and reach $i + T$.

8.2.2 Strict exponential distortion

This ideas of §8.2.1 will be used in §8.2.3 in the proof of Proposition 8.2. But before that we shall further illustrate these ideas by constructing an example of a subgroup of strict exponential distortion (i.e. distance in the subgroup $\approx$ exp[distance in the group]; see Varopoulos, 2000a for more on that phenomenon).

We shall consider here a simply connected soluble Lie group of the form $G = N \leftthreetimes H$ where N, H are both nilpotent and we write $\mathfrak{g}$, $\mathfrak{n}$, $\mathfrak{h}$ for the corresponding Lie algebras. We shall denote (see §3.8.2) by $\mathfrak{n}\otimes\mathbb{C} = \mathfrak{n}_{\alpha_1} \oplus \mathfrak{n}_{\alpha_2} \oplus \cdots$ the root space decomposition under the action $\mathrm{ad}\,\mathfrak{h}$ and block together $\mathfrak{n}_{\alpha_1} \oplus \cdots = \mathfrak{n}_L \otimes \mathbb{C}$ all the root spaces with roots that satisfy $\mathrm{Re}\,\alpha_1 = \mathrm{Re}\,\alpha_2 = \cdots = L$ for some $0 \neq L \in \mathfrak{h}^*$. We assume of course that such an L and $\alpha_1, \ldots$ exist.

We shall fix $0 \neq \xi \in \mathfrak{n}_L$ and consider $e^{t\xi} \in N$, the one-parameter subgroup generated by ξ. In the next few lines we shall show that there exists then a constant C such that

$$\left|e^{T\xi}\right|_G \leqslant C\log(|T| + 10); \quad T \in \mathbb{R}. \tag{8.20}$$

This was exactly the situation earlier for the one-parameter subgroup $x \mapsto x + b$ of translations in the group of affine motions. Observe that for the proof of (8.20) we may as well assume that $T \geqslant 100$.

To prove (8.20) let us fix $\zeta \in \mathfrak{h}$ such that $L\zeta = 1$. Then a direct application of §3.9.1 implies that the norm of $\mathrm{Ad}(e^{t\zeta})$ as an operator in $\mathrm{GL}(\mathfrak{n}_L)$ satisfies

$$\left|\mathrm{Ad}\left(e^{-t\zeta}\right)\right| \leqslant e^{-ct}; \quad t > C \tag{8.21}$$

for appropriate constants.

The path that we shall use to join the identity to $e^{T\xi}$ in G will then be given by the product of two time-changed one-parameter subgroups

$$p(s) = \phi_1(s)\,\phi_2(s) = \exp\left(\alpha(s)T\xi\right)\,\exp\left(A\beta(s)\log T\zeta\right); \quad 0 < s < 1, \tag{8.22}$$

where $A > 0$, $\alpha(\cdot)$, $\beta(\cdot) \in C^\infty$ with $\alpha(0) = \beta(0) = \beta(1) = 0$, $\alpha(1) = 1$. We shall show that these may be chosen so that the speed of that path $dp(\partial/\partial s) = \dot{p} \in TG$ satisfies

$$|\dot{p}(s)| \leqslant C\log T; \quad 0 < s < 1 \tag{8.23}$$

for some appropriate $C > 0$. This will clearly imply (8.20).

To see this we shall choose the α increasing from 0 to 1, $\beta \geqslant 0$ and also

$$\begin{aligned} \alpha(s) &= 0; \quad s \in \left[0, \tfrac{1}{3}\right], \\ \alpha(s) &= 1; \quad s \in \left[\tfrac{2}{3}, 1\right], \\ \beta(s) &= 1; \quad s \in \left[\tfrac{1}{10}, \tfrac{9}{10}\right], \end{aligned} \tag{8.24}$$

that is, β takes off fast and stays equal to 1 before it sinks again. As for α, it takes all its variation from 0 to 1 in the range where $\beta = 1$. As for $A > 0$, it will be fixed at the end and it will be large enough. We then have

$$\begin{gathered} \left|\dot{\phi}_1(s)\right| \leqslant CT, \quad \left|\dot{\phi}_2(s)\right| \leqslant A\log T; \quad 0 < s < 1, \\ |\dot{p}(s)| \leqslant A\log T; \quad s < \tfrac{1}{3} \text{ or } s > \tfrac{2}{3}. \end{gathered} \tag{8.25}$$

But from (8.21) we have, as well,

$$\left|\mathrm{Ad}\,\phi_2^{-1}\right| \leqslant CT^{-cA}; \quad \tfrac{1}{3} < s < \tfrac{2}{3}, \tag{8.26}$$

where the constants c and C in (8.25) and (8.26) are independent of A. If we use (8.24), (8.25), (8.26) with A large enough, and formula (8.6), we finally see that (8.23) holds and this completes the proof of our assertion.

8.2.3 Proof of Proposition 8.2

In this subsection, $G = N \leftthreetimes K$ will be simply connected, N will be nilpotent, $K \cong \mathbb{R}^a$ Abelian and all the notation for the roots of the action of $\mathfrak{k}$ on $\mathfrak{n}$ and $\zeta \in \mathfrak{k}$ will be as in (8.16), (8.17) in Proposition 8.2. The notation of the homotopies F_N and F_K will be as in Example 8.1 but of course, here, for F_K we may simply take $F_K(x, \lambda) = \lambda x$, with $x \in K \cong \mathbb{R}^a$.

For $g = nk$, with $n \in N, k \in K$, we define $F_G(g,\lambda) = \widetilde{F}_N(n,\lambda)\widetilde{F}_K(k,\lambda)$ as in (8.14). The issue is to prove that in our special group we can modify F_G and construct a new homotopy retract F as in (8.12) that improves the exponential estimate (8.13) and gives instead

$$|\mathrm{d}F| \leqslant C(1+|g|)^C; \quad g \in G \tag{8.27}$$

for appropriate constants. The proof closely follows the ideas of §8.2.2.

With the above notation we shall write $\sigma(t) = \exp(t\zeta) \in G$ which is but a straight line in $K = \mathbb{R}^a$ and set

$$F(g,\lambda) = F_G(g,\alpha(\lambda))\,\sigma(A\beta(\lambda)|g|); \quad g \in G,\ 0 \leqslant \lambda \leqslant 1. \tag{8.28}$$

We shall choose large enough $A > 0$ at the end, and the functions $0 \leqslant \alpha, \beta \in C^\infty$ that perform the time change will be exactly as in §8.2.2, with $\alpha = 0$ for $0 \leqslant \lambda \leqslant 1/3$, $\alpha = 1$ for $2/3 < \lambda \leqslant 1$ and $\beta(0) = \beta(1) = 0$, $\beta = 1$ for $1/10 \leqslant \lambda \leqslant 9/10$.

With this definition, with constants C that are independent of A, we clearly have

$$|\mathrm{d}F| \leqslant CA(|g|+C); \quad \lambda < \tfrac{1}{3} \text{ or } \lambda > \tfrac{2}{3}; \tag{8.29}$$

note that the first factor in (8.28) is $F_G(g) = g$ or $F_G(g) = e$ in that range. To see this we apply formula (8.6) with $\phi_1 = F_G$ and $\phi_2 = \sigma$ and in our range and by the choice of ζ in (8.17) we can assert that $|\mathrm{d}\phi_1|,\ |\mathrm{d}\phi_2| \leqslant CA(|g|+C)$ and $|\operatorname{Ad}\phi_2^{-1}| \leqslant C$ (see §3.9.1). For the range $1/3 < \lambda < 2/3$ formula (8.6) gives

$$\begin{aligned} \mathrm{d}F &= \operatorname{Ad}\sigma^{-1}\mathrm{d}F_G + \mathrm{d}\sigma \\ &= \operatorname{Ad}\sigma^{-1}\operatorname{Ad}\tilde{F}_K^{-1}\tilde{F}_N + \mathrm{d}\tilde{F}_K + \mathrm{d}\sigma; \quad \tfrac{1}{3} < \lambda < \tfrac{2}{3}, \end{aligned} \tag{8.30}$$

because K is commutative and the $\operatorname{Ad}\sigma$-action on $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{k}$, the Lie algebra of G, actually reduces to the action on $\mathfrak{n}$. Now, on the right-hand side of (8.30), the second and third terms are bounded by $C|g|+C$ and in the first term the factor $\operatorname{Ad}\tilde{F}_K^{-1}\mathrm{d}\tilde{F}_N$ can be estimated as in Example 8.1 by $\exp(C|g|)$ with C independent of A. If, however, we use (8.16), (8.17), it follows that the norm of $\operatorname{Ad}\sigma^{-1}$ on $\mathfrak{n}$ can be estimated by $\exp(-cA|g|)$ with c independent of A. It follows that if A is chosen large enough, we have $|\mathrm{d}F| \leqslant C|g|+C$ in (8.30). This completes the proof of the proposition.

8.3 Semidirect Products

8.3.1 The definition of the semidirect product

Let N, H be two connected Lie groups. Then the usual definition of their semidirect product is given by the formula

$$(n_1,h_1)(n_2,h_2) = \left(n_1 n_2^{h_1}, h_1 h_2\right); \quad n_1,n_2 \in N,\ h_1,h_2 \in H \tag{8.31}$$

and by a homomorphism $s: H \to \mathrm{Aut}(N)$, the group of automorphisms of N, and where we write $n^h = s(h)n$, with $n \in N$, in (8.31). The notation is $N \leftthreetimes H$, though $N \leftthreetimes_s H$ would have been more accurate.

We shall assume in what follows that the groups N and H are simply connected. In that case, s is determined by a homomorphism $\hat{s}: H \to \mathrm{GL}(\mathfrak{n})$ where $\mathfrak{n}$ is the Lie algebra of N provided that $\hat{s}(H) \subset \mathrm{Aut}(\mathfrak{n})$, that is, that $\hat{s}(h)$ is an automorphism of the Lie algebra $\mathfrak{n}$. Taking the differential $\sigma = d\hat{s}$ we then have $\sigma: \mathfrak{h} \to \mathrm{Der}(\mathfrak{n})$ a Lie algebra homomorphism where $\mathfrak{h}$ is the Lie algebra of H and $\sigma(\xi) \in \mathrm{Der}(\mathfrak{n}) = \mathrm{D}(\mathfrak{n})$, the algebra of derivations of $\mathfrak{n}$ (we recall that in any algebra A, the map $\mathrm{D}\colon A \to A$ is a derivation if $\mathrm{D}(x \cdot y) = (\mathrm{D}x) \cdot y + x \cdot (\mathrm{D}y)$).

For any $\mathrm{D} \in \mathrm{D}(\mathfrak{n})$ we have $e^{\mathrm{D}} \in \mathrm{Aut}(\mathfrak{n})$ and therefore we can reverse the above process. So from a Lie algebra homomorphism $\sigma: \mathfrak{h} \to \mathrm{D}(\mathfrak{n})$ we can reconstruct $\hat{s}$ and s. For all this see Hochschild (1965, Chapter IX); Varadarajan (1974, §§2.13, 2.14); Helgason (1978, §II.5).

For the Lie group structure of $G = N \leftthreetimes H$ the Lie algebra is $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{h}$ and then with the above notation we can identify

$$\mathrm{ad}\,\xi = \sigma(\xi), \quad \hat{s}(\exp\xi) = e^{\mathrm{ad}\,\xi} = \mathrm{Ad}\exp\xi \in \mathrm{Aut}(\mathfrak{n}); \quad \xi \in \mathfrak{h}.$$

We regard all these as operators acting on $\mathfrak{n}$. For this reason in the above definition, in what follows we shall abuse notation and write $\sigma = \mathrm{ad}$, $e^{\mathrm{ad}} = \hat{s} \circ \exp = \mathrm{Ad} \circ \exp$. The reader may wish to go back to the definitions in the above references to clarify the above formulas.

8.3.2 Definition of replicas

The notation $G = N \leftthreetimes H$, $\mathfrak{n}$, $\mathfrak{h}$, $\sigma = \mathrm{ad}$, $e^{\mathrm{ad}} = \mathrm{Ad} \circ \exp$ are as in the previous subsection. The Lie algebra homomorphism $\mathrm{ad}: \mathfrak{h} \to \mathrm{D}(\mathfrak{n})$ will then be 'modified' to another similar homomorphism $\mathrm{ad}_0 = \sigma_0 : \mathfrak{h} \to \mathrm{D}(\mathfrak{n})$ which will be called a *replica* of σ and from this we can get $\hat{s}_0 \circ \exp = \mathrm{Ad}_0 \circ \exp = e^{\mathrm{ad}_0}$ and a new semidirect product $G_0 = N \leftthreetimes_0 H$ will be defined. This semidirect product will be called a replica of the original product G. The above idea fits with the more general set-up of replicas that can be found in Varadarajan (1974, §3.1).

This general notion of replicas is due to C. Chevalley and, like everything else that he did, is ‘worth the detour’. The reader is urged to check this out.

For us here we shall consider only concrete explicit examples and we shall assume throughout that $\mathfrak{n}$ and $\mathfrak{h}$ are nilpotent Lie algebras and that

$$\mathfrak{n}\otimes\mathbb{C}=\mathfrak{n}_{\alpha_1}\oplus\mathfrak{n}_{\alpha_2}\oplus\cdots \tag{8.32}$$

is the root space decomposition of the action of $\sigma(\mathfrak{h})=\operatorname{ad}\mathfrak{h}$ on the complexification of $\mathfrak{n}$ (see Jacobson, 1962, §II.4; see also §3.8).

8.3.3 The semisimple replica

To give the first significant example of a replica, for every $\xi\in\mathfrak{h}$ we shall decompose $\sigma(\xi)=\sigma_S(\xi)+\sigma_N(\xi)$ into its semisimple and nilpotent parts. Here $\sigma_S(\xi)$ is a replica of $\sigma(\xi)$ in the sense of Chevalley, and the fact that $\sigma_S(\xi)\in \mathrm{D}(\mathfrak{n})$ follows from the general theory (see Varadarajan, 1974, §3.1, especially §3.1.14). We can verify this directly by using the root space decomposition (8.32) as follows. The complexification of the mapping $\sigma(\xi)$ reduces on each $\mathfrak{n}_{\alpha_j}$ to the mapping

$$\alpha_j(\xi)\mathrm{I}+\mathrm{T}_j, \tag{8.33}$$

where I is the identity, T_j is nilpotent and the complexification of $\sigma_S(\xi)$ is defined by $\alpha_j(\xi)\mathrm{I}$ on each $\mathfrak{n}_{\alpha_j}$. On the other hand, $[\mathfrak{n}_{\alpha_j},\mathfrak{n}_{\alpha_k}]\subset\mathfrak{n}_{\alpha_j+\alpha_k}$ if $\alpha_j+\alpha_k$ is a root, and that bracket is zero otherwise. From this we see that when $\sigma(\xi)$ is a derivation of $\mathfrak{n}$, the complexification of $\sigma_S(\xi)$, and therefore $\sigma_S(\xi)$ itself, is a derivation. The fact that $\xi\to\sigma_S(\xi)$ is a Lie algebra homomorphism ($\mathfrak{h}\to\mathrm{D}(\mathfrak{n})$) is also clear because we have in general $\alpha_j([\xi_1,\xi_2])=0$ for any root. Therefore we have

$$\sigma([\xi_1,\xi_2])_S=[\sigma_S(\xi_1),\sigma_S(\xi_2)]=0;\quad \xi_1,\xi_2\in\mathfrak{h}. \tag{8.34}$$

We shall write $\operatorname{ad}_S=\sigma_S$ and $G_S=N\leftthreetimes_S H$ for the corresponding replica.

The C–NC classification It is clear from the above definition that $N\leftthreetimes_S H$ is a C- or NC-group if $N\leftthreetimes H$. This of course holds because the corresponding algebras have exactly the same roots defined by Lie’s theorem in §2.3.3.

8.3.4 A class of special soluble groups

We shall say that $\mathfrak{g}$ is given by a *real semisimple action* if $\mathfrak{g}=\mathfrak{n}\leftthreetimes\mathfrak{h}$, where $\mathfrak{n}$ and $\mathfrak{h}$ are nilpotent, and where the action of $\mathfrak{h}$ on $\mathfrak{n}$ is semisimple with real roots. This means, for an appropriate basis of $\mathfrak{n}$, that $\operatorname{ad}\xi$, with $\xi \in \mathfrak{h}$, takes the diagonal form

$$\operatorname{ad}\xi = \operatorname{Diag}(L_1\xi, \dots, L_n\xi), \tag{8.35}$$

where the roots L_j are real, that is, they are $L_1, \dots, L_n \in \mathfrak{h}^*$, linear functionals on $\mathfrak{h}$. No complexification is involved here.

Let us go back to the replica $N \leftthreetimes_S H$ that we constructed in the previous subsection and the complexification of $\sigma_S(\xi)$ that acts on the root spaces $\mathfrak{n}_{\alpha_j}$ of (8.32) by $\sigma_S(\xi) = \alpha_j(\xi)\mathrm{I}$, with $\xi \in \mathfrak{h}$. We shall then define on each $\mathfrak{n}_{\alpha_j}$ the action

$$\sigma_0(\xi) = \operatorname{Re}\alpha_j(\xi)\mathrm{I}; \quad \xi \in \mathfrak{h}. \tag{8.36}$$

In other words, and loosely speaking, we replace each root by its real part. The mapping $\sigma_0(\xi)$ maps $\mathfrak{n}$ into itself (i.e. is a real mapping), and by the same argument as before, though even easier, we see that it is a replica of σ (see the exercise below). We shall write

$$\operatorname{ad}_0 = \sigma_0, \qquad \operatorname{Ad}_0 \circ \exp = e^{\operatorname{ad}_0}, \qquad \mathfrak{n} \leftthreetimes_0 \mathfrak{h} = \mathfrak{g}_0. \tag{8.37}$$

Exercise The sum of all the $\mathfrak{n}_{\alpha_j}$ for the roots α_j with the same real part L, is of the form $\mathfrak{n}_L \otimes \mathbb{C}$ for some $\mathfrak{n}_L \subset \mathfrak{n}$ (because it is a real space, i.e. stable under conjugation). We have $\mathfrak{n} = \mathfrak{n}_{L_1} \oplus \mathfrak{n}_{L_2} \oplus \cdots$ and $[\mathfrak{n}_{L_j}, \mathfrak{n}_{L_k}] \subset \mathfrak{n}_{L_j+L_k}$. Our assertion follows because $\sigma_0(\xi)$ reduces to a real scalar transformation on each $\mathfrak{n}_L$. This is also consistent with the general set-up of replicas that we mentioned, and in Varadarajan (1974, §3.1.15), one can find a complete description of the replicas of diagonal transformations.

This replica $\mathfrak{g}_0$ is then given by a real semisimple action. We shall denote by $G_0 = N \leftthreetimes_0 H$ the corresponding simply connected group that is a replica of G. As before it is of course clear that G_0 is a C- or NC-group if G is.

We shall now state one of the main results of the chapter and also the raison d'étre of the replica ad_0.

Proposition 8.3 *Let Q be some simply connected soluble Lie group. Then there exists $G = N \leftthreetimes A$, some simply connected group, whose algebra $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ is given by real semisimple action, where $\mathfrak{n}$ is nilpotent and $\mathfrak{a}$ is Abelian and such that Q is polynomially homotopically equivalent to G and furthermore, G is* C- *(resp.* NC*) if Q is.*

The proof of Proposition 8.3 will be completed in §8.5.1 below. We pointed out previously, however, in Proposition 8.2, that this is the missing link needed to complete the proof of the NC-theorem (Theorem 7.10). Replicas will be used in the proof but before then some preparation is needed.

8.3.5 Riemannian structures on semidirect products

We shall consider again $G = N \leftthreetimes H$ the semidirect product of two connected Lie groups and let $\mathfrak{n}$, $\mathfrak{h}$, $\mathfrak{g}$ denote the Lie algebras of N, H and G respectively. We shall fix left-invariant Riemannian structures on N and on H by specifying $|\ |_{\mathfrak{n}}$ and $|\ |_{\mathfrak{h}}$ some Euclidean norms. These will be kept fixed throughout. We define, then, the corresponding Euclidean norm $|\ |_{\mathfrak{g}}^2 = |\ |_{\mathfrak{n}}^2 + |\ |_{\mathfrak{h}}^2$ on $\mathfrak{g}$.

Now G as a differential manifold can be identified with the product $N \times H$ and therefore at every $g = nh \in G$ ($n \in N, h \in H$) the tangent space $\mathrm{T}_g(G)$ can be identified with $\mathrm{T}_n(N) \times \mathrm{T}_h(H)$. To push this identification a step further, we write

$$\mathrm{T}_g(G) \ni X = Y + Z; \quad Y \in \mathrm{T}_n(N),\ Z \in \mathrm{T}_h(H), \tag{8.38}$$

and then, with the notation of §8.1.1,

$$Y = \dot{\rho}(\nu, n), \quad Z = \dot{\rho}(\mu, h); \quad \nu \in \mathfrak{n},\ \mu \in \mathfrak{h}, \tag{8.39}$$

which identifies $X = Y + Z$ with $\nu + \mu$.

We shall now give the explicit expression of the Riemannian norm on $\mathrm{T}_g(G)$ for the left-invariant structure induced by $|\ |_{\mathfrak{g}}$. To do this we consider first $\dot{\rho}(\mu, g)$ which is the L_g translate of $\mu \in \mathfrak{h} \subset \mathfrak{g}$ and in the identification $\mathrm{T}(G) = \mathrm{T}(N) \times \mathrm{T}(H)$ this is Z as in (8.39).

Now let $\nu_1 \in \mathfrak{n}$ and consider $\rho(t) = \rho(t, \nu_1, g) \in G$, with $t \in \mathbb{R}$ where ν_1 is considered as an element of $\mathfrak{g}$. Then (see §8.1)

$$\rho(t) = n \exp\big(t(\mathrm{Ad}\,h)\nu_1\big)h = \rho\big(t, (\mathrm{Ad}\,h)\nu_1, n\big)h; \quad g = nh \in G. \tag{8.40}$$

This means that if $\nu_1 = (\mathrm{Ad}\,h^{-1})\nu$ in (8.40) we have $\dot{\rho} = Y$.

The above can be reformulated by saying that when we identify $X \in \mathrm{T}_g(G)$ in (8.38, 8.39) with the elements of $\mathfrak{g}$ by the left translation L_g, we have

$$d\mathrm{L}_g^{-1} X = (\mathrm{Ad}\,h^{-1})\nu + \mu; \quad g = nh \in G. \tag{8.41}$$

Therefore for the left-invariant Riemannian structure on G we have

$$|X|^2 = |(\mathrm{Ad}\,h^{-1})\nu|_{\mathfrak{n}}^2 + |\mu|_{\mathfrak{h}}^2, \tag{8.42}$$

and in particular $\mathrm{T}N$ and $\mathrm{T}H$ are orthogonal.

Riemannian norms on replicas We shall preserve all the notation of the previous subsection. Then let $G_1 = N \leftthreetimes_1 H$ be some replica of $G = N \leftthreetimes H$ and let Ad_1 be defined for that replica as in §8.3.2. For fixed $n \in N$, with $h \in H$, we shall denote $(n, h) = n \leftthreetimes h$ (resp. $n \leftthreetimes_1 h$) their product in G (resp. G_1). For the identification of the tangent space at (n, h) as in (8.38) we then have the following two Riemannian norms on TG that are given by (8.42):

$$\begin{aligned} |X|^2 &= |(\operatorname{Ad} h^{-1})\nu|^2_{\mathfrak{n}} + |\mu|^2_{\mathfrak{h}}, \\ |X|^2_1 &= |(\operatorname{Ad}_1 h^{-1})\nu|^2_{\mathfrak{n}} + |\mu|^2_{\mathfrak{h}}. \end{aligned} \tag{8.43}$$

8.3.6 Quasi-isometric and polynomially equivalent replicas

Let $G_1 = N \leftthreetimes_1 H$, $G_2 = N \leftthreetimes_2 H$ be two replicas of G and let Ad_1 and Ad_2 (from $H \to \mathrm{GL}(\mathfrak{n})$) be the two corresponding mappings. Let us assume that there exists $C > 0$ such that

$$\left\|(\operatorname{Ad}_i h)(\operatorname{Ad}_j h^{-1})\right\| \leqslant C; \quad h \in H,\ i, j = 1, 2. \tag{8.44}$$

Then it is clear from (8.43) that for left-invariant Riemannian structures on G_1, G_2, the identity mapping is a quasi-isometry. An example of this situation is supplied by the replicas G_S and G_0 of §§8.3.3, 8.3.4.

Similarly, let us assume for the two replicas that there exist constants such that

$$\left\|(\operatorname{Ad}_i h)(\operatorname{Ad}_j h)^{-1}\right\| \leqslant C(1 + |h|_H)^C; \quad h \in H,\ i, j = 1, 2. \tag{8.45}$$

Then, again, formula (8.43) shows that the identity mapping between G_1 and G_2 is a polynomial equivalence. An example of this situation is supplied by the groups G and G_S of §8.3.3.

Exercise 8.4 Prove the above statement. The commutation between the semi-simple and nilpotent components in §8.3.3 shows that the left-hand side of (8.45) can be estimated by the operator norm on $\mathfrak{n}$ of $e^{\sigma_N(\xi)}$ with $h = \exp \xi$, where $\xi \in \mathfrak{h}$, and therefore also estimated by $(1 + |\xi|)^C$. We can then use the distance distortion (2.60), and (8.45) follows.

Exercise 8.5 If we denote the direct product by $G_T = N \times H$ we obtain the trivial replica of $G = N \leftthreetimes H$. Use (8.9) and the above to prove that G_T is polynomially equivalent to G if and only if G is an R-group. Of course, G_T is also polynomially equivalent to a Euclidean space (see Example 7.4).

Exercise 8.6 For any two replicas as above on the subset $[g = nh]$, with $|h| \leqslant C$, prove that the two Riemannian structures are quasi-isometric with constants that depend only on C.

8.4 Polynomial Sections

8.4.1 Definitions and examples

We shall consider a homomorphism

$$\pi : G \longrightarrow \frac{G}{N} = K, \tag{8.46}$$

where G, N, K are simply connected Lie groups and N is a closed normal subgroup of G. We then say that the mapping $\sigma : K \to G$ is a *section* if it is smooth and $\pi \circ \sigma = \text{Identity on } K$. Such sections always exist (see Varadarajan, 1974, §3.18). We shall say that σ as above is a *polynomial section* if there exist constants such that

$$|d\sigma| \leqslant c(1+|k|)^C; \quad k \in K, \tag{8.47}$$

where $d\sigma \colon \mathrm{T}K \to \mathrm{T}G$.

Example 8.7 If G is nilpotent we can use the exponential mapping $\exp\colon \mathfrak{g} \to G$ and Example 7.5, among other things, to see that every homomorphism (8.46) admits a polynomial section.

Example 8.8 (Soluble groups) See also Varopoulos (1994b, §1). Let us assume in (8.46) that G is a simply connected soluble group and that $N \lhd G$ is its nilradical. Then $G/N = K \cong \mathbb{R}^d$ is Abelian (see Varadarajan, 1974, §3.18). We can then find a section σ that is polynomial.

This example and the considerations that follow in the next one will not be essential to us in the sense that they are not used in the proofs of Theorems 7.10 and 7.11. We shall outline the proofs, however, because they already bring in some of the ideas that will be needed later.

Exercise 8.9 (The use of the Cartan subgroups in Example 8.8) We recall from §3.4 that we can find $H \subset G$, a closed connected *nilpotent* subgroup such that $NH = G$. It is of course by no means true in general that $N \cap H = \{e\}$. With the above notation, π reduces to a projection π_H:

$$\pi_H : H \longrightarrow \frac{H}{H \cap N} = K; \quad \sigma_H : K \longrightarrow H, \tag{8.48}$$

and from Example 8.7 there exists a polynomial section σ_H of π_H in the sense of (8.47) for which $\pi_H \circ \sigma_H = \text{Identity of } K$. However, $\sigma = \sigma_H \colon K \to H \subset G$ can be considered as a mapping from K to G, and then it is actually a polynomial section of π. Our assertion follows. Notice also, and this will be essential later, that $H \cap N$ is connected. This follows from the simple-connectedness of K and the lifting homotopy theorem for fibre spaces (see Hilton, 1953,

§V.1.2). But, using Varadarajan (1974, §3.18), we have already given a proof of this by direct methods in §2.14.2.

Example 8.10 (R-groups) Let the notation be as in Example 8.8. Then π, σ can be used to identify

$$G \underset{\theta^{-1}}{\overset{\theta}{\rightleftarrows}} N \times K. \tag{8.49}$$

This is achieved by writing $g \in G$ uniquely as a product $g = nx$ ($n \in N, x \in \Sigma = \sigma(K)$), so that we have

$$\begin{aligned} \theta(g) = (n, \pi(g)); \quad n = n(g) = g(\sigma \circ \pi(g))^{-1} \in N, \\ \theta^{-1}(n,k) = n\sigma(k) \in G. \end{aligned} \tag{8.50}$$

As seen in §8.1, the fact that σ is a polynomial does not suffice to guarantee that θ is polynomial. This, however, is a typical situation when formula (8.6), with $M = N \times K$, can be used.

In the case when G is an R-group, to show that θ is polynomial it suffices to show that $\theta_N : g \to n(g)$ as a mapping $G \to N$ is polynomial. But since $N \subset G$ is a subgroup it suffices to prove that $\theta_N : G \to G$ is polynomial. Then going back to the expression of $n(g)$ we see again that formula (8.6) applies with $M = G$ with $\phi_1 =$ Identity and $\phi_2 = (\sigma \circ \pi)^{-1}$. Then by (8.9) we have a proof of our assertion on θ, and since the same argument works for θ^{-1}, we see that (8.49) gives a polynomial equivalence. In other words, we have a proof of one of the assertions made in §7.3.1.

8.4.2 The strict polynomial section. Definition and statement of the results

Here we shall need to strengthen the notion of a polynomial section of the projection $\pi : G \to K$ of (8.46). Let σ be some polynomial section as in (8.47) and let us denote, throughout, $\Sigma = \sigma(K) \subset G$ which is a closed (embedded; see Hirsch, 1976, Chapter 1) submanifold. We shall then say that σ or Σ is a *strict polynomial section* if Σ is a polynomial retract of G.

To be precise, we assume that there exists $F : G \times [0,1] \to G$ that is locally Lipschitz and satisfies (see §7.4)

$$|dF(g,t)| \leqslant C(1+|g|)^C; \quad g \in G,\ 0 \leqslant t \leqslant 1, \tag{8.51}$$

for the product Riemannian structures and fixed constants. Furthermore, F is such that

$$F(g,1)=g, \quad F(g,0)=\sigma\circ\pi(g), \quad F(x,\lambda)=x; \\ g\in G,\ x\in\Sigma,\ 0\leqslant\lambda\leqslant 1. \tag{8.52}$$

We now come to one of the main technical tools used in this and the following chapter. We assume as in Example 8.8 that G is a soluble simply connected group, that N is the nilradical and that $H\subset G$ is a closed connected nilpotent subgroup such that $G=NH$ (see Varadarajan, 1974, §3.18.12). We can then use the action of H on N that is defined by inner automorphisms $(: n\to hnh^{-1})$ to construct the semidirect product $N\leftthreetimes H$ and the canonical projection (see §3.4.2)

$$\pi:\widetilde{G}=N\leftthreetimes H\longrightarrow G=NH. \tag{8.53}$$

Furthermore, the group $\widetilde{G}$ is C- (resp. NC-) if G is.

Exercise Prove this last point by elaborating §§2.3.3, 2.3.4 as follows. Let $\mathfrak{q}$ be some soluble algebra and let $\mathfrak{n},\mathfrak{h}\subset\mathfrak{q}$ be nilpotent subalgebras such that $\mathfrak{q}=\mathfrak{n}+\mathfrak{h}$ and $\mathfrak{n}$ is an ideal. Then the non-zero roots $\lambda_i(x)$ in (2.9) can be identified naturally with the non-zero roots of the $\operatorname{ad}\mathfrak{h}$-action on $\mathfrak{n}^c$. *Hint*. Start with $v_i,\ldots,$ some basis (over $\mathbb{C}$) of $\mathfrak{n}^c$ that triangulates the $\operatorname{ad}\mathfrak{q}$-action on $\mathfrak{n}^c$. Then complete with appropriate $\chi_1,\ldots\in\mathfrak{h}^c$ and obtain thus a basis of $\mathfrak{q}^c$ that triangulates the $\operatorname{ad}\mathfrak{q}$-action on $\mathfrak{q}^c$.

To avoid confusion in the notation and the proofs that follow, we shall in $\widetilde{G}$ denote the product $\tilde{g}=n\leftthreetimes h\in\widetilde{G}$, with $n\in N$, and $h\in H$ (this notation is not standard and is simply convenient for the occasion). Then

$$\pi(n\leftthreetimes h)=nh=\text{the product in } G, \tag{8.54}$$

and also, to be absolutely explicit, we can use the notation of §8.3.1 and then $\tilde{G}=N\leftthreetimes_s H$ with $s(h)n=hnh^{-1}$, with $h\in H, n\in N$.

Proposition 8.11 *With the above notation the projection π admits a strict polynomial section.*

The proof of the proposition will occupy the rest of this section. Observe, however, right away that once we have this proposition we have a polynomial homotopy equivalence between $\widetilde{G}$ and G (see §7.4 and also §12.2 below). We can combine this with the replicas of §8.3.4 and we almost have a proof of Proposition 8.3. The only essential element that is missing is that H is not necessarily Abelian. We shall come back to this when we finalise in §8.5.

8.4.3 One-parameter subgroups and notation

We shall denote by $\mathfrak{g}, \mathfrak{n}, \mathfrak{h}$ the Lie algebras of the groups G, N, H of the proposition. We shall fix $(\zeta_1, \dots, \zeta_r) \subset \mathfrak{n} \cap \mathfrak{h}$ and then complete it to $\eta_1, \dots, \eta_p \in \mathfrak{h}$ so that $(\zeta_1, \dots, \zeta_r, \eta_1, \dots, \eta_p) \subset \mathfrak{h}$ is an ordered basis of $\mathfrak{h}$ for which we can define exponential coordinates of the second kind in H as in §7.3.1.

Exercise Show that this is possible. Use the following facts: $\mathfrak{h}$ is nilpotent and $\mathfrak{n} \cap \mathfrak{h}$ is an ideal in $\mathfrak{h}$. The construction of this basis is then much easier than the general construction that is given in Varadarajan (1974, §3.18.12) because we can start from any basis of $\mathfrak{n} \cap \mathfrak{h}$ that is appropriately constructed in successive steps in the central series and complete that with a similar basis of $\mathfrak{h}$ (mod $\mathfrak{h} \cap \mathfrak{n}$).

With the notation from §7.3.1 we shall write

$$\left.\begin{aligned} e(\tau\zeta) &= e(\tau_1\zeta_1) \cdots e(\tau_r\zeta_r) \in H; \quad \tau = (\tau_1, \dots, \tau_r) \in \mathbb{R}^r, \\ e(t\eta) &= e(t_1\eta_1) \cdots e(t_p\eta_p) \in H; \quad t = (t_1, \dots, t_p) \in \mathbb{R}^p, \end{aligned}\right\} \tag{8.55}$$

where, as in §7.3.1, $e(t\xi) = e^{t\xi}$, with $t \in \mathbb{R}$, with $\xi \in \mathfrak{g}$, denotes the one-parameter subgroup generated by $\xi \in \mathfrak{g}$. With this notation and that of (8.54) we then have

$$G = [Ne(t\eta);\ t \in \mathbb{R}^p], \tag{8.56}$$

$$\widetilde{G} = [N \leftthreetimes e(\tau\zeta)e(t\eta);\ \tau \in \mathbb{R}^r,\ t \in \mathbb{R}^p], \tag{8.57}$$

and we shall define

$$\Sigma = [N \leftthreetimes e(t\eta);\ t \in \mathbb{R}^p] \subset \widetilde{G}, \tag{8.58}$$

$$\sigma(ne(t\eta)) = n \leftthreetimes e(t\eta) \in \widetilde{G}. \tag{8.59}$$

Exercise Verify that in (8.56) an element $g = ne(t\eta)$ uniquely determines $t \in \mathbb{R}^p$. This is necessary for the above definition to be legitimate. It is obvious because the images $d\phi(\eta_j)$ with $\phi: G \to G/N = H/H \cap N$ give coordinates on the Euclidean space $H/H \cap N$. Note also that these coordinates stay bounded by $C|g|$.

Let us now fix $g \in G$ and use the notation of §8.1.1 to define

$$\dot{v}_j = \dot{\rho}(v_j, g), \quad \dot{\eta}_j = \dot{\rho}(\eta_j, g) \in T_g(G); \quad j = 1, 2, \dots,$$

where $v_1, \dots, v_n \in \mathfrak{n}$ is a basis of the algebra $\mathfrak{n}$. The left-invariant Riemannian structures on G and $\widetilde{G}$ will then be given by requiring that

$$(v_1, \dots, v_n, \eta_1, \dots, \eta_p) \subset \mathfrak{g} \tag{8.60}$$

is orthonormal and

$$(\nu_1,\dots,\nu_n,\zeta_1,\dots,\zeta_r,\eta_1,\dots,\eta_p)\subset\tilde{\mathfrak{g}}=\mathfrak{n}\leftthreetimes\mathfrak{h} \tag{8.61}$$

is orthonormal.

8.4.4 Proving that σ is a polynomial section

By the above definitions, to prove that σ is a polynomial section it suffices to prove that there exist constants such that

$$\begin{aligned}&\left|d\sigma(\dot{\nu}_j)\right|\leqslant C(1+|g|)^C,\\&\left|d\sigma(\dot{\eta}_j)\right|\leqslant C(1+|g|)^C;\quad g\in G,\ j=1,2,\dots.\end{aligned} \tag{8.62}$$

The first assertion in (8.62) is easy. Indeed, if $g=nh$ for $n\in N$, $h\in H$, with $h=e(t\eta)$ as in (8.56), then $\sigma(g)=n\leftthreetimes h=\tilde{g}\in\Sigma\subset\widetilde{G}$. As in §8.1.1 we can then define

$$\begin{aligned}&\rho_j(t)=\rho_j(t,\nu_j,g)=ge^{t\nu_j}\in G,\\&\tilde{\rho}_j(t)=\rho_j(t,\nu_j,\tilde{g})=\tilde{g}e^{t\nu_j}=\tilde{n}\leftthreetimes h\in\widetilde{G},\end{aligned} \tag{8.63}$$

where here $\tilde{n}=he^{t\nu_j}h^{-1}\in N$ depends on t. Clearly the above cosets of one-parameter groups correspond by the mappings π (: $\tilde{\rho}_j(t)\underset{\pi}{\rightarrow}\rho_j(t)$), and by (8.63) it follows that $\tilde{\rho}_j(t)\subset\Sigma$ and therefore that $\sigma\rho_j=\tilde{\rho}_j$. The conclusion is that

$$d\sigma(\dot{\nu}_j)=\dot{\rho}(\nu_j,\tilde{g}), \tag{8.64}$$

where the $\dot{\nu}_j$ on the left is identified to an element of TG and on the right we have an element of T$\widetilde{G}$. This completes the proof of the first assertion of (8.62).

8.4.5 Proof of the second assertion of (8.62)

This is more subtle because although we can define $\dot{\rho}(\eta_j,\tilde{g})\in\mathrm{T}\widetilde{G}$ as before, this vector is not in general tangent to Σ. To get round this difficulty we first need to introduce appropriate additional notation. Let $g\in G$ be as before, where we use the coordinates of (8.56) to write

$$g=ne(t\eta);\quad n\in N,\ t=(t_1,\dots,t_p)\in\mathbb{R}^p, \tag{8.65}$$

in a well-determined unique way. For $t'\in\mathbb{R}$, we shall set $t^{(j)}=(t_1,\dots,t_j+t',t_{j+1},\dots,t_p)$, that is, we add t' to the jth coordinate. We shall then set

$$E_j(t')=e(t\eta)e(t'\eta_j)=h_1(t')e\left(t^{(j)}\eta\right)=h_1(t')h_2(t');\quad h_1,h_2\in H. \tag{8.66}$$

In (8.66), $h_1 \in [H,H]$ is obtained by the commutation operations that are needed to bring the additional one-parameter subgroup to the jth place and then also to bring the error term so obtained at the beginning of the product. Here $h_2(t') = e(t^{(j)}\eta)$ is a function of t' with the other coordinates fixed.

(a) Now, to obtain the tangent vector that corresponds to η_j in TG at g we have to take the derivative at $t' = 0$ of $nE(t') \in G$. With the notation of (8.1), this is $dL_n\dot{E}_j(0)$.

(b) To obtain the tangent vector that corresponds to η_j in T$\widetilde{G}$ at $\tilde{g} = n \leftthreetimes \mathrm{e}_0$ with $\mathrm{e}_0 = \mathrm{e}(t\eta)$ we have to take the derivative at $t' = 0$ of $n \leftthreetimes E_j(t')$ which again, with the notation (8.1), is $dL_n\dot{E}_j(0)$.

Here we have abused notation slightly and denoted by L_n the left translation by elements of N in both G and $\widetilde{G}$; moreover, H was considered as a subgroup of both G and of $\widetilde{G}$. In each case, $\dot{E}_j(0) \in \mathrm{T}_{\mathrm{e}_0}H$ and this, as well as the two $\mathrm{d}L_n\dot{E}_j(0)$ in (a) and (b), has length 1.

The problem that arises is that, unlike the previous case in §8.4.4, it is not true in general that $\varphi_0(t') = n \leftthreetimes E_j(t') \subset \Sigma$ and therefore the tangent vector that we have obtained in (b) is not necessarily tangent to Σ. To cope with this difficulty we modify φ_0 and with the two factors h_1, h_2 of (8.66) we write

$$\varphi(t') = nh_1(t') \leftthreetimes h_2(t') \in \widetilde{G}. \tag{8.67}$$

This makes sense since $h_1 \in [H,H] \subset N$ (see the discussion following (8.72) below). Now we clearly have

$$\varphi(t') \subset \Sigma, \qquad \pi\big(\varphi(t')\big) = \pi\big(\varphi_0(t')\big), \tag{8.68}$$

and therefore, with the notation of §8.1.1, we have

$$d\sigma\,\dot{\eta}_j = \dot{\varphi}(0). \tag{8.69}$$

It suffices then to prove that

$$|\dot{\varphi}(0)| \leqslant C(1+|g|)^C; \quad g \in G \tag{8.70}$$

and the second assertion (8.62) follows. Furthermore, since the Riemannian structure on $\widetilde{G}$ is left invariant, we can use L_n^{-1} and reduce the problem to the case when, in (8.65), $n = e$ is the identity. In other words, in (8.67) we can assume

$$\varphi(t') = h_1(t') \leftthreetimes h_2(t') \in \widetilde{G}. \tag{8.71}$$

But in this special case,

$$\varphi(t') \in (N \cap H) \leftthreetimes H = \widetilde{G}_N \tag{8.72}$$

because $h_1 \in [H,H] \subset N \cap H$. Note that the subgroup $[H,H]$ generated by all the group commutators $xyx^{-1}y^{-1}$, with $x,y \in H$, is the analytic subgroup of H whose Lie algebra is $\mathfrak{h}^2 = [\mathfrak{h},\mathfrak{h}]$ and $\mathfrak{h}^2 \subset \mathfrak{n}$ (see Varadarajan, 1974, §3.8.3; this description of the commutator subgroup was explained in Varadarajan, 1974, §3.18.8). The group $\widetilde{G}_N$ is nilpotent because the ad-action of $\mathfrak{h}$ on $\mathfrak{n} \cap \mathfrak{h}$ is clearly nilpotent, and this together with the nilpotency of H implies that ad$\,\xi$ (ξ in the algebra of the group) is nilpotent (see Varadarajan (1974, §3.5) or Chevalley (1955, V, §2.2, Proposition 4)). The reason for this is that in that algebra the roots defined in (2.9) by Lie's theorem have to vanish on $\mathfrak{h}$ and on $\mathfrak{h} \cap \mathfrak{n}$. On the other hand, by their construction, the two mappings h_1, h_2 are products of one-parameter subgroups of the nilpotent group H; they are therefore polynomial mappings in the sense that $\dot{h}_1$, $\dot{h}_2$ can be estimated by $1+|t|^C$, where again we use the notation of §8.1.1. Our assertion (8.70) follows from this by applying (8.7), (8.9), or by applying Example 7.5 to the nilpotent group (8.72). At the end, we estimate $|t|$ by $|g|$ as explained in §8.4.3 (see the exercise in §8.4.3).

8.4.6 Proof that σ is strictly polynomial

The issue is to construct the homotopy that satisfies the conditions of the definition in (8.51), (8.52). We shall use (8.57), that is, the fact that every $\tilde{g} \in \widetilde{G}$ can be written uniquely as

$$\tilde{g} = n \leftthreetimes h = n \leftthreetimes e(\tau\zeta)e(t\eta). \tag{8.73}$$

Moreover, the maps $\widetilde{G} \to H$, $\tilde{g} \to h$ in (8.73) are polynomial, and the dependence of τ, t on h is polynomial because H is nilpotent. The only difficulty is that in general the mapping $\tilde{g} \to n$ in (8.73) is not polynomial.

We shall now define

$$\begin{aligned} &F(\tilde{g},\lambda) = n \leftthreetimes e(\lambda\tau\zeta)e(t\eta) \in \widetilde{G};\\ &\tilde{g} = n \leftthreetimes h,\ 0 \leqslant \lambda \leqslant 1,\ \tau \in \mathbb{R}^r,\ t \in \mathbb{R}^p, \end{aligned} \tag{8.74}$$

where, to clarify the notation, we recall that $e(\lambda\tau\zeta) = e(\lambda\tau_1\zeta_1)e(\lambda\tau_2\zeta_2)\cdots$, and that τ, t are the polynomial functions of $\tilde{g}$ defined in (8.73). It is clear that (8.52) is satisfied and to prove that this gives a strict polynomial section it suffices to estimate dF and verify (8.51) for the product Riemannian structure $\widetilde{G} \times [0,1] = M$. To do this we shall rewrite

$$\begin{gathered} F(\tilde{g},\lambda) = n \leftthreetimes e(\tau\zeta)e(t\eta)e((\lambda-1)\tau)h^* \in \widetilde{G},\\ h^* = h(\tau,t,\lambda) \in [H,H] \subset N \cap H, \end{gathered} \tag{8.75}$$

where to obtain h^* we write each one-parameter group

$$e(\lambda \tau_j \zeta_j) = e(\tau_j \zeta_j) e((\lambda - 1)\tau_j \zeta_j); \quad j = 1, \ldots \tag{8.76}$$

and then use the necessary commutators in H of the form $[x,y] = xyx^{-1}y^{-1}$, with $x, y \in H$, to bring the second factors in (8.76), and the correcting terms coming from the commutators, at the end of the product in (8.75). What counts is that from the above definitions we have that the mapping

$$\widetilde{G} \times [0,1] \longrightarrow H \cap N; \quad (\tilde{g}, \lambda) \longrightarrow h^* \tag{8.77}$$

is polynomial. This follows from our original remarks just after (8.73) that make the mapping from $\widetilde{G} \times [0,1]$ into H polynomial and the nilpotency of H. We can also rewrite

$$\begin{aligned} F(\tilde{g}, \lambda) &= \tilde{g} e((\lambda - 1)\tau) h^* = \tilde{g} \phi_2, \\ H \cap N \ni \phi_2(m) &= e((\lambda - 1)\tau) h^*; \quad m \in M = \widetilde{G} \times [0,1]. \end{aligned} \tag{8.78}$$

In (8.78) the product is taken in $\widetilde{G} = N \leftthreetimes H$ and ϕ_2 lies in the cofactor H. More precisely, $\phi_2 \in H \cap N \subset H \subset \widetilde{G}$. In other words, if $\phi_1(m) = \tilde{g}$ we are in the set-up where formula (8.6) applies. However, (8.7) also applies because $\mathrm{Ad}(H \cap N)$ is polynomial on $\tilde{\mathfrak{g}} = \mathfrak{n} \leftthreetimes \mathfrak{h}$ in the sense of (8.7). Indeed, for $x \in H \cap N$ and $\tilde{\xi} = \nu + \chi \in \tilde{\mathfrak{g}}$, with $\nu \in \mathfrak{n}$, $\chi \in \mathfrak{h}$, we have $(\mathrm{Ad}x)\tilde{\xi} = (\mathrm{Ad}x)\nu + (\mathrm{Ad}x)\chi$. In applying (8.7), $H \cap N$ is considered as a subgroup of the cofactor H in $\widetilde{G}$ and therefore also as a subgroup of $\widetilde{G}$ and $\mathrm{Ad}x$ in the above formula refers to the action of $\widetilde{G}$ on its Lie algebra. The estimate

$$|dF| \leqslant C(1 + |\tilde{g}|)^C; \quad \tilde{g} \in \widetilde{G},\ 0 \leqslant \lambda \leqslant 1, \tag{8.79}$$

therefore follows and this proves our assertion. The nilpotency of N, H and (8.9) are used in the above.

8.4.7 A variant of the argument

The argument in §§8.4.4–8.4.6 can be adapted in the following situation.

Let $\widetilde{G} = N \leftthreetimes H$, where N, H are assumed simply connected and H is nilpotent. Furthermore, let $H_1 \subset H$ be some normal connected closed subgroup that centralises N. This means that for all $h \in H_1$, if we restrict $\mathrm{Ad}h$ to the Lie algebra of N we obtain $\mathrm{Ad}_{\mathfrak{n}}(h) = \text{Identity}$. We then define $\pi_H : H \to H/H_1 = H_2$ and extend this to a homomorphism $\pi : \widetilde{G} \to N \leftthreetimes H_2 = G$ by setting $\pi(n \leftthreetimes h) = n \leftthreetimes \pi_H(h)$, where the semidirect product is defined by the induced action of H_2 on N. For every section $\sigma_H : H_2 \to H$ (i.e. $\pi_H \circ \sigma_H = \text{Identity}$) we can then define the corresponding section $\sigma : G \to \widetilde{G}$. We then have the following proposition.

Proposition 8.12 *Let the notation be as above and assume that* $[H,H] \subset H_1$. *Then we can construct a polynomial section* $\sigma_H : H_2 \to H$ *in such a way that*

(i) σ *is polynomial*;
(ii) σ *is strictly polynomial*.

Exercise 8.13 Adapt the proofs that we gave in §§8.4.3–8.4.6 to prove the first part of the proposition.

Here are the changes in the definitions that have to be made.

Part (i) of Proposition 8.12 Now $(\zeta_1, \ldots, \zeta_r) \subset \mathfrak{h}_1$ is a basis of the Lie algebra of H_1 and we complete it to $(\zeta_1, \ldots, \zeta_r, \eta_1, \ldots, \eta_p) \subset \mathfrak{h}$ to form a basis of the Lie algebra of H, for which we can define exponential coordinates of the second kind in H as in §8.4.3. Expressions (8.55) are defined exactly as before and the definitions (8.57) and (8.58) are used here again and are identical, but now instead of (8.56) and (8.59) we set

$$\begin{aligned} G &= [N \leftthreetimes \pi_H(e(t\eta));\ t \in \mathbb{R}^p], \\ \sigma(n \leftthreetimes \pi_H(e(t\eta))) &= n \leftthreetimes e(t\eta) \in \widetilde{G}. \end{aligned} \tag{8.80}$$

In all of this, the notation that we introduced in (8.54) is of course used.

With this notation essentially nothing changes in §§8.4.3–8.4.4. Adapting the argument of §8.4.5 is very simple. Indeed, with the notation of (8.66), $h_1(t') \in [H,H] \subset H_1$ by our hypothesis. The rest of the argument applies with

$$\varphi_0(t') = n \leftthreetimes h_1(t') h_2(t'), \qquad \varphi(t') = n \leftthreetimes h_2(t') \tag{8.81}$$

instead of (8.67), where the definitions of h_1, h_2 are as in (8.66). The details are in fact simpler than before and are left for the reader.

Exercise 8.14 (Part (ii) of Proposition 8.12) Exactly as for part (i), the proof is an easy adaptation of §8.4.6. We change the notation as explained in Exercise 8.13. The group $\widetilde{G}$ is still $N \leftthreetimes e(\tau\zeta)e(t\eta)$, as in (8.57), with the new definitions of ζ and η of Exercise 8.13 and G, σ as in (8.80). The definition of the retract is as in (8.75):

$$F(\tilde{g}, \lambda) = n \leftthreetimes e(\tau\zeta)e(t\eta)e((\lambda-1)\tau\zeta)h^*, \qquad \tilde{g} = n \leftthreetimes e(\tau\zeta)e(t\eta),$$

with the only difference being that $h^* \in [H,H] \subset H_1$, and where, as in (8.78), we set $\phi_2 = e((\lambda-1)\tau\zeta)h^* \in H_1$. To conclude the polynomial estimate (8.79) for this new F we argue as at the end of §8.4.6 where we use $\mathrm{Ad}_{\tilde{\mathfrak{g}}}(H_1)$ with $\tilde{\mathfrak{g}}$ the Lie algebra of $\widetilde{G}$. For this we use the nilpotency of H and the fact that H_1 centralises N. The end of proof is identical and the reader can finish things up without any difficulty.

8.5 Dénouement

8.5.1 Proofs of Proposition 8.3 and Theorem 7.10

If we combine Proposition 8.11 with §§8.3.4 and 8.3.6 we see that if Q is some simply connected soluble group then there exist N, H, two simply connected nilpotent groups, and there exists also an action of $\mathfrak{h}$, the Lie algebra of H, on $\mathfrak{n}$, the Lie algebra of N that is given by (real) diagonal matrices

$$\operatorname{ad}\xi = \operatorname{diag}(L_1\xi, \dots, L_n\xi); \quad \xi \in \mathfrak{h}. \tag{8.82}$$

The $L_j \in \mathfrak{h}^*$ are the roots of that action, and in this construction the semidirect product $G = N \leftthreetimes H$ is polynomially homotopically equivalent to Q. Furthermore, G is C- (resp. NC-) if Q is and in this construction we can even assume that H is Abelian since Proposition 8.12 applies (cf. §8.5.1.2 below). We have thus a proof of Proposition 8.3. What has to be seen therefore is that the conclusion of Theorem 7.10 holds for this new groups G. Towards that and in view of Propositions 8.2 and 8.3, we shall make two additional easy reductions.

8.5.1.1 A reduction when Q is an NC-group In this case, as we have already said, the group G that we constructed above is also NC. The easy case here is when Q is an R-group and all the roots in (8.82) are zero: $L_i = 0$. In that case $G = N \oplus H$ is a direct sum and is a nilpotent group.

If G is not an R-group, we fall back on the constructions (and notation) of the exercise in §8.3.4, denote by $\mathfrak{n}_{L_j} = \mathfrak{n}_j \subset \mathfrak{n}$ the root spaces in the Lie algebra $\mathfrak{n}$ of N that correspond to the roots L_j in (8.82) and write $\mathfrak{n}_{\mathrm{R}} = \sum_{L_j \neq 0} \mathfrak{n}_j$ and $\mathfrak{n}_0$ for the root space that corresponds to the zero root $L_0 = 0$ if such a root exists. Otherwise we set $\mathfrak{n}_0 = 0$. It is obvious that $[\mathfrak{n}_0, \mathfrak{h}] = 0$ and that $\mathfrak{h}_{\mathrm{R}} = \mathfrak{n}_0 + \mathfrak{h}$ is a nilpotent subalgebra. Furthermore (see §3.8), $\mathfrak{n}_{\mathrm{R}}$ is an ideal and $\mathfrak{g}$, the Lie algebra of G, can be written $\mathfrak{g} = \mathfrak{n}_{\mathrm{R}} \leftthreetimes \mathfrak{h}_{\mathrm{R}}$ and therefore $G \simeq N_{\mathrm{R}} \leftthreetimes H_{\mathrm{R}}$ for the corresponding groups. This gives a new decomposition for the groups G with an action of $\mathfrak{h}_{\mathrm{R}}$ on $\mathfrak{n}_{\mathrm{R}}$. This action may well not be semisimple but if we then take the semisimple replica of this new semidirect product we have a group G exactly as in (8.82) but here all the roots $L_j \neq 0$ are non-zero, as was one of the prerogatives in Proposition 8.2: see (8.17).

8.5.1.2 Reduction to make $\mathfrak{h}$ Abelian We start with $G = N \leftthreetimes H$ as in (8.82) and set $H_1 = [H, H]$ the commutator subgroup: it is an analytic subgroup with Lie algebra $[\mathfrak{h}, \mathfrak{h}]$. We then consider $\pi\colon H \to H/H_1 = A$: we can make the construction of Proposition 8.12 because by (8.82) H_1 acts trivially on N. Furthermore, the roots of the action of A on N are given by $L_j \circ (d\pi)$, which is well defined, and we have, with these roots, a new group $N \leftthreetimes A$ that satisfies (8.82). In addition, A is Abelian, as needed in Proposition 8.2.

Putting together (8.82) with the two above reductions we can finally give the proof.

Proof of Theorem 7.10 Let Q be some simply connected soluble NC-group.

(i) If Q is an R-group then Q is polynomially homotopically equivalent to a nilpotent group (see Example 8.10 or Exercise 8.5).

(ii) If Q is not an R-group then Q is polynomially homotopically equivalent to a special soluble group that satisfies the conditions of Proposition 8.2.

In both cases therefore, either by Proposition 8.2 or by Example 8.10, the group so obtained, and therefore also the original group Q, is polynomially retractible. This completes the proof of the NC-theorem (Theorem 7.10). □

.

8.5.1.3 A class of special soluble groups We have just seen that every simply connected soluble group G is polynomially homotopically equivalent to some special soluble group of the form $N \leftthreetimes A$ that satisfies the conditions of Proposition 8.3 where $A = \mathbb{R}^\rho$ for some $\rho \geqslant 0$. In view of Theorem 7.11 for C-groups, it is of interest to note that it is possible to choose

$$\rho = \dim(G/\text{nilradical of } G).$$

To see this we just have to follow the proofs given in this chapter with special emphasis on §8.5.1.2, which shows that ρ can be chosen to be $\dim(\mathfrak{h}/[\mathfrak{h},\mathfrak{h}])$. We then use the remark in §3.4.4. The details will be left to the reader.

Open problems Let us write $\rho_0(G)$ for the minimum possible value of $\rho \geqslant 0$ for which G is polynomially homotopically equivalent to $N \leftthreetimes \mathbb{R}^\rho$. For instance, $\rho_0(G) = 0$ if and only if G is an NC-group.

Open problem 1 Can the geometry of the roots of G (in the sense of Chapter 2) be used to determine $\rho_0(G)$?

Open problem 2 Is it conceivable that we can classify simply connected soluble groups up to polynomial homotopy equivalence? Or at least obtain invariants similar to $\rho_0(G)$?

8.5.2 Comments on the proof of Theorem 7.10

In the previous subsection we gave a proof of the theorem by combining Propositions 8.2 and 8.3. These show that any soluble simply connected group is polynomially homotopically equivalent to a group that satisfies the conditions of Proposition 8.3. Then we apply Proposition 8.2.

There is, however, an alternative approach to the theorem which is the one that was originally adopted in Varopoulos (2000b). This approach avoids the use of the replicas of §8.3.3 and also of the mapping $N \leftthreetimes H \to NH$ in (8.53) that uses Cartan subgroups. This approach relies on the following two facts:

(i) We can write G, some soluble simply connected NC-group, as $G = N_R \leftthreetimes Q_R$, where N_R is nilpotent and Q_R is a soluble R-group. This was explained in §3.8. Note, however, that Cartan subgroups are used for that construction.

(ii) The soluble simply connected R-group Q_R is polynomially retractable. This was proved in Example 8.10. In the proof that we gave of this fact we again made use of Cartan subgroups. Alternative but more or less equivalent proofs of this fact exist (see Appendix F and also the original proof in Varopoulos, 2000b).

One feature of this alternative approach of the NC-theorem that is worth noting is that the proof is done in two stages. It is first done for R-groups in (ii) and although not trivial, it certainly is not surprising. Then we use (i) and an argument that is entirely analogous to §8.2.3 to conclude the proof in general.

This alternative approach is 'messier' but it should be noted that it generalises and gives the first half of the final theorem in the area (see §1.6.3 and Appendix F).

8.5.3 Comments on the definition of a strict section and the polynomial homotopy equivalence

We shall end this chapter by making some comments on the conditions of the polynomial sections in (8.52). There we actually demanded more than was necessary for our purposes. But – and this presents some independent interest – with the same notation, instead of (8.52), suppose we impose the following three conditions: $F(g,1) = g$, $F(g,0) \in \Sigma$, $F(x,\lambda) = x$; for $g \in G$, $x \in \Sigma$, $0 \leqslant \lambda \leqslant 1$. Then we can assert that G admits the $\mathscr{F}$ property of Definition 7.12 if and only if K does. (We leave the proof of this to the reader.)

The above remark is related to a more general problem in homotopy theory: in loose terms the issue is what, if anything, stays fixed in a homology

$H(m,t) \in M$, for $m \in M$, $0 \leqslant t \leqslant 1$ (i.e. points $m_0 \in M$ for which $H(m_0,t) = m_0$ for $0 \leqslant t \leqslant 1$)? For instance, in the original definition in §7.4.1 it might have been more consistent with the context to impose the additional condition that $F(O,t) = O_1$ (with the notation that we used there). As it happens (this will be examined in more detail in Chapter 12), the modification makes no difference for our purposes because the additional condition comes for free for the homotopies that we shall be using.

9

Algebra and Geometry on C-Groups

Overview of Chapters 9 and 10

In these two chapters we shall give the proof of the C-theorem (Theorem 7.11) for simply connected soluble C-groups. The proof is very long and, at least in this approach to the problem, this is probably inevitable.

To clarify how the proof is done in the two different chapters, the following comments are in order. *Grosso modo*, we could say that in Chapter 9 a number of intricate geometric constructions are made in a special class of soluble groups, which we call SSG for short. Then, in Chapter 10, these constructions are used to give the proof of the theorem for these SSG. The method that allows us to pass from these SSG to the general case is based on the polynomial homotopy equivalences developed in Chapter 8.

However, this overall organisation is relative. For instance, in Chapter 9 we already explain, in broad terms, how the proof finishes in order to justify the reasons for the constructions. And in Chapter 10 we give further refinements of the constructions of Chapter 9.

Section 9.1 is special and stands apart from the theory of Lie groups because it consists of linear algebra and finite geometry. The SSA are explicitly defined there.

The first basic construction is carved out in §§9.3 and 9.4. The overall description is given in §9.3, and the details in §9.4. In §9.5 we give the second basic construction. This involves a different idea but, once we have it, the details are much easier to describe than those of the first construction. Section 9.2 is also special and in it we describe a number of examples that existed before the C–NC classification and the proof of the theorem.

Before embarking on the details of the constructions, it is important for the reader to understand why in §§9.3–9.4 we work so hard to construct such a specific piecewise linear S (i.e. a triangulation of the d-dimensional sphere) embedded in $\mathbb{R}^{2d+1}$. It is here that spending some time in §9.3 on the general description of the first basic construction, and in §9.2 on the examples, will pay off.

However, given how long the construction is, the reader basically has just two options.

The first is to understand how the examples in §9.2 work – this is very limited but is also very easy – and then simply believe that the general construction described in §9.3 works. In this case the reader can skip altogether the tricky §9.4.

In the second option, which will require time and energy, the reader will also have to develop a certain 'tolerance' (preferably a 'liking') for the finite geometry of §9.1 and the piecewise linear constructions of §9.4. Otherwise, we warn, these two sections will be unbearably hard to digest.

Finally, for the purpose of getting a global perspective and the prospect of avoiding many of the geometric constructions that we are about to describe, one could have a quick look at the introduction to Part III and at the epilogue of the book.

9.1 The Special Soluble Algebras

9.1.1 Algebraic considerations

We shall start by giving the formal definition of a special soluble algebra (SSA). These have already been used in Proposition 8.3.

Definition 9.1 We say that $\mathfrak{g}$ is a *special soluble algebra* if $\mathfrak{g} = \mathfrak{n} \rightthreetimes \mathfrak{a}$ where $\mathfrak{n}$ is nilpotent and $\mathfrak{a} \neq \{0\}$ Abelian and where the action of $\mathfrak{a}$ on $\mathfrak{n}$ is semisimple with real roots.

We shall elaborate on this definition and codify some notation that will stay fixed throughout. We write $V = \mathfrak{a}^* \neq \{0\}$ for the dual space but exclude the case $\mathfrak{a} = 0$ from the definition. We shall fix $E \subset V$, a finite set which will be called the *set of the roots*. When $\mathfrak{n} \neq \{0\}$ then $E \neq \emptyset$ and we can decompose

$$\mathfrak{n} = \bigoplus_{e \in E} \mathfrak{n}_e, \quad \text{the root space decomposition,} \tag{9.1}$$

where (see §2.1 and Jacobson, 1962, §II.4)

$$\mathfrak{n}_e = [v \in \mathfrak{n};\ (\operatorname{ad}\xi)v = e(\xi)v,\ \xi \in \mathfrak{a}] \tag{9.2}$$

and where $\mathfrak{n}_e$ is assumed to be $\neq 0$. This implies the basic relation

$$[\mathfrak{n}_e, \mathfrak{n}_{e'}] \subset \begin{cases} \mathfrak{n}_{e+e'} & \text{if } e+e' \in E, \\ 0 & \text{if } e+e' \notin E. \end{cases} \tag{9.3}$$

In the above notions the elements of E are distinct and are counted once, that is, no notion of multiplicity is involved but $\dim \mathfrak{n}_e \geqslant 1$. The additional point to be stressed is that in this definition we are not only defining the algebra $\mathfrak{g}$ but also the splitting $\mathfrak{n} \leftthreetimes \mathfrak{a}$ and therefore also (9.1).

Definition 9.2 Let $\mathfrak{n}' \subset \mathfrak{n}$ be some subalgebra and assume that $[\mathfrak{n}', \mathfrak{a}] \subset \mathfrak{n}'$. Then $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}$ is a subalgebra of $\mathfrak{g}$. It is an SSA with the induced structure and will be called an *SS-subalgebra* of $\mathfrak{g}$. From the definition and elementary linear algebra it follows that

$$\mathfrak{n}' = \bigoplus_{e \in E} \left(\mathfrak{n}' \cap \mathfrak{n}_e\right). \tag{9.4}$$

In the above definition the roots of $\mathfrak{g}'$ form a subset $E' \subset E$. We have, for instance, $E' = \emptyset$ in the trivial case $\mathfrak{n}' = \{0\}$.

We stress the point that in the above definition $\mathfrak{a} \neq \{0\}$ stays put and fixed. It is only the $\mathfrak{n}' \subset \mathfrak{n}$ that varies.

Definition 9.3 Let $\mathfrak{a}_2 = \bigcap_{e \in E} \operatorname{Ker} e$, that is, $\mathfrak{a}_2 = V_1^{\perp}$ is the orthogonal subspace for the duality between $\mathfrak{a}$ and V, with $V_1 = \operatorname{Vec}(e, e \in E)$. Let $\mathfrak{a}_1 \subset \mathfrak{a}$ be some complementary subspace so that $\mathfrak{a} = \mathfrak{a}_1 \oplus \mathfrak{a}_2$. Then clearly

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{a}_2, \qquad \mathfrak{g}_1 = \mathfrak{n} \leftthreetimes \mathfrak{a}_1, \qquad V_1 = \left(\frac{\mathfrak{a}}{\mathfrak{a}_2}\right)^* = \mathfrak{a}_1^*. \tag{9.5}$$

When $E \neq \{0\}$ this means that once we have fixed the complement $\mathfrak{a}_1$ we obtain the decomposition (9.5) where $\mathfrak{g}_1 = \mathfrak{n} \leftthreetimes \mathfrak{a}_1$ is also an SSA in its own right (because $\mathfrak{a}_1 \neq 0$) where now the roots E_1 of $\mathfrak{g}_1$ span $V_1 = \mathfrak{a}_1^*$. Note that $\mathfrak{a}_2$ could be zero, but $\mathfrak{a}_1 \neq 0$ unless $E = \{0\}$.

The two extreme cases are $\mathfrak{g} = \mathfrak{n} \times \mathfrak{a}$ and the action of $\mathfrak{a}$ on $\mathfrak{n}$ vanishes and then the roots $E = \{0\}$. At the other extreme, $\mathfrak{a}_2 = 0$ and E spans $\mathfrak{a}^*$ and we then say that $\mathfrak{g}$ is *irreducible*.

Definition 9.4 We say that $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ is an *SSA of Abelian type* if $\mathfrak{n}$ is Abelian.

Definition 9.5 We recall here the definition of a *Heisenberg algebra*. This is the nilpotent algebra $\mathfrak{h} = \mathfrak{h}_s = (\nu, \mu, \zeta)$ that depends on the parameter $0 \neq s \in \mathbb{R}$ and is generated by three vectors that satisfy the relations

$$[\nu, \mu] = s\zeta; \quad [\nu, \zeta] = [\mu, \zeta] = 0. \tag{9.6}$$

By rescaling we may assume that $s = 1$.

Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ be some SSA. We shall say that $\mathfrak{g}$ is of *Heisenberg type* if there exists $0 \neq L \in E$ such that $-L \in E$ and such that the root space decomposition (9.1) becomes

$$\mathfrak{n} = \mathfrak{n}_L \oplus \mathfrak{n}_{-L} \oplus \mathfrak{n}_0, \tag{9.7}$$

where $\mathfrak{n}_0$ is the root space with zero root $0 \in E$. Furthermore, we assume that (9.7) is a Heisenberg algebra, that is, that $\mathfrak{n}_L, \mathfrak{n}_{-L}$ are both one-dimensional and $\mathfrak{n}_0 = [\mathfrak{n}_L, \mathfrak{n}_{-L}] \neq 0$. In that definition $\mathfrak{n}_0$ is central in $\mathfrak{n}$.

Definition 9.6 We shall say that $\mathfrak{g}$ some SSA is a *special soluble* C*-algebra* (SSCA) if it is a C-algebra.

This abbreviation is a trifle 'heavy'. Already the abbreviation SSA is bad enough for possibly different but obvious and rather sinister reasons (that must be clear to all familiar with twentieth-century German history). But it is useful. To wit,

the SSA algebra $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ is an SSCA if it satisfies the C-condition of §2.2. This means that with the notation $X = E \setminus \{0\}$, *which will be adopted throughout*, we have

$$0 \in \text{convex hull of } X = \mathrm{CH}(X) = \sum_{x \in X} \lambda_x x, \quad \text{with } \lambda_x \geqslant 0, \sum \lambda_x = 1. \tag{9.8}$$

To see this use the fact that the nilradical of $\mathfrak{g}$ contains $\mathfrak{n}$, or use §2.3.3.

This definition forces X to have at least two points. The aim of this first section will be to prove the following basic result.

Theorem 9.7 (The algebraic structure theorem) *Let* $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ *some* SSCA *as above. Then there exists* $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}$ *some* SS*-subalgebra that is also a* C*-algebra and furthermore* $\mathfrak{g}'$ *is of Abelian type or of Heisenberg type.*

Remark 9.8 The theorem and its proof provide a different point of view of the notion. We could instead say that we start with $\mathfrak{a}$, some non-zero Abelian algebra, and identify it to a commuting family of linear mappings $\alpha\colon \mathfrak{n} \to \mathfrak{n}$, with $\alpha \in E$, where $\mathfrak{n}$ is some nilpotent algebra and we demand that

(i) α is a semisimple transformation with real roots;
(ii) α is a derivation of $\mathfrak{n}$.

In this light, the case $\mathfrak{n} = 0$, and $E = \emptyset$, is trivial.

The proof of the theorem will be given in §9.1.6, but before we can do that we must develop the necessary machinery.

Observe also that one can 'skip' all this and go straight to §9.1.7, where we recapitulate this structure theorem and prepare it for use in our geometric theory.

9.1.2 Bracket-reduced SSA

Definition 9.9 All the notation of the previous subsection is preserved and $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ is an SSA. We shall write throughout

$$\begin{aligned} A &= [e \in E;\ \mathfrak{n}_e \cap [\mathfrak{n},\mathfrak{n}] = 0], \\ B &= E \backslash A = [e \in E;\ \mathfrak{n}_e \cap [\mathfrak{n},\mathfrak{n}] \neq 0], \qquad (9.9)\\ \mathfrak{n} &= \mathfrak{n}_A + \mathfrak{n}_B = \bigoplus_{\alpha \in A} \mathfrak{n}_\alpha \oplus \bigoplus_{\beta \in B} \mathfrak{n}_\beta. \qquad (9.10)\end{aligned}$$

Either A or B (or both if $\mathfrak{n} = 0$) could be $\emptyset$, and $B = \emptyset$ if and only if $\mathfrak{n}$ is Abelian. More precisely, $[\mathfrak{n},\mathfrak{n}]$ is stable under the $\mathfrak{a}$-action and therefore by (9.4), $\mathfrak{n}_B \supset [\mathfrak{n},\mathfrak{n}]$ and B is exactly the set of the roots of the action of $\mathfrak{a}$ on $[\mathfrak{n},\mathfrak{n}]$. When $0 \in E$ we write $\mathfrak{n}_0 \neq \{0\}$, the corresponding root space and it is clear from (9.3) that

$$[\mathfrak{n}_\alpha, \mathfrak{n}_0] \subset \mathfrak{n}_\alpha \cap [\mathfrak{n},\mathfrak{n}] = 0; \quad \alpha \in A. \qquad (9.11)$$

With this notation it is also clear that if $A' \subset A$ and $\mathfrak{n}_{A'} = \bigoplus_{\alpha \in A'} \mathfrak{n}_\alpha$ (this is taken to be 0 if $A' = \emptyset$) and writing

$$\mathfrak{n}' = \mathfrak{n}_{A'} + \mathfrak{n}_B, \qquad \mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}, \qquad (9.12)$$

then $\mathfrak{n}' \subset \mathfrak{n}$ is a subalgebra and $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a} \subset \mathfrak{g}$ is an SS-subalgebra of $\mathfrak{g}$.

If we set $\mathfrak{n}'' = \mathfrak{n}_A + [\mathfrak{n},\mathfrak{n}]$ then $\mathfrak{n}'' \leftthreetimes \mathfrak{g}$ is an SS-subalgebra with exactly the same roots.

Definition 9.10 The notation is as before. We say that $\mathfrak{g}$ is *bracket reduced* if $\mathfrak{n}_B = [\mathfrak{n},\mathfrak{n}]$ or, equivalently, $\mathfrak{n} = \mathfrak{n}_A + [\mathfrak{n},\mathfrak{n}]$.

When $\mathfrak{g}$ is bracket reduced then $\mathfrak{n}$ is generated by $\mathfrak{n}_A$. More precisely, if we use the notation of §2.1 we have

$$\mathfrak{n}_B = [\mathfrak{n},\mathfrak{n}] = \sum [\mathfrak{n}_A, [\mathfrak{n}_A, [\cdots \mathfrak{n}_A] \cdots]] = \sum_{j \geqslant 2} \mathfrak{n}_A^j, \qquad (9.13)$$

$$B \subset \bigcup_{j \geqslant 2} (A + A + \cdots), \qquad (9.14)$$

where j is the length of summation. To see this write

$$[\mathfrak{n},\mathfrak{n}] = \big[\mathfrak{n}_A + [\mathfrak{n},\mathfrak{n}], \mathfrak{n}_A + [\mathfrak{n},\mathfrak{n}]\big] = [\mathfrak{n}_A,\mathfrak{n}_A] + [\mathfrak{n},[\mathfrak{n},\mathfrak{n}]] = [\mathfrak{n}_A,\mathfrak{n}_A] + \mathfrak{n}^3. \qquad (9.15)$$

Therefore,

$$[\mathfrak{n},[\mathfrak{n},\mathfrak{n}]] = \left[\mathfrak{n}_A + \mathfrak{n}^2, [\mathfrak{n}_A,\mathfrak{n}_A] + \mathfrak{n}^3\right] = \mathfrak{n}_A^3 + \mathfrak{n}^4, \tag{9.16}$$

and so on until the nilpotency kills $\mathfrak{n}^p$ for large enough p. Formula (9.14) follows. From (9.11), (9.13) it follows that for a bracket-reduced algebra for which $0 \in E$ we have

$$\mathfrak{n}_0 \subset \mathfrak{z}(\mathfrak{n}) = \text{the centre of } \mathfrak{n}. \tag{9.17}$$

Two special cases

If $\mathfrak{g}$ is a bracket-reduced SSA and if $B = \emptyset$, then by definition, $\mathfrak{g}$ is of Abelian type.

Let us instead make the assumption $B = \{0\}$. This, by (9.13) and (9.17), implies

$$0 \neq \mathfrak{n}_B = [\mathfrak{n},\mathfrak{n}] = \mathfrak{n}_0 \subset \mathfrak{z}(\mathfrak{n}). \tag{9.18}$$

But (9.13), (9.18) imply

$$[\mathfrak{n},\mathfrak{n}] = \sum [\mathfrak{n}_a, \mathfrak{n}_{-a}], \tag{9.19}$$

where the summation extends through all $a \in E$ such that $0 \neq a$, $-a \in A$ because if $a, b \in E$ are such that $[\mathfrak{n}_a, \mathfrak{n}_b] \neq 0$ then $a + b \in B$ and thus $a + b = 0$. The relation (9.19) implies in particular that $\mathfrak{n}$ contains an SS-subalgebra of Heisenberg type. We can summarise.

Proposition 9.11 *Let $\mathfrak{g}$ be some bracket-reduced SSA and assume that $B = \emptyset$. Then $\mathfrak{g}$ is of Abelian type. If $B = \{0\}$ then $\mathfrak{g}$ contains an SS-subalgebra of Heisenberg type.*

Example Let $\mathfrak{h}$ be as in Definition 9.5 and $\mathfrak{a} \neq 0$ be Abelian. Then, in (9.7), for $\mathfrak{g} = \mathfrak{h} \leftthreetimes \mathfrak{a}$ in (9.9), (9.10) we have $A \neq \emptyset, B = \{0\}$.

The bracket reduction Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ be some SSA and let $E \subset V = \mathfrak{a}^*$ be the set of its roots. We shall consider the class of SS-subalgebras $\mathfrak{g}_1 = \mathfrak{n}_1 \leftthreetimes \mathfrak{a}$ such that $E_1 \subset V$, the set of roots of $\mathfrak{g}_1$ is the same set $E_1 = E$. In other words, we have not 'lost any roots' by passing from $\mathfrak{g}$ to $\mathfrak{g}_1$. We shall consider $\mathfrak{g}_0 = \mathfrak{r} \leftthreetimes \mathfrak{a}$, some minimal element in that class (in the sense that no strict SS-subalgebra of $\mathfrak{g}_0$ has the same set of roots). This SS-subalgebra is of course not uniquely determined.

Proposition 9.12 *If $\mathfrak{g}_0$ is as above, then $\mathfrak{g}_0$ is a bracket-reduced SSA and has the same roots E as $\mathfrak{g}$.*

Indeed, let $\mathfrak{r}_A + \mathfrak{r}_B$ be the decomposition of $\mathfrak{r}$ as in (9.9), (9.10) for a disjoint union $A \cup B = E$. If we assume by contradiction that $\mathfrak{g}_0$ is not bracket reduced then $E \neq \emptyset$ and

$$\mathfrak{r}' = \mathfrak{r}_A + [\mathfrak{r}, \mathfrak{r}] \subsetneqq \mathfrak{r} \tag{9.20}$$

and $\mathfrak{g}' = \mathfrak{r}' \leftthreetimes \mathfrak{a}$ has the same roots E. This contradicts the minimality of $\mathfrak{g}_0$ and proves the proposition.

Remark One conclusion that cannot a priori be deduced from this proposition is that the inclusion (9.14) holds in general (i.e. without making the hypothesis that the algebra is bracket reduced). The reason is that the sets of roots E of $\mathfrak{g}$ and $\mathfrak{g}_0$ are the same but the decompositions $E = A \cup B$ of Definition 9.9 may be different.

9.1.3 Combinatorics

9.1.3.1 Definitions Here V will denote throughout some *non-zero* finite-dimensional real vector space and $E \subset V$ will denote a finite subset. We shall then denote $X = E \backslash \{0\}$.

We enumerate three properties that the set E may or may not possess:

$\mathscr{A}_1$: E spans V: $\operatorname{Vec}(E) = V$;
$\mathscr{A}_2$: the convex hull of X contains 0:

$$0 \in \mathrm{CH}(X) = [\textstyle\sum_{x \in X} \lambda_x x;\ \lambda_x \geqslant 0,\ \sum \lambda_x = 1]; \tag{9.21}$$

$\mathscr{A}_3 = \mathscr{A}_1 \cap \mathscr{A}_2$: that is, both $\mathscr{A}_1$ and $\mathscr{A}_2$ hold for E; it will also be convenient to write $\mathscr{A}_3 = \mathscr{A}$.

If $E \in \mathscr{A}_i$ and $0 \in E$, then $E \backslash \{0\}$ is still an $\mathscr{A}_i$-set. Notice that these properties imply that $X \neq \emptyset$.

Definition 9.13 Let $E \subset V$ be as above. We say that it is a *minimal $\mathscr{A}_i$-set* for some $i = 1,2,3$ if E is an $\mathscr{A}_i$-set and if for every $x \in X$, $E \backslash \{x\}$ is not an $\mathscr{A}_i$-set.

It is clear that E is a minimal $\mathscr{A}_1$-set if $X = E \backslash \{0\}$ is a basis of V.

Definition 9.14 (Simplexes and minimal $\mathscr{A}_2$ sets) We recall the standard definition of a *simplex*. Let $E = (x_0, x_1, \ldots, x_k) \subset V$ be distinct points. We say that E are the *vertices of a simplex* in V and denote the corresponding simplex σ if, for the convex hull of E (: $\mathrm{CH}(E)$),

$$\sigma = [E] = \mathrm{CH}(E) \quad \text{we have } k = \text{ topological dimension of } \sigma. \tag{9.22}$$

We then write

$$\operatorname{Int}\sigma = \left[\textstyle\sum_{j=0}^{k} \lambda_j x_j;\ \lambda_j > 0,\ \sum \lambda_j = 1\right]. \tag{9.23}$$

There should no confusion between $\operatorname{Int}\sigma$ and $\mathring{\sigma}$, the topological interior of σ, which could be $\emptyset$. When $\operatorname{Int}\sigma = \mathring{\sigma}$, we say that the simplex is *non-singular*. When $k = 0$ we say that the simplex is *degenerate*. The following property of simplexes is well known.

Let $E \subset V$ and let $x \in \mathrm{CH}(E)$. Then there exist $Y = (y_1, y_2, \ldots, y_k) \subset E$ that are vertices of a simplex $\sigma = \mathrm{CH}(Y)$ and such that $x \in \operatorname{Int}(\sigma)$.

Exercise 9.15 (Carathéodory's theorem; see Grünbaum, 1967) Prove this by induction on $\dim V = n \geqslant 1$. Indeed, we may assume that $P = \mathrm{CH}(E)$ has a non-empty interior and $x \in \mathring{P} \subset V$, for otherwise, E lives in an affine subspace of lower dimension and we use the induction there. Let $x \neq x_0 \in E$ be arbitrary. Then the line segment that joins x_0 to x, once extended, cuts ∂P at some point $x' \in \mathrm{CH}(E')$ for some $E' \subset E$, and E' lies in some affine subspace of V of lower dimension. The inductive hypothesis applies therefore to x' and E' and we are done. Here we have assumed that E is finite which is good enough for us.

From the above it clearly follows that if E is a minimal $\mathscr{A}_2$-set then $X = E\backslash\{0\}$ are the vertices of a non-degenerate simplex and $0 \in \operatorname{Int}[X]$.

Proposition 9.16 *Let E be some minimal $\mathscr{A} = \mathscr{A}_3$-set. Then we can decompose $X = X_1 \cup X_2$ into two disjoint subsets and we can decompose $V = V_1 \oplus V_2$ into a direct sum of subspaces in such a way that the following hold*:

(a) *$X_i = X \cap V_i$ and V_i is spanned by X_i, $i = 1, 2$;*
(b) *$X_1 \neq \emptyset$ and X_1 form the vertices of a non-degenerate simplex $[X_1]$ and $0 \in \operatorname{Int}[X_1]$;*
(c) *when $V_2 \neq \{0\}$, X_2 is a basis of V_2 (but it could be that $V_2 = \{0\}$ and then $X_2 = \emptyset$).*

To see this we express $0 = \sum_{j=1}^{k} \lambda_j x_j$ as a convex combination with $X_1 = (x_1, \ldots, x_k) \subset X$ and k as small as possible. By the minimality of k and Exercise 9.15 it follows that X_1 form the vertices of a non-degenerate simplex and $0 \in \operatorname{Int}[X_1]$.

Let $X_2 = X\backslash X_1$ and $V_i = \mathrm{Vec}(X_i)$ with the convention that $V_2 = \{0\}$ if $X_2 = \emptyset$. Then we must have $V_1 \cap V_2 = \{0\}$. Otherwise, assume for contradiction that $0 \neq z \in V_1 \cap V_2$. Now z is a linear combination of points of X_1 and a non-trivial linear combination of points of X_2. This implies that an element $u \in X_2$ can be 'knocked out' and $X_1 \cup (X_2\backslash\{u\})$ still spans V. This contradicts the $\mathscr{A}$-minimality of X. This proves that $V = V_1 \oplus V_2$. But then if $V_2 \neq \{0\}$ it follows that X_2 is a basis of V_2 because of the $\mathscr{A}$-minimality of X. This completes the proof of the proposition.

9.1.4 $\mathscr{A}$-couples

Definition 9.17 Let $A \subset E \subset V$ be two finite subsets where A is an $\mathscr{A}_i$-subset for some $i = 1,2,3$ and $V \neq \{0\}$ as before, and where

$$E \subset \left[a_1 + \cdots + a_j;\ j \geqslant 1,\ a_k \in A,\ 1 \leqslant k \leqslant j\right]. \tag{9.24}$$

We then say that $A \subset E$ is an *$\mathscr{A}_i$-couple*. We shall say that $A \subset E$ is a *minimal $\mathscr{A}_i$-couple* if for all $0 \neq a \in A$ the set $E \backslash \{a\}$ is not an $\mathscr{A}_i$-set.

Comments For arbitrary finite sets $A \subset E \subset V$, condition (9.24) implies that $A \in \mathscr{A}_i$ if and only if $E \in \mathscr{A}_i$. Notice, on the other hand, that (9.24) could hold for two sets $A \subset E$ but not for $A \subset (E \cup \{0\})$. But more to the point, the motivation example of $\mathscr{A}_i$-couples is supplied by E, the set of roots of a bracket-reduced SSA where A is as in §9.1.2, and condition (9.24), being none other than (9.14). This makes the link with bracket-reduced algebras. For instance, a bracket-reduced algebra is SSCA if and only if the corresponding $A \subset E$ in §9.1.2 is an $\mathscr{A}_2$-couple. It is clear that when $A \subset E$ is a minimal $\mathscr{A}_i$-couple then A is a minimal $\mathscr{A}_i$-set. In what follows we shall write $B = E \backslash A$. It is then obvious that when $A \subset E$ is a minimal $\mathscr{A}_1$ couple then B is either $\emptyset$ or $\{0\}$. We also have, however, the following proposition.

Proposition 9.18 *If $A \subset E \subset V$ is a minimal $\mathscr{A}_i$-couple for some $i = 1,2,3$ then B is either $\emptyset$ or $\{0\}$.*

The $\mathscr{A}_2$ case We can write $0 = \sum_{a \in A \backslash \{0\}} \lambda(a) a$ where $\lambda(a) > 0$ by the minimality condition. Use (9.24) and assume for contradiction that

$$0 \neq b = \sum_{a \in A \backslash \{0\}} \mu(a) a \in B; \quad \mu(a) = 0,1,2,\ldots \text{ not all } 0. \tag{9.25}$$

Both λ and μ are considered as non-negative functions on A and we can define

$$\eta_0 = \sup\left[\eta > 0;\ \eta \mu(\cdot) \leqslant \lambda(\cdot)\right]. \tag{9.26}$$

There exists then $0 \neq a_0 \in A$ such that $\eta_0 \mu(a_0) = \lambda(a_0)$. We have then

$$0 = \eta_0 b + \sum_{a \in A \backslash \{0\}} \big(\lambda(a) - \eta_0 \mu(a)\big) a. \tag{9.27}$$

But $b \neq a_0$ because $b \notin A$ and (9.27) implies that $E \backslash \{a_0\}$ is an $\mathscr{A}_2$-set and contradicts the minimality.

Aside What we have used here is the exchange principle while keeping track of the signs.

The $\mathscr{A}_3$ case By our condition, A is a minimal $\mathscr{A}_3$-set and with the notation of Proposition 9.16 we can write $V = V_1 \oplus V_2$ and, for $i = 1,2$, $A_i = V_i \cap A$, $X_i = A_i \backslash \{0\}$, giving $V_i = \text{Vec}(A_i), 0 \in \text{Int}(X_1)$.

We shall give the proof of Proposition 9.18 by contradiction and distinguish two separate cases:

(i) assume there exists $B \ni b \neq 0$ with $b \in V_1$;
(ii) assume there exists $B \ni b \neq 0$ with $b \notin V_1$.

By our definition (9.24), every $b \in B$ can be written $b = b_1 + b_2$ where, when $b_i \neq 0$, we have $b_i =$ [a sum of elements of X_i], with $i = 1,2$. If we use the canonical projection $V_1 \oplus V_2 = V \to V_2$ we see from Proposition 9.16 that in that decomposition, if $b \in V_1$ then $b_2 = 0$. We can therefore write

$$b = b_1 = \sum_{a \in X_1} \mu(a)a, \quad \mu(a) = 0,1,2,\ldots.$$

The argument that we used in the $\mathscr{A}_2$ case can be repeated because in Proposition 9.16(b), $0 \in \text{Int}X_1$ and we can find some $a_0 \in X_1$ such that 0 lies in the barycentre of the set $Y = (X_1 \backslash \{a_0\}) \cup \{b\}$ (: $0 \in \text{CH}(Y)$: see (9.27)). For the point a_0 we also have $\mu(a_0) \neq 0$ and therefore Y can be used to span $V_1 = \text{Vec}(X_1) = \text{Vec}(Y)$. In this exchange that we have made between a_0 and b we have not affected the set X_2 at all. The conclusion is that the set $Y \cup X_2$ has the $\mathscr{A}_3$-property. Since $E \backslash \{a_0\} \supset Y \cup X_2$, because $b \notin A$ and therefore $b \neq a_0$, this *contradicts* the minimality condition of the couple $A \subset E$.

In case (ii), if $b \notin V_1$ we write again the decomposition

$$b = b_1 + b_2 = \sum_{a \in X_1} \phi(a)a + \sum_{a \in X_2} \phi(a)a, \tag{9.28}$$

and there exists $a_2 \in X_2$ such that $\phi(a_2) \neq 0$. We shall consider the projection $V \ni u \to u(\text{mod} V_1) \in V_2$. With this projection we see from (9.28) that the set $(X_2 \backslash \{a_2\})(\text{mod} V_1)$, together with $b(\text{mod} V_1)$, spans V_2. On the other hand, since X_1 spans V_1, it follows that the set $(A \backslash \{a_2\}) \cup \{b\}$ spans V. This set is contained in $E \backslash \{a_2\}$ because $b \notin A$, and since $0 \in \text{CH}(X_1) \subset \text{CH}(E \backslash \{a_2\})$, the required contradiction with the minimality of the couple $A \subset E$ follows.

9.1.5 The $\mathscr{A}$ algebras

We recall that $\mathscr{A}_3$ is simply denoted by $\mathscr{A}$. Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ be some SSA and $E \subset V = \mathfrak{a}^*$ is the set of roots as in §9.1.1. We shall assume that $\mathfrak{g}$ is bracket reduced and $A \cup B = E$ are as in Definition 9.9 and $\mathfrak{n} = \mathfrak{n}_A + \mathfrak{n}_B$. We shall further assume that $E \in \mathscr{A}$ in V. To wit, the conditions imposed on $\mathfrak{g}$ are the following.

Definition 9.19 If $\mathfrak{g}$ is a bracket-reduced SSA and $E = A \cup B$ is an $\mathscr{A}$-set of $V = \mathfrak{a}^*$, then we call $\mathfrak{g}$ a *bracket-reduced* $\mathscr{A}$*-algebra* (BR$\mathscr{A}$A). See the comments in §9.1.4.

Definition 9.20 We say that $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ is a *model* if it is a BR$\mathscr{A}$A and if $A \subset E$ is a minimal $\mathscr{A}$-couple.

Definition 9.21 Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$, $\mathfrak{n} = \mathfrak{n}_A + \mathfrak{n}_B$, $A \cup B = E$ (with the notation of (9.10) be some BR$\mathscr{A}$A and assume that it has the following minimality property: if $\mathfrak{g}_1 = \mathfrak{n}_1 \leftthreetimes \mathfrak{a}$ is an SS-subalgebra and if $\mathfrak{g}_1 \neq \mathfrak{g}$ then $\mathfrak{g}_1$ is *not* a BR$\mathscr{A}$A. We then say that $\mathfrak{g}$ is a *minimal* BR$\mathscr{A}$A.

This definition is formulated in such a way that every $\mathfrak{g}$ BR$\mathscr{A}$A contains some SS-subalgebra that is a minimal BR$\mathscr{A}$A. Indeed, in the class of SS-subalgebras of $\mathfrak{g}$ that are BR$\mathscr{A}$A-algebras we pick one up of minimal dimension.

Proposition 9.22 *Every minimal* BR$\mathscr{A}$A *algebra is a model.*

Proof Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ with $\mathfrak{n} = \mathfrak{n}_A + \mathfrak{n}_B$, $A \cup B = E$ (as in Definition 9.9) be some minimal BR$\mathscr{A}$A. And let us suppose for contradiction that $A \subset E$ is not a minimal $\mathscr{A}$-couple.

There exists then $0 \neq \alpha_0 \in A$ such that $E' = E \backslash \{\alpha_0\}$ is an $\mathscr{A}$-set and we can define

$$\mathfrak{n}' = \sum_{\alpha \in A'} \mathfrak{n}_\alpha + \sum_{\alpha \in B} \mathfrak{n}_\alpha; \quad A' = A \backslash \{\alpha_0\}. \tag{9.29}$$

If we define $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}$, by (9.12) we have an SS-subalgebra which, considered as an SSA in its own right, has roots E' and this is an $\mathscr{A}$-set. We now apply Proposition 9.12 and construct $\mathfrak{g}'_0 \subset \mathfrak{g}$, some bracket-reduced subalgebra with the same roots E'. This means that $\mathfrak{g}'_0$ is a BR$\mathscr{A}$A and since $\mathfrak{g}'_0 \subsetneqq \mathfrak{g}$ we have the required contradiction. □

9.1.6 Proof of the algebraic structure theorem (Theorem 9.7)

Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ be some SSCA as in Definition 9.6. The first step is that we can use this definition to write $\mathfrak{g} = (\mathfrak{n} \leftthreetimes \mathfrak{a}_1) \oplus \mathfrak{a}_2$, where the action of $\mathfrak{a}_2$ is trivial and $\mathfrak{g}_1 = \mathfrak{n} \leftthreetimes \mathfrak{a}_1$ is irreducible in the sense of §9.1.1. For the proof of Theorem 9.7 we may therefore assume that $\mathfrak{g}$ is an irreducible C-algebra. Furthermore, we can apply Proposition 9.12 and assume that $\mathfrak{g}$ is bracket reduced.

These two conditions are equivalent to saying that $\mathfrak{g}$ is an SS$\mathscr{A}$A-algebra by the definition of the $\mathscr{A} = \mathscr{A}_3$ condition (see the comments in §9.1.4). But then Proposition 9.22 applies and it follows that $\mathfrak{g}$ contains a model $\mathfrak{m} = \mathfrak{n}' \leftthreetimes \mathfrak{a}$ with $\mathfrak{n}' = \mathfrak{n}_A + \mathfrak{n}_B$ as in (9.12). By Proposition 9.18 there are exactly two possibilities:

(a) $B = \emptyset$: then by Proposition 9.11, $\mathfrak{m}$ is an SSCA of Abelian type;

(b) $B = \{0\}$: then, by Proposition 9.11, $\mathfrak{m}$ contains some SSA of Heisenberg type.

This completes the proof of the theorem.

9.1.7 The two alternatives

In this final subsection we shall recapitulate what we have done and express it in concrete terms without the use of the additional technical terminology that we used in the proofs.

We shall consider $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{a}$ some SSA that is also a C-algebra. This implies that $\mathfrak{a} \neq \{0\}$. One of the following two alternatives is realised.

The Abelian alternative There exists $\mathfrak{n}' \subset \mathfrak{n}$ a subalgebra that is $\mathfrak{a}$ stable; that is, $[\mathfrak{n}', \mathfrak{a}] \subset \mathfrak{n}'$ such that

(i) $\mathfrak{n}'$ is Abelian;

(ii) the roots $L_1, \dots, L_p \in V = \mathfrak{a}^*$ of $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}$ are non-zero and are vertices of a simplex σ and $0 \in \operatorname{Int} \sigma$ and $p \geqslant 2$;

(iii) the root spaces $\mathfrak{n}_{L_j} = \mathfrak{n}_j \subset \mathfrak{n}'$, with $j = 1, 2, \dots, p$, are all one-dimensional and $\mathfrak{n}' = \mathfrak{n}_1 \oplus \cdots \oplus \mathfrak{n}_p$.

Example 9.23 When $p = 2$ this is the C-algebra considered in §2.3.2(iii) and Example 7.16.

The Heisenberg alternative There exist $0 \neq L \in V = \mathfrak{a}^*$ some root of $\mathfrak{g}$ and two non-zero vectors $v_+, v_- \in \mathfrak{n}$ such that

$$(\operatorname{ad} \xi) v_\pm = (\pm L\xi) v_\pm; \quad \xi \in \mathfrak{a} \tag{9.30}$$

and such that $\zeta = [v_+, v_-] \neq 0$, $[v_\pm, \zeta] = 0$. It also follows of course from the above that $(\operatorname{ad} \xi)\zeta = 0$, with $\xi \in \mathfrak{a}$. Furthermore, with $\mathfrak{n}' = (v_+, v_-, \zeta)$, $\mathfrak{g}' = \mathfrak{n}' \leftthreetimes \mathfrak{a}$ is a subalgebra of $\mathfrak{g}$ because $[\mathfrak{n}', \mathfrak{a}] \subset \mathfrak{n}'$.

The general overview We shall summarise here the notation that will be used throughout and which will hold for both the above alternatives. We shall set

$$\begin{aligned} \mathfrak{a}_0 &= \bigcap_{1\leqslant j\leqslant p} \operatorname{Ker} L_j \quad \text{in the Abelian case,} \\ \mathfrak{a}_0 &= \operatorname{Ker} L \qquad\qquad \text{in the Heisenberg case.} \end{aligned} \tag{9.31}$$

With this definition, $\mathfrak{a}_0$ acts trivially on $\mathfrak{n}'$ in both cases. If we denote by $\mathfrak{a}'$ some complement of $\mathfrak{a}_0$ in $\mathfrak{a}$, the formulas that should be retained and which apply to both cases are (see (9.5))

$$\begin{aligned} \mathfrak{g} &= \mathfrak{n} \leftthreetimes (\mathfrak{a}' \oplus \mathfrak{a}_0)\,; \quad \mathfrak{n}' \subset \mathfrak{n}, \\ \mathfrak{g}' &= \mathfrak{n}' \leftthreetimes (\mathfrak{a}' \oplus \mathfrak{a}_0) = (\mathfrak{n}' \leftthreetimes \mathfrak{a}') \oplus \mathfrak{a}_0. \end{aligned} \tag{9.32}$$

Furthermore, in the Abelian case, the roots $L_1, \dots, L_p$ can be identified to the vertices of a non-singular simplex $\sigma \subset (\mathfrak{a}')^*$, the dual space of $\mathfrak{a}'$, and $0 \in \operatorname{Int}\sigma = \mathring{\sigma}$. This means, in particular, that $\dim \mathfrak{a}' = p-1$.

In the Heisenberg case, $\dim \mathfrak{a}' = 1$ and the roots are $0, L, -L \in (\mathfrak{a}')^*$, with $L \neq 0$.

Remark 9.24 In both cases $\mathfrak{a}_0$ splits off trivially as a direct factor in $\mathfrak{g}'$ but *not* in $\mathfrak{g}$; and $\mathfrak{a}_0$ in general acts non-trivially on $\mathfrak{n}$. This fact is a source of serious complications in the geometry of the group in §9.5.

9.2 Geometric Constructions on Special Soluble Groups: Examples

In this section I shall codify notation and give the proof of Theorem 7.11 for some examples that can be found in Gromov (1991).[1]

9.2.1 Notation

Throughout, we shall use the terminology SSG for special soluble groups for simply connected groups whose Lie algebra is an SSA (see §9.1.1). Right through we shall resort to the standard practice that consists in *tacitly* identifying the Lie algebra and the corresponding simply connected group when the group is Abelian. (For example, the additive group $G = \mathbb{R}$ is identified with the vector space $\mathbb{R}$ which is the Lie algebra $\mathfrak{g}$. Strictly speaking of course, the

[1] In that reference the author refers back to Epstein et al. (1992) so, at least to me, the precise individual who initiated these examples, and the corresponding proofs, is not clear.

exp mapping $x \mapsto e^x$ sends $\mathfrak{g}$ to the multiplicative group $\mathbb{R}_+^*$. But for us here, $\mathbb{R}_+^* \simeq \mathbb{R}$.)

We shall denote the group that corresponds to $\mathfrak{n}' \leftthreetimes \mathfrak{a}'$ of (9.32) in the case of the Abelian alternative by $G_r = N \leftthreetimes A = \mathbb{R}^r \leftthreetimes \mathbb{R}^{r-1}$, for $r = 2, 3, \ldots$. For the Abelian group N we shall identify the group with the Lie algebra $\mathfrak{n}'$ and the action of A on $\mathfrak{n}'$ is given by the diagonal matrix

$$\operatorname{Ad} u = \operatorname{Diag}\left(\mathrm{e}^{L_1 u}, \ldots, \mathrm{e}^{L_r u}\right); \quad u \in A, \tag{9.33}$$

where A and $\mathfrak{a}$ are identified and where $L_1, \ldots, L_r$ are vectors in $\mathbb{R}^{r-1}$ which is identified to the dual space of $\mathfrak{a}'$ (see (9.32)). These vectors are the vertices of a non-singular simplex $\sigma^* = [L_1, \ldots, L_r]$ with $0 \in \operatorname{Int} \sigma^* = \mathring{\sigma}^*$.

With the above notation we shall write $x = (x_1, \ldots, x_r) \in \mathbb{R}^r = N$ and $(u_1, \ldots, u_{r-1}) \in \mathbb{R}^{r-1} = A$ for the Euclidean coordinates so that we can give on G the coordinates $(x_1, \ldots, x_r, u_1, \ldots, u_{r-1})$ (cf. §7.3.1).

The important new object that we shall consider is the cube

$$\square_R^r = \left(x = \left(x_1, \ldots, x_r\right);\, 0 \leqslant x_j \leqslant R,\ j = 1, \ldots, r\right) \subset N, \quad R > 1. \tag{9.34}$$

The size R of this cube throughout will be a free parameter that will vary and will be made to tend to ∞.

9.2.2 The special case G_2

Here the coordinates on $G = G_2$ are $(x_1, x_2, u) \in \mathbb{R}^3$ that is identified with $\mathbb{R}^2 \leftthreetimes \mathbb{R}$ and

$$\operatorname{Ad} u = \begin{pmatrix} e^{\alpha u} & 0 \\ 0 & e^{-\beta u} \end{pmatrix}; \quad u \in \mathbb{R}, \text{ where } \alpha, \beta > 0. \tag{9.35}$$

This is the group that we considered in §2.3.2 and more explicitly in Exercise 7.17.

The vertices of $\square_R^2$ are then identified to the four points $(\varepsilon_1 R, \varepsilon_2 R, 0) \in G$, with $\varepsilon_i = 0, 1$, which we shall denote by a, b, c, d. For some $C > 0$, to be specified later, we shall write $U = (0, 0, C \log R)$ and $a_\pm = a \pm U, \ldots, d_\pm = d \pm U$.

We shall consider the four paths in Figure 9.1, consisting of three line segments each, indicated below by their successive vertices:

$$\gamma_{ab} = a a_+ b_+ b, \quad \gamma_{bc} = b b_- c_- c, \quad \gamma_{cd} = c c_+ d_+ d, \quad \gamma_{da} = d d_- a_- a. \tag{9.36}$$

These are polygonal loops: γ_{ab} joins a to b, and similarly for the other loops. The important property that we shall need is that the total length of these paths for the Riemannian structure on G is $\approx \log R$.

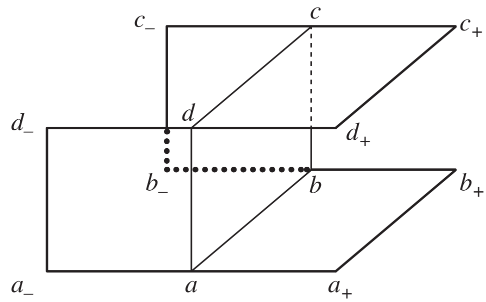


Figure 9.1 A piecewise linear path of total length $\approx \log R$ that goes through the vertices a, b, c, d of $\square_R^2$.

It is clear that to see this, it suffices to verify that the four straight-line segments

$$(a_+b_+), \qquad (c_+d_+), \qquad (b_-c_-), \qquad (d_-a_-) \tag{9.37}$$

have Riemannian length $\lesssim \log R$ because, by the choice of U, the other components of the paths clearly have lengths $\approx \log R$.

For the first two segments in (9.37), by §8.3, it follows that their Riemannian length is just their Euclidean length R multiplied by $e^{-C\log R}$. That length tends therefore to 0 as $R \to \infty$ as long as the constant $C > 0$ has been chosen large enough. If we switch the coordinates x_1 to x_2 and α to $-\beta$, we see that the same thing holds for the other two segments in (9.37). This proves our assertion.

The polynomial arcs $\gamma_{ab}, \dots$ can be smoothed out by rounding out the corners to the C^∞ arcs $\tilde{\gamma}_{ab}, \dots$ as shown in Figure 9.2. Thus we obtain S which is a C^∞ embedded circle (: $S^1 \subset \mathbb{R}^3$) which, apart from having length $\approx \log R$, also has the following property.

Transversality property

$$S \cap [|u| \leqslant 1] = [x_1 = \varepsilon_1 R,\ x_2 = \varepsilon_2 R,\ |u| \leqslant 1,\ \varepsilon_i = 0, 1]. \tag{9.38}$$

These are four segments in the u coordinate that stick out from the vertices of $\square_R^2$.

9.2.3 A first application of the construction

We shall go back to the differential form $\omega = \mathrm{d}x_1 \wedge \mathrm{d}x_2$ of Example 7.16 and assume here that $\alpha = \beta = 1$ in (9.35). From our construction it is now easy to

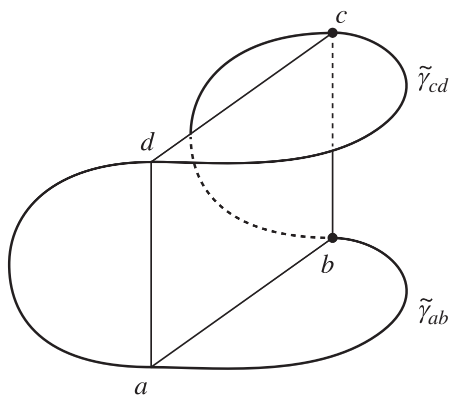


Figure 9.2 The smoothed-out path of Figure 9.1 which is an embedded circle that lives in $G = \mathbb{R}^3$.

see that it is not possible to solve $d\theta = \omega$ with a polynomially growing form θ on $\mathbb{R}^3$; here θ is a form of degree 1 (i.e. $\theta \in T^*\mathbb{R}^3$).

To see this we use Stokes' theorem on the surface $\Sigma \subset \mathbb{R}^3$ that is the union of the five rectangles $abcd$ and $aa_+b_+b, bb_-c_-c, cc_+d_+d, dd_-a_-a$. The boundary of that surface is $S_0 = \gamma_{ab} \cup \gamma_{bc} \cup \gamma_{cd} \cup \gamma_{da}$. The contradiction under the assumption that θ is polynomial is then immediate by $\int_\Sigma \omega = 4R^2$ because we are here only integrating in $\square_R^2$. On the other hand, $|\int_{\partial\Sigma} \theta| \leqslant c\,(\log R)^c$ because $\partial\Sigma = S_0$ lies in the ball of radius $\lesssim \log R$ and we can use the bounds on $|\theta|$ and also the length of $S_0 \leqslant C\log R$. Stokes' theorem on the other hand gives $\int_\Sigma \omega = \int_{\partial\Sigma} \theta$, provided of course that the correct orientations have been chosen in Σ and $\partial\Sigma$.

The above argument can, without too much trouble, be extended to the general case $\alpha, \beta > 0$ in (9.35) and to the differential forms that we defined in Exercise 7.17 but we shall not pursue the matter further now (see §9.3.5 below).

9.2.4 The use of transversality and the filling property

The C-condition (i.e. that the two roots $L_1 = \alpha, L_2 = -\beta$ are of opposite signs) is essential for transversality because it is this that allows us to switch the coordinates x_1 to x_2 and vice versa without backtracking in the u-coordinate.

For the requirement that the total length of S is $\leqslant C\log R$ we could have made the four satellite rectangles $aa_+b_+b, \ldots$ all on the same side by backtracking in u and the argument in §9.2.2 about the lengths in (9.37) works just the same.

For this the C-condition is not needed. For the application that we gave in §9.2.3 the C-condition $\alpha = +1$, $\beta = -1$ is, on the other hand, used to prove that ω is a bounded form (see Exercise 7.17).

Now, however, in our next exploitation of the construction of §9.2.2 we shall make essential use of this transversality.

We shall show that the $\mathscr{F}_2$-property of Definition 7.12 fails in G_2. More precisely, we shall assume that there exists some Lipschitz map $\widehat{\Phi}$ defined on the unit disc $D = [x \in \mathbb{R}^2;\ |x| \leqslant 1]$ such that

$$\widehat{\Phi}\colon D \longrightarrow G; \quad \widehat{\Phi}(\partial D) = S \tag{9.39}$$

and that there exist constants such that

$$\operatorname{Lip}\widehat{\Phi} \leqslant C(\log R)^C; \quad R \geqslant 1. \tag{9.40}$$

We shall show that a contradiction can be obtained from this.

This contradiction will here be obtained under the additional assumption on $\widehat{\Phi}$ that it is possible to perturb $\widehat{\Phi}$ and modify it slightly to a new $\widetilde{\Phi} \in C^\infty$ that is not only smooth, but also gives an embedding of D to a smooth manifold with boundary $\widetilde{\Phi}(D) = \widetilde{D} \subset G$, and for which (9.38), (9.39) and (9.40) also hold.

Let $\pi\colon N \leftthreetimes A \to A$ denote the canonical projection. This projects (x_1, x_2, u) to the third coordinate. We shall need to use the following fundamental fact from differential topology.

For almost all $[u \in A;\ |u| < 1]$, $l_u = \pi^{-1}(u)$ is a 'neat' submanifold of $\widetilde{D}$ and its boundary consists of two points among (a,u), (b,u), (c,u), (d,u). Here, neat submanifold means a C^∞ curve that joins two different points among the above four points. The values of $u \in A$ for which this holds are the regular points, and the 'almost all' comes from Sard's theorem in differential topology (see de Rham, 1960; Hirsch, 1976).

The possibility of regularising and obtaining $\widetilde{D}$ relies on the Whitney embedding theorem (see Hirsch, 1976, Chapter 2). We shall come back to that in greater generality in §9.3.4. (The well-informed reader must have noticed that there are problems with the use of Whitney's theorem simply because the dimensions do *not* add up! Here $\dim G_2 = 3$ and $\dim \widetilde{D} = 2$. We shall explain how to get round that difficulty in §9.3.4.)

The endgame From the above it follows that for every regular point the Euclidean length of l_u, that is, the length for the Riemannian structure of $\mathbb{R}^3$, is $\geqslant 2R$. We shall now use the classical coarea formula in the Euclidean space $\mathbb{R}^3$. This implies that

$$\text{Euclidean surface area of } \widetilde{D} \cap [|u| \leqslant 1] \geqslant 2R \tag{9.41}$$

because the set of regular points is of full measure. The direct verification of (9.41) is in fact elementary.

On the other hand, by (8.43), the two Riemannian metrics on $[\,|u|<1] \subset G$, namely the Euclidean metric of $\mathbb{R}^3$ and the one coming from the group structure, are comparable within fixed constants. In fact the Euclidean Riemannian structure comes from the trivial replica G_T (see Exercises 8.5 and 8.6). The conclusion is that the Riemannian two-dimensional volume

$$\mathrm{Vol}_2\left(\widetilde{D}\cap |u|<1\right) \geqslant cR. \tag{9.42}$$

This contradicts the fact that $\tilde{\Phi} \in \mathrm{Lip}\,(c(\log R)^c)$. We have proved therefore that the $\mathscr{F}_2$-property fails in G_2.

Remarks Several remarks and references can be given on the above argument. But we shall defer these until we treat the general case in Chapter 10. It is worth pointing out, however, that we have not used the full thrust of the $\mathscr{F}_2$-property to obtain the above contradiction. In fact we have used only the original conditions (cf. (1.2))

$$\mathrm{Vol}_2\left(\widehat{\Phi}(D)\right) \leqslant C\big(1+\mathrm{Vol}_1\,\widehat{\Phi}(\partial D)\big)^C. \tag{9.43}$$

9.2.5 Generalisations and the Heisenberg alternative

The example that we examined in §9.2.2 has two special features. First it is a $G_r = \mathbb{R}^r \leftthreetimes \mathbb{R}^{r-1}$ group as defined in §9.2.1. We shall generalise this construction to all SSG that are C-groups of the form G_r. Some indication of how this is done will be given, however, at the end of this section in §9.2.7 but the general construction will be deferred to §§9.3 and 9.4.

The other special feature of G_2 in §9.2.2 is that it is of 'rank 1'. This means that it is of the form $N \leftthreetimes A$ where $A = \mathbb{R}$ is one-dimensional and N is some nilpotent group. It is this feature that we shall maintain here and consider the Heisenberg alternative $\mathfrak{n}' \leftthreetimes \mathfrak{a}'$ in (9.30) and the corresponding group $K = H \leftthreetimes A = H \leftthreetimes \mathbb{R}$ where H is the Heisenberg group that is simply connected and corresponds to the Heisenberg algebra in (9.6).

To find the correct generalisation of $\square_R^2 \subset \mathbb{R}^2$ from §9.2.3, we observe that if $X = (R,0), Y = (0,R) \in \mathbb{R}^2$ then $0, X, X+Y, X+Y-X, X+Y-X-Y = 0$ are the four vertices of $\square_R^2 \subset \mathbb{R}^2$ where we use the group product to multiply (i.e. add since $\mathbb{R}^2$ is Abelian) in the group $\mathbb{R}^2$. In multiplicative notation the vertices of $\square_R^2$ are therefore the successive segments of the word $[X,Y] = XYX^{-1}Y^{-1}$ that represents the identity.

For the Heisenberg group H we shall use the notation of (9.6) and set $X =$

$e^{R\nu}$, $Y = e^{R\mu}$, with $R > 1$, for $\nu, \mu \in \mathfrak{h}$ in the Lie algebra. What replaces the above commutator is then the following representation of the identity:

$$XY^2XY^{-1}X^{-2}Y^{-1} = e. \tag{9.44}$$

To see (9.44) we simply use the BCH formula (see Varadarajan, 1974, §2.15) to verify that $[X,Y] = YXY^{-1}X^{-1} \in$ centre of H and write the left-hand side as $XYX^{-1}Y^{-1}YXY^{-1}X^{-1}$. The successive segments of the word (9.44) (after cyclic permutation) are then denoted by the points a, b, c, d, e, f in H; they are drawn in Figure 9.3. These points are distinct because they clearly project to distinct points in $H/(\text{the centre})$.

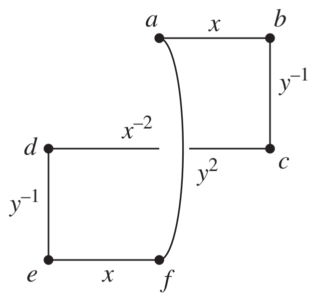


Figure 9.3 The successive components of the word (9.44) in H.

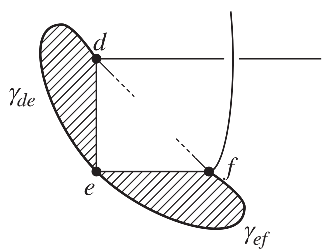


Figure 9.4 For the arcs $\gamma_{ef}, \gamma_{de}, \ldots$ we spill out of H in the positive or negative $A^{\pm} = (h \in H,\ u \in A;\ \pm u > 0)$ as the case might be.

The aim now is to find a new $S = S_K \subset K$ that is an embedded C^∞ circle S^1 that has Riemannian length $\mathrm{Vol}_1(S) \leqslant C \log R$ and passes through the six vertices. We shall denote by $\gamma_{ef}, \ldots$ the arc of S between e and f, and so on. Then $S = \gamma_{ef} \cup \cdots$, as in Figure 9.4. The transversality condition now says that with the coordinates of K,

$$(h,u); \quad h \in H,\ u \in A = \mathbb{R}, \tag{9.45}$$

we have the following result.

Theorem 9.25 (Transversality property) *The set* $S \cap [|u| \leqslant 1]$ *is obtained by multiplying the segment* $[-1,1] \subset A$ *by* a, b, c, d, e, f *on the left in* K.

The best way to verify that we can guarantee these properties is to use for the arcs $\gamma_{ab}, \ldots$ the same parametrisation as in §8.2.2. For example, γ_{ef} is parametrised by

$$\exp(\alpha(s)R\nu)\exp(C_0\beta(s)\log R\xi), \quad 0 < s < 1, \tag{9.46}$$

where ξ is the generator of $\mathfrak{a}$ the Lie algebra of A and where $0 \leqslant \alpha, \beta \in C^\infty$, $\beta(0) = \beta(1) = 0$ and these functions satisfy (8.24). The fact that the length of γ_{ef} is $O(\log R)$ then follows if C_0 in (9.46) is large enough as in (8.25), and the transversality property (Theorem 9.25) is a consequence of the definition of α, β in (8.24).

The exploitation that we made of S in §9.2.4 can now be repeated verbatim for $S_K \subset K$. We construct $\widehat{\Phi}\colon D \to K$ that has properties (9.39), (9.40). We then regularise as in §9.2.4 and obtain $\widetilde{\Phi}$ to be close to $\widehat{\Phi}$ and for which $\widetilde{D} = \widetilde{\Phi}(D)$ is transversal to the canonical projection $\pi\colon K \to A$ for almost all $|u| \leqslant 1$ for coordinates that are as (9.45): see de Rham (1960), Hirsch (1976). The endgame in §9.2.4 is now modified as follows.

9.2.6 The endgame in the Heisenberg case

As in §9.2.4, we write $l_u = \pi^{-1}(u)$ for every regular point of π $u \in A$, for $|u| < 1$. The set l_u is a one-dimensional submanifold in the manifold K which is a C^∞ manifold $\simeq H \times \mathbb{R} \simeq \mathbb{R}^4$. Furthermore, l_u has as a boundary two distinct points (a_1, u), (a_2, u), with $a_i = a, b, \ldots, f$ as in Figure 9.3. We can again assign on K with the product Riemannian structure of the group Riemannian structure on H with $\mathbb{R}$. This again is the trivial replica of Exercise 8.5 and will be denoted by K_T. On $|u| < 1$, the group Riemannian structure K and the product structure K_T are comparable and lie within fixed constants. This fact was observed previously in §9.2.4. We conclude from the above observations the following two points.

Remark 9.26 (Coarea) The surface areas of $\widetilde{D} \cap (|u| < 1)$ measured for the two Riemannian structures K and K_T are comparable within a constant independent of R:

$$\text{area in } K \approx \text{area in } K_T. \tag{9.47}$$

We shall now use the fact that by the polynomial distortion of Example 7.5

and §2.14 on the nilpotent group H, the mutual distance between the points a, b, c, d, e, f is at least CR^C for some constant $C > 0$ and $R \geqslant 1$. This implies that the length of l_u in the Riemannian metric K_T is at least CR^C. To see this mutual distance we can also project to $H/[H,H]$, as we did before.

The coarea formula applies again in the product $H \times \mathbb{R}$ and it implies that the right-hand side, and therefore also the left-hand side, of (9.47) is larger than CR^C. We finally have the exact analogue of (9.42):

$$\mathrm{Vol}_2\left(\widetilde{D}\right) \geqslant CR^C; \quad R \geqslant 1 \tag{9.48}$$

for appropriate constants. The above argument that uses the coarea formula will be elaborated further in Chapter 10.

One conclusion of (9.48) is that we cannot have $\widetilde{\Phi} \in \mathrm{Lip}\left(C(\log R)^C\right)$ for any constants and the final result is the following proposition.

Proposition 9.27 *The group K in the Heisenberg alternative does not have the $\mathscr{F}_2$-property.*

Remark 9.28 The above argument can be extended without any difficulty to deal with any group of rank 1, that is, any soluble simply connected group that is a C-group and of the form $N \leftthreetimes A$ with N nilpotent and $A \simeq \mathbb{R}$ (see Varopoulos, 2000b, §3.1.2). Finally, as I have already pointed out, these are examples that are found in Gromov (1991).

9.2.7 The five-dimensional example of G_3

In the next section we shall give the formal definition of the generalisation of the embedded circle $S \subset G_2$ of §9.2.2 and construct an embedded $(r-1)$-dimensional sphere $S^{r-1} \subset G_r$ that generalises to higher dimensions the properties of S in G_2. This extension is natural, and in some sense it is not even very difficult to imagine, but to describe it precisely, a formidable array of notation will have to be introduced and the constructions become long and tedious to explain. Furthermore, no pictures can be drawn in that generality. We propose therefore to finish this paragraph in the following informal way. We shall draw a few pictures in $G_3 = N \leftthreetimes A = \mathbb{R}^3 \leftthreetimes \mathbb{R}^2$ as far as this is possible, and we hope that these will help the reader to improve their intuitive grasp on the constructions that will be made in §9.3 and §9.4 below.

Figure 9.5 shows $\square_R^3$ the R-cube sitting in $N = \mathbb{R}^3$. In Figure 9.6, the dotted lines indicate smooth arcs that spill out of $\mathbb{R}^3$ into G_3 (which is in $\mathbb{R}^5$). Their Riemannian length is $\approx \log R$ and they are the generalisations of the four arcs $\tilde{\gamma}_{ab}, \ldots$ in Figure 9.2. Their construction relies on (9.33) and the geometry of the simplex σ of §9.2.1.

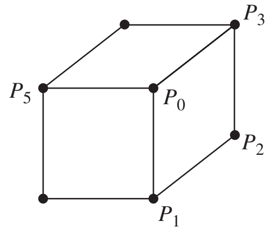


Figure 9.5

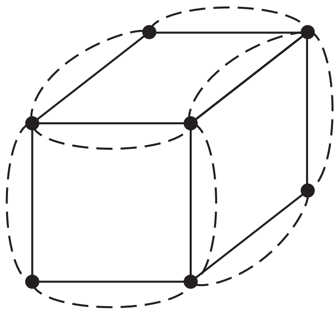


Figure 9.6

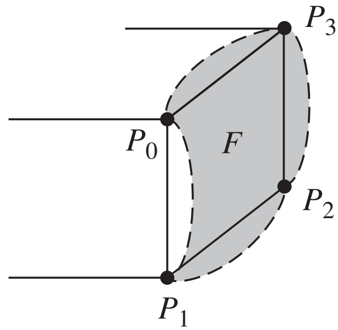


Figure 9.7

In Figure 9.7 we indicate one face of $\square_R^3$ in Figure 9.6 with its dotted sides which we fill in with a film $F \subset G_3$ (surface). This film is a deformation of the side $P_0P_1P_2P_3$. It spills out in such a way that area $\mathrm{Vol}_2(F)$ is also $\mathrm{O}\big((\log R)^C\big)$. The boundary of F consists of the four dotted edges of Figure 9.6. Now F corresponds to the face $P_0P_1P_2P_3$. We glue together the six deformed faces so obtained and obtain the required embedded sphere $S^2 = F_1 \cup \cdots \cup F_6$.

Figure 9.8 illustrates the transversality near the vertices. Here we look at the three deformed faces F, F', F'' that meet at P_0. The heavily shaded neighbourhood of P_0 in S^2 is *flat*, that is,it lies in $P_0 \times$ (some neighbourhood of 0 in $\mathbb{R}^2 = A$).

9.3 The First Basic Construction (I): The Description

9.3.1 An introduction and guide for the reader

In this subsection we shall deal exclusively with the group $G_r = \mathbb{R}^r \leftthreetimes \mathbb{R}^{r-1} = N \leftthreetimes A$ for $r \geqslant 2$ that was introduced in §9.2.1. The case $G = G_r$ comes about

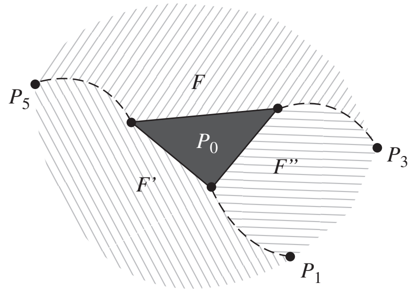


Figure 9.8

from the Abelian alternative of §9.1.7. The notation for the roots is as in (9.33), the coordinates $x_1,\ldots,u_1,\ldots$ are as in §9.2.1 and $\square^r_R \subset N$ is as in (9.34). In §9.2.2 we constructed an embedded circle $S = S^1 \subset G_2$ and in §9.2.7 we exhibited some pictures (Figures 9.5–9.8), and indicated how this construction generalises to G_3.

Here we shall describe the formal construction but since the notation needed for the construction is heavy we shall start with the auxiliary §9.3.2 where we shall codify some terminology, which is essential to keep the notation within acceptable limits both here and in the next section.

In §9.3.3 we shall describe precisely the properties that we shall require from $S \subset G_r$, which will be some embedded smooth $(r-1)$-dimensional sphere S^{r-1}. In §9.3.4 we shall show how these properties are used to prove that G_r does not have the $\mathscr{F}_r$-property. It is not until §9.4 that we shall grind out the details and the formal definition of S.

This choice of presentation makes it possible for the reader to go through this section and get a precise idea of what is happening without having to embark on the details of the construction in §9.4, which the reader could skip in a first reading. In going through this section the reader may find the pictures drawn in the previous section helpful.

9.3.2 Terminology and conventions

Throughout, M will be some Riemannian manifold and in our case it will almost exclusively be the Riemannian structure on some connected Lie group G. We shall consider mappings

$$\Phi\colon X \to M \tag{9.49}$$

where, here, X will be some metric space and not necessarily a Riemannian manifold. Typical examples for X are

$$\begin{aligned} &X = \square^r, \quad \square^r_R \text{ with } R = 1, \\ &X = \partial\,\square^r \text{ the boundary of } \square^r \text{ in } \mathbb{R}^r, \\ &X = \partial_0\,\square^r = [(x_1,\dots,x_r)\,;\, x_j = 0,1] \text{ the set of vertices of } \square^r, \\ &X = \partial_s\,\square^r_R \text{ the } s\text{-dimensional boundary.} \end{aligned} \tag{9.50}$$

Here $0 \leqslant s \leqslant r-1$ and for $s = r-1$ is the ordinary boundary, for $s = 1$ it is the set of edges of $\square^r_R$ and so on. The formal definition will be given in §9.4.1 below.

The functions (9.49) will be Lipschitz with Lipschitz constant l, that is, $\Phi \in \mathrm{Lip}(l)$. Throughout, $R \gg 1$ will be a free parameter that will be made to tend to ∞, and we shall denote

$$\Phi \in \mathrm{LL}(R) \Leftrightarrow \Phi \in \mathrm{Lip}\left(C(\log R)^C\right) \tag{9.51}$$

for some fixed constants $C > 0$ (i.e. $l = C(\log R)^C$); the two L's represent Lipschitz and log. If Φ in (9.49) is $\mathrm{LL}(R)$ we shall write

$$\Phi(X) \text{ is an } \mathrm{LL}(R) - X \subset M. \tag{9.52}$$

As an illustration, $\partial_0\,\square^r_R \subset N \subset G$ in §9.2.1 is then

$$\Phi_0\left(\partial_0\,\square^r\right) = \partial_0\,\square^r_R \text{ is an } \mathrm{LL}(R) - \partial_0\,\square^r \subset G \tag{9.53}$$

because, by §§8.2.1–8.2.2 and (9.33), the mutual distance of the vertices of $\square^r_R$ in G is $\lesssim \log R$. The mapping Φ_0 in (9.53) is the ordinary dilution $x \to Rx$ in $N = \mathbb{R}^r$.

What we shall do here is extend the mapping Φ_0 of (9.53) to higher-dimensional boundaries $\partial_s\,\square^r$ and construct, in particular,

$$S = \Phi\left(\partial\,\square^r\right) \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset G \tag{9.54}$$

for some function $\Phi \in \mathrm{LL}(R)$ that extends Φ_0, that is, $\Phi\left(\partial_0\,\square^r\right) = \Phi_0\left(\partial_0\,\square^r\right)$. For sure, in the construction (9.54) we are *not* going to have $S \subset N$. The pictures in §9.2 illustrate these constructions.

To show the flexibility of this terminology let us go back to the $\mathscr{F}_r$-property of §7.5. Then, to show that the manifold M admits this property, we must prove that for all Φ such that

$$E = \Phi\left(\partial\,\square^r\right) \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset M, \tag{9.55}$$

there exists $\hat{\Phi}$ such that

$$\hat{E} = \hat{\Phi}\left(\square^r\right) \text{ is an } \mathrm{LL}(R) - \square^r \subset M \tag{9.56}$$

and is such that

$$\hat{\Phi}\big|_{\partial\Box^r} = \Phi. \tag{9.57}$$

Or, if we use notation from cubic singular homology theory (see Massey, 1991) we can simply say

$$\partial\hat{E} = E. \tag{9.58}$$

We have abused terminology here mostly because we have made no mention of the constants. The formally correct way of saying things would have been this.

Let $C > 0$ be given, let $R \gg 10^{10}$ and let $\Phi \in \mathrm{LL}(R)$ with constants C (i.e. $\Phi \in \mathrm{Lip}(C(\log R)^C)$) be such that (9.55) holds. Then there exists C_1 that depends on C only (not on R nor on Φ) and there exists also $\widehat{\Phi} \in \mathrm{LL}(R)$ with constant C_1 (i.e. $\widehat{\Phi} \in \mathrm{Lip}\left(C_1(\log R)^{C_1}\right)$) that satisfies (9.56) and for which (9.57) holds.

The equality (9.58), if interpreted in terms of cubic singular homology, also is an abuse of terminology in homology theory because signs related to orientation are assigned to the various faces of the boundary operator ∂. The correct way to read (9.58) is that the support of $\partial\hat{E}$ is E.

Abuse or not, this terminology is very convenient and it will be used systematically. Its briefness outweighs the imprecision and we feel confident that no confusion will arise.

To fill in We shall express (9.58) in words by saying that $\hat{E}$ fills in E. This again is an abuse unless we bring in the two functions $\Phi, \widehat{\Phi}$ in (9.57).

The following statement is a sample of how we abuse terminology:

To show that G_r does not admit the $\mathscr{F}_r$-property it suffices to construct some

$$S \text{ that is an } \mathrm{LL}(R) - \partial\Box^r \subset G_r \tag{9.59}$$

that cannot be filled in by some

$$\hat{S} \text{ an } \mathrm{LL}(R) - \Box^r \subset G_r. \tag{9.60}$$

This is what was done in G_2 in §9.2.2. We can already see why it is convenient to use $\log R$ for the free parameter.

9.3.3 The description of the embedded sphere $S = S^{r-1} \subset G_r$

In this subsection we shall start from the final statement of the previous subsection and describe precisely what we need to construct.

The *sine qua non* properties of the construction are the following:

$$S = \Phi(\partial\Box^r) \text{ is an } \mathrm{LL}(R) - \partial\Box^r \subset G. \tag{9.61}$$

Let us for the last time write explicitly what (9.61) means: there exists

$$\Phi\colon \partial\square^r \longrightarrow G, \tag{9.62}$$

some Lip $\left(C(\log R)^C\right)$ mapping for fixed C but with R that is allowed to vary (of course Φ depends on R). The other essential condition on S is the following transversality condition.

Transversality With the notation of §9.3.2 we have

$$\begin{aligned} S\cap[(n,u)\in N\leftthreetimes A;\ |u|\leqslant 1] &= \bigcup[D(P);\ P\in\partial_0\square^r], \\ D(P) &= [(x,u)\in G;\ x=\Phi_0(P),\ |u|\leqslant 1]. \end{aligned} \tag{9.63}$$

Here the coordinates of §9.2.1 have been used and (9.63) means that in the cylinder above $|u|\leqslant 1$, the set S splits into a finite number of discs sticking out from the vertices $\Phi_0\left(\partial_0\square^r\right)=\partial_0\square^r_R$ as in (9.53). This is of course the obvious generalisation of (9.38) or of Figure 9.1.

Comments The above two properties will in fact suffice to give the proof that property $\mathscr{F}_r$ fails for G_r. This will be done in Chapter 10 and the proof is but a straightforward extension of §9.2.4 when $r=2$. These two properties suffice to show the existence of S, and we do not need to assume that Φ is smooth or that S is an embedded manifold. Therefore, the smoothing procedure that we shall explain later is not essential for our purposes. However, it is possible to guarantee a posteriori that

$$\Phi\in C^\infty;\quad S \text{ is an embedded } S^{r-1}\subset G, \tag{9.64}$$

that is, that S is an *embedded* $(r-1)$-dimensional sphere. The set-up (9.64) is exactly what we assumed in §9.2.2 for $r=2$ and we actually used this in the proof that we gave in §9.2.4. Figures 9.1–9.8 in §9.2 illustrate this smoothing well.

Regularisation First of all, it is easy to see that we can smooth Φ and preserve the transversality (9.63). The fact that S can be made an *embedded* S^{r-1} is more subtle and is in fact exactly the Whitney embedding theorem (see Hirsch, 1976, Chapter 2) which applies because

$$\dim(G_r)=2r-1\geqslant 2\left(\dim S^{r-1}\right)+1=2r-1. \tag{9.65}$$

This means that by a small perturbation of the original $C^\infty-S$ we can make it an embedded manifold. Since the perturbations that we performed are arbitrarily small, property (9.61) is preserved and we are done.

9.3.4 The endgame. Whitney regularisation and the $\mathscr{F}$-property

Once we have regularised S and we assume that S is an embedded $(r-1)$-dimensional sphere we can play the same endgame as in §9.2.4 and conclude that G_r does not admit the $\mathscr{F}_r$ property.

The general validity or non-validity of the $\mathscr{F}_r$ properties will be of central concern in Chapter 10 and they will be examined there in a systematic way. Here, however, we shall briefly indicate how the Whitney theorem, Sard's theorem and transversality (i.e. standard facts from differential topology, for which see de Rham, 1960 and Hirsch, 1976) can be used to settle this question very simply by extending the argument of §9.2.4. We proceed as follows.

We assume by contradiction that $\mathscr{F}_r$ holds on G_r. There exists then

$$D = \Phi(B) \text{ is an } \mathrm{LL}(R) - B \subset G_r, \tag{9.66}$$

where B is the unit ball in $\mathbb{R}^r$ and we can clearly assume that $\Phi \in C^\infty$ by simple regularisation, say by convolution. The Whitney theorem can then be used, and we replace D by $\widetilde{D}$ such that $\widetilde{D}$ is some smooth submanifold not of G_r, because there is no room in the dimensions as in (9.65), but it can be a smooth submanifold in $\mathbb{R}^a \times G_r$, for some positive a, and such that $\partial\widetilde{D} = \partial D = S$. For this it suffices that $a + 2r - 1 = \dim(\mathbb{R}^a \times G_r) \geqslant 2r + 1 = 2\dim D + 1$, that is, $a \geqslant 2$. In other words, we have to 'spill out' of G_r.

Once this is done, the group projection $G_r \to A = \mathbb{R}^{r-1}$ induces $\pi_a \colon \mathbb{R}^a \times G_r \to \mathbb{R}^{r-1}$ simply by sending to 0 the factor $\mathbb{R}^a$. For the regular values $u \in A$, $|u| \leqslant 1$ then $l_u = \pi_a^{-1}(u)$ are *neat submanifolds*[2] of dimension 1, that is, C^∞ arcs that have their boundary on the set (9.63). This last statement does not quite make sense of course because the set (9.63) is a subset of G_r and not of $\mathbb{R}^a \times G_r$. But the way to make a correct and precise statement out of this is rather obvious (see §10.4.3 below).

The length of these arcs is therefore $\geqslant cR$ for the Euclidean distance $\mathbb{R}^a \times \mathbb{R}^r \times \mathbb{R}^{r-1}$ on $\mathbb{R}^a \times G_r$. As before, the Euclidean distance and the Riemannian distance lie within constants in $[|u| < 1]$ and the contradiction with (9.60) follows as in (9.42) because by Sard's theorem the set of regular points is of full measure in $[u \in A;\ |u| \leqslant 1]$.

9.3.5 A different strategy

The first construction of the set S was done in §9.2.2 for the group D_2, and everything was explicit and simple: see (9.35) and Figures 9.1 and 9.2. In §9.2.3

[2] Yet another standard notion from differential topology; see Hirsch (1976, Chapter 1).

we combined this with the differential form ω of Example 7.16 and Exercise 7.17 to give a proof that the group D_2 satisfies the conditions of the C-theorem (Theorem 7.11). In that proof the C-condition of D_2 was used for the construction of the form ω which has to be closed and of polynomial growth. The properties needed from S, however, are much more flexible and much easier to describe. This fact has been pointed out already in §9.2.4, and no use of the transversality condition was made (in §9.2.4 we used the term 'backtracking'). More to the point, the C-condition on the group is not used in the construction of $S \subset D_2$.

Exactly the same strategy works for general D_r groups, with $r \geqslant 2$, and, following this strategy, a different proof can be given that all these groups satisfy the conditions of Theorem 7.11. The constructions of ω and $S \subset D_r$ involved here are *considerably* simpler than what we are about to do in the next paragraph. It is not obvious, however, we can complete the proof of Theorem 7.11 with this strategy because these simpler constructions do not seem to fit with the method that we shall adopt in §9.5 and Chapter 10. As a consequence, as things stand we have to proceed as in the next section.[3]

9.4 The First Basic Construction (II): Details and Computations

The issue here is to give the formal explicit definition of the function Φ and the set S of (9.54) that satisfy the properties that we explained in §9.3.3. To achieve this we shall start by constructing an appropriate *simplicial decomposition* of $\square^r$. The mapping Φ will then be a *simplicial mapping*, that is, piecewise affine. This is illustrated in Figure 9.1. To define the mapping Φ for $r \geqslant 2$ we need a complicated construction in *piecewise linear topology*. We shall offer two different constructions, the second at the end in §9.4.8, in order to give a choice to the reader. But unfortunately they are both quite long.

9.4.1 Notation on the unit cube

The unit cube in $\mathbb{R}^r$ was defined in §9.2.1 and can be taken to be (9.34) with $R = 1$, that is,

$$\square = \square^r = \big(x = (x_1, \dots, x_r) \in \mathbb{R}^r;\ 0 \leqslant x_j \leqslant 1,\ j = 1, \dots, r\big), \tag{9.67}$$

[3] **Added in proof:** In the Epilogue at the end of this book we discuss another approach to the problem that was worked out after the writing of the book was completed. In this new approach, we do not need the piecewise linear constructions of the next section which, though elementary, are long and tedious to describe.

where we shall drop the exponent r when no confusion can arise. The vertices are then

$$\partial_0 \square = (P = (\varepsilon_1, \dots, \varepsilon_r)\,;\ \varepsilon_j = 0, 1,\ j = 1, \dots, r)\,. \tag{9.68}$$

For every $I \subsetneqq [1, 2, \dots, r]$ we denote by $s = |I|$ the cardinality of that set and for fixed $P = (\varepsilon_1, \dots, \varepsilon_r) \in \partial_0 \square$ we denote the corresponding face by

$$F = F(I, P) = \left[0 \leqslant x_i \leqslant 1,\ i \in I;\ x_j = \varepsilon_j,\ j \notin I\right] \subset \partial \square\,. \tag{9.69}$$

The union of all s-dimensional faces is denoted by $\partial_s \square$ and $\partial_0 \square \subset \partial_1 \square \subset \cdots \subset \partial_{r-1} \square = \partial \square =$ (the topological boundary). Each face F is an affine cube and we shall denote the centre of F by $\xi_F \in F$. The notation is simplified further by writing $\partial_s = \partial_s \square$.

9.4.2 The simplicial decomposition of $\partial \square^r$

The constructions in this section will depend on a decomposition of $\partial \square$ into simplexes, that is, a simplicial decomposition $\mathscr{S}$ which we shall now describe in two different ways.

We shall denote by $P \in \partial_0$ some vertex of the cube and to fix ideas we shall assume below that $P = 0 \in \mathbb{R}^r$. Then we obtain as follows all the simplexes of dimension $r-1$ in $\mathscr{S}$ that admit P as a vertex. We first choose $I \subset [1, \dots, r]$, a subset of cardinality $r-1$, and fix some ordering $[i_1, \dots, i_{r-1}] = I$, that is, the i's are distinct. The simplex that corresponds is then

$$\sigma = \left[0 \leqslant x_{i_1} \leqslant x_{i_2} \leqslant \cdots \leqslant x_{i_{r-1}} \leqslant \tfrac{1}{2}\right]. \tag{9.70}$$

It is also clear that two different orderings give rise to simplexes with disjoint interiors (because an ordering is given by assigning $x \leqslant y$ or $y \leqslant x$ for all (x, y), and for different orderings these two differ for at least one pair).

We shall denote by $\mathscr{E}_s(P)$ the simplexes of dimension s with P as one of the vertices. The above description gives us all the simplexes in $\mathscr{E}_{r-1}(0)$. The definition of the simplexes in $\mathscr{E}_{r-1}(P)$ for the other vertices $P \in \partial_0$ is identical by making the obvious changes $x_j \to 1 - x_j$ whenever necessary.

The above simplexes cover $\partial \square$ because each $x \in \partial \square$ belongs to one of the simplexes of $\mathscr{E}_{r-1}(P)$, where $P = (\varepsilon_1, \dots, \varepsilon_r) \in \partial_0 \square$ is determined by $|x - P| = \max_i |x_i - \varepsilon_i| \leqslant 1/2$. The simplicial decomposition $\mathscr{S}$ of $\partial \square$ is then obtained by all the subsimplexes of all simplexes of $\mathscr{E}_{r-1}(P)$ for all $P \in \partial_0$.

An alternative description of $\mathscr{S}$ can be given as follows. The set of vertices $\mathscr{S}_0$ of the complex $\mathscr{S}$ is $\{\xi_F;\ F = \text{the faces (9.69)}\}$ the centres of the faces.

It follows in particular that $\partial_0 \subset \mathscr{S}_0$. Let us now fix some $P \in \partial_0$. The one-dimensional simplexes $\mathscr{E}_1(P)$ with one vertex at P are then the segments $[P, \xi_{F_1}]$ for all one-dimensional faces $F_1 \subset \partial_1$, as in (9.69), that contain P.

For every $\sigma_1 \in \mathscr{E}_1(P)$ and F_2, a two-dimensional face that contains σ_1, we construct $\sigma_2 =$ convex hull $[\sigma_1, \xi_{F_2}] \in \mathscr{E}_2(P)$, and so on inductively for $\mathscr{E}_j(P)$, the j-dimensional simplexes that contain P. More explicitly for $\sigma_j \in \mathscr{E}_j(P)$ and F_{j+1}, a $(j+1)$-dimensionless face that contains σ_j, we construct

$$\sigma_{j+1} = \text{convex hull } (\sigma_j, \xi_{F_{j+1}}) \in \mathscr{E}_{j+1}(P). \tag{9.71}$$

It follows that the simplexes in $\sigma \in \mathscr{E}_{r-1}(P)$ are in one-to-one correspondence with sequences of subsets

$$I_1 \subset \cdots \subset I_{r-1} \subset [1, 2, \ldots, r] \tag{9.72}$$

with $|I_j| = j$. These of course give rise to simplexes

$$\{P\} = \sigma_0 \subset \sigma_1 \subset \cdots \subset \sigma_{r-1} = \sigma; \quad \sigma_j \in \mathscr{E}_j(P). \tag{9.73}$$

With this interpretation of the construction, if $\sigma = \sigma' \in \mathscr{E}_{r-1}(P)$ then the corresponding sequence, $\sigma'_0 \subset \sigma'_1 \subset \cdots \sigma'_{r-2} \subset \sigma'$, has to be identical. To see this observe that because the $\xi_{F_{r-1}}$ that is used for both σ and σ' is the same, we must have $\sigma_{r-2} = \sigma'_{r-2}$. Then we argue by induction moving downwards in the dimension.

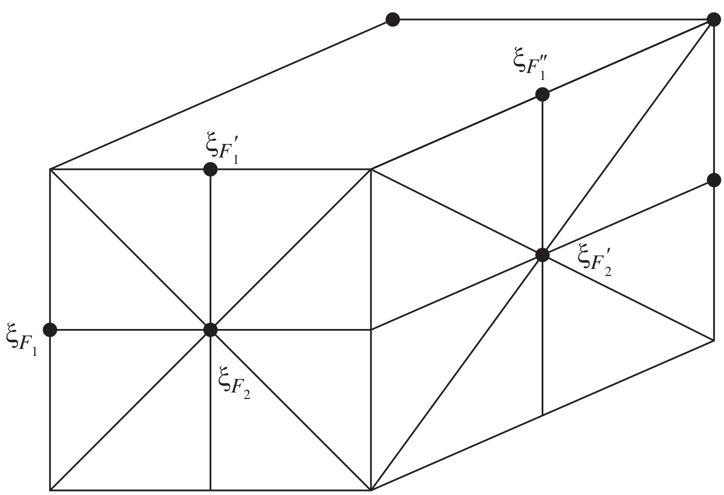


Figure 9.9 The triangulation of $\partial \square^3$.

9.4.3 The mappings. General description

We shall closely follow §9.2.1 and denote $G_r = N \leftthreetimes A$ with $N = \mathbb{R}^r, A = \mathbb{R}^{r-1}$ but we shall change the notation and instead write $A = V$ so the relevant space becomes $\mathbb{R}^r \times V$. We use this new notation to highlight the fact that the vector space structures of both $N = \mathbb{R}^r$ and $V = \mathbb{R}^{r-1}$ play a crucial role in the construction that we shall make. Two mappings

$$f_N : \partial \square^r \longrightarrow N, \qquad f_V : \partial \square^r \longrightarrow V \tag{9.74}$$

will be defined and examined in the next two sections. These two mappings will depend on the free parameter $R > 0$ of §9.3.2 and put together they will combine to give a mapping

$$f = (f_N, f_V) : \partial \square^r \longrightarrow \mathbb{R}^r \times V = G_r. \tag{9.75}$$

This mapping will have the properties described in §9.3. Namely

$$f(\partial \square^r) \text{ is an } \mathrm{LL}(R) \subset G_r \tag{9.76}$$

and it enjoys the $\mathrm{LL}(R)$ and transversality properties of §9.3.3.

Two different variants of the construction will be given and the reader can choose the one that suits them best. The first one is given in §§9.4.4–9.4.7, the second in §9.4.8. The actual problem with these constructions is that a lot of heavy notation is needed; otherwise they are but the obvious generalisation to higher dimensions of what was done in §9.2 for $r = 2, 3$. The reader is advised to take their time, to draw a few pictures, and to cross-check back and forth the material that follows from here up to §9.4.8, or, as a final resort, simply to believe the end result! This 'end result' is summarised in (9.75) and (9.76) but also a more precise description of the mapping f near the vertices of $\square^r$ will be needed. This additional information is described in detail in §§9.4.5–9.4.6. What this says is that near, say, the vertex $0 \in \square^r$, f_N stays constant (we say 'stuck' at 0) and f_V is one-to-one near 0 and piecewise linear on the simplexes of $\mathscr{E}_{r-1}(0)$. See Figure 9.1 for the case $r = 2$.

9.4.4 The mapping that does the stretching

In this subsection we shall define mappings $f : E \to \partial \square^r$ where $E = \partial_j$, with $j = 0, 1, \ldots$. These mappings are successive extensions of each other and eventually E will be the whole of $\partial \square$. A property that all these mappings will have is that for $F \subset E$, some face of $\square^r$ as in (9.69), we have $f(F) \subset F$. We shall say that f has the *δ-retract property* for $0 < \delta < 1/2$ if $f(x) = f(x')$ for two

$x=(x_1,\dots,x_r)$, $x'=(x'_1,\dots,x'_r)\in F$ where the coordinates only differ when they are δ-close to the end points 0 or 1. More precisely, if we denote

$$J_0=\big[j;\ 0\leqslant x_j\leqslant\delta\big],\qquad J'_0=\big[j;\ 0\leqslant x'_j\leqslant\delta\big],$$
$$J_1=\big[j;\ 1-\delta\leqslant x_j\leqslant 1\big],\qquad J'_1=\big[j;\ 1-\delta\leqslant x'_j\leqslant 1\big],$$

and if we assume $J_0=J'_0$, $J_1=J'_1$ and $x_i=x'_i$, $i\notin J_0\cup J_1$, then $f(x)=f(x')$.

We now move on to define the mappings that do the '*stretching*'. A typical stretching mapping, or more precisely a δ-stretching mapping, on $[0,1]$ is given by

$$f(x)=\begin{cases}0 & \text{for } x\in[0,\delta],\\ 1 & \text{for } x\in[1-\delta,1],\\ \text{affine} & [\delta,1-\delta],\end{cases}$$

so the three pieces fit to give a continuous function on [0,1].

Our mappings on $\partial\,\square$ will be defined inductively on $\partial_0\subset\partial_1\subset\cdots$ and they will be required to have the δ-retract property. We define $f_0\colon\partial_0\to\partial_0$ to be the identity. With $f_s\colon\partial_s\to\partial_s$ assumed already defined, we first define $f_{s+1}\colon F\to F$ for every $(s+1)$-dimensional face, (9.69), on the collar of width δ of the boundary ∂F. More precisely, this collar, $C_\delta=C_\delta(F)$, for $F=F(I,P)$ of (9.69), is defined as the set $x=(x_1,\dots,x_r)\in F$ for which there exists at least one index in I, say $i_0\in I$, such that either $x_{i_0}\in[0,\delta]$, or $x_{i_0}\in[1-\delta,0]$. In the first case, we set $\overline{x}_{i_0}=0$ and in the second $\overline{x}_{i_0}=1$. We also set $x'_i=x_i$ if $i\neq i_0$, and $x'_{i_0}=\overline{x}_{i_0}$. We then extend the definition of f_s to C_δ by setting $f_{s+1}(x)=f_s(x')$. By putting all the $C_\delta(F)$ together, we obtain in this way an extension to a δ-neighbourhood of ∂_s.

One easily verifies that this definition is unambiguous in case there are several candidates for the index i_0. To see this, one uses the δ-retract property that, by induction, f_s satisfies on all the faces in ∂_s. The inductive construction clearly shows that f_{s+1} is Lipschitz on $C_\delta(F)$ if f_s is already Lipschitz on ∂_s. (For this verification we must estimate the difference $f_{s+1}(x)-f_{s+1}(y)$ when $x,y\in C_\delta(F)$ and *in addition* x,y differ only on one of their coordinates, and say $x_i=y_i$, $i\neq 1$. If $x_2,\dots,,x_r\in[\delta,1-\delta]$ and x,y are close, then that difference vanishes. If not, then by our construction we can use the inductive hypothesis.) And of course, foremost, f_{s+1} is an extension of f_s, that is, $f_{s+1}\big|_{\partial_s}=f_s$. The definition of f_{s+1} on the complement $F\backslash C_\delta$ will be done below but, from what we already have, it is clear that f_{s+1} has the δ-retract property.

It remains to define f_{s+1} on $F_\delta=F\backslash C_\delta$, that is, the complementary set of the collar C_δ in F. This could be done in any way whatsoever as long as it respects the Lipschitz property and $f_{s+1}(\xi_F)=\xi_F$. To fix ideas, we shall make

f_{s+1} continuous and affine on every line segment in F between ξ_F and C_δ on which f_{s+1} is already defined.

The normalisation and the definition of f_N With $R \geqslant 1$ the free parameter of §9.3.2 we shall now define the mapping

$$f_N = R f_{r-1} \colon \partial \square \longrightarrow \partial \square_R, \tag{9.77}$$

where of course $\square_R = R\square \subset \mathbb{R}^r$ in the R-cube $= [0 \leqslant x_i \leqslant R;\ i = 1, \dots, r]$ and $f_N \in \mathrm{Lip}(cR)$ for some c and there exists some small $c_1 > 0$ such that f_N stays constant in the corresponding neighbourhoods of the vertices

$$f_N\,[x;\ |x - P| < c_1] = f_N(P); \quad P \in \partial_0 \square. \tag{9.78}$$

The definition of f_V and the transversality We shall now give the definition of f_V in (9.75), (9.76). For this we use the simplicial decomposition $\mathscr{S}$ of §9.4.2 and, for every vertex $\xi_F \in \mathscr{S}_0$ which is the centre of the face F, we set $f_V(\xi_F) = \zeta_I \in V$ where I is the subset I in the definition (9.69) of the face F and ζ_I is an *appropriate choice of vectors of* V. The definition of f_V is then completed on the whole of $\partial \square$ by requiring that it is an affine mapping on each simplex of $\mathscr{S}$ (i.e. $f(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 f(x_1) + \lambda_2 f(x_2)$; $\lambda_1, \lambda_2 \geqslant 0$, $\lambda_1 + \lambda_2 = 1$). In this definition two different faces defined by the same index I but with different vertex P in (9.69) give rise to the same $f_V(\xi_F) \in V$ and this mapping is continuous.

The choice of the vectors $\zeta_I \in V$ will be crucial for (9.76) and for the transversality condition §9.3.3. That choice will also depend on the free parameter $R > 1$ by scaling as

$$\zeta_I = c_0\,(\log R)\,\zeta_I^*; \quad I \subsetneqq [1, \dots, r],\ R \geqslant 1, \tag{9.79}$$

where c_0 will be chosen later and will be large enough. The $\zeta_I^* \in V$ on the other hand will depend only on the roots $L_1, \dots, L_r$ of (9.33) and not on R, and we shall set $\zeta_\emptyset^* = 0$. The construction of the other ζ_I^* will be given in §9.4.6 and for this purpose in §9.4.5 we shall introduce some notation.

9.4.5 Affine mappings in conical domains

Let $V = \mathbb{R}^{r-1}$ as before and let $e_1, \dots, e_n \in V$ $(n = r - 1)$ be a basis. For vectors $x_1, \dots, x_m \in V$ we shall use the notation

$$\mathrm{CC}(x_1, \dots, x_m) = \sum_{j=1}^{m} \rho_j x_j; \quad \rho_j \geqslant 0 \tag{9.80}$$

for the convex conical domain they generate. For every non-empty subset $I \subset [1,2,\ldots,n]$ we shall define once and for all $e_I = \sum_{i\in I} \lambda_i^{(I)} e_i$ with fixed positive coefficients $\lambda_i^{(I)} > 0$. For consistency set $e_\emptyset = 0$. We shall further write $\mathrm{CC} = \mathrm{CC}(e_1,\ldots,e_n)$ and $\mathrm{CC}_{\mathscr{J}} = \mathrm{CC}(e_{J_1}, e_{J_2},\ldots,e_{J_n})$ where

$$\mathscr{J} : J_1 \subset J_2 \subset \cdots \subset J_n = [1,2,\ldots,n] \tag{9.81}$$

is an arbitrary sequence of subsets with $|J_j| = j$. The following facts must be verified.

The conical domains $\mathrm{CC}_{\mathscr{J}}$ as $\mathscr{J}$ runs through all the choices (9.81) form a tessellation of CC. This means that the conical domains $\mathrm{CC}_{\mathscr{J}}$ have non-empty disjoint interiors and that their union is CC.

Exercise 9.29 Verify this. As in (9.72), (9.73) there is a one-to-one correspondence between $\mathscr{J}$ in (9.81) and $\mathrm{CC}_{\mathscr{J}}$. To see that we have a covering of CC we take a slice with a hyperplane, consider the simplex $\sigma = [e_1,\ldots,e_n]$, and write $\dot{e}_I = \left(\sum \lambda_i^{(I)}\right)^{-1} e_I \in \operatorname{Int}\sigma$ with $I = J_n$ and the simplexes

$$\sigma_j = [e_1,\ldots,e_{j-1},\dot{e}_I,e_{j+1},\ldots,e_n] \subset \sigma,$$

that is, we replace e_j by $\dot{e}_I$. These simplexes cover σ. Then we argue by induction on the dimension.

The next obvious fact is that if $e'_1,\ldots,e'_n \in V$ is another basis and $e'_I = \sum \mu_i^{(I)} e'_i$ some choice of the e'_I as before, then there is a continuous bijection from $\mathrm{CC}(e_1,\ldots,e_n)$ to $\mathrm{CC}(e'_1,\ldots,e'_n)$ that for each $\mathscr{J}$, as in (9.81), is the restriction of a linear mapping from $\mathrm{CC}(e_{J_1},\ldots,e_{J_n})$ to $\mathrm{CC}(e'_{J_1},\ldots,e'_{J_n})$.

Exercise 9.30 Verify this. It is a simple consequence of the fact that for any sequence $\mathscr{J}$ in (9.81), $e_{J_1},\ldots,e_{J_n}$ is a basis of V and the linear mapping is obtained by $e_{J_i} \to e'_{J_i}$. That the linear mappings glue together and define a mapping on $\mathrm{CC}(e_1,\ldots,e_n)$ follows because if $\mathscr{J} : J_1 \subset \cdots$, $\mathscr{K} : K_1 \subset \cdots$ are as in (9.81) then $\mathrm{CC}_{\mathscr{J}} \cap \mathrm{CC}_{\mathscr{K}} = \mathrm{CC}(e_{I_1},\ldots,e_{I_n})$ where $I_i = J_i$ if $J_i = K_i$ and $I_i = \emptyset$ otherwise.

The tessellation of V Let $\sigma = [u_1,\ldots,u_r] \subset V$ be some non-singular simplex of V such that $0 \in \operatorname{Int}\sigma$. A tessellation as before can then be obtained for the whole space V. This is done by using the conical domains $\mathrm{CC}_i = \mathrm{CC}(u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_r)$ where the vectors that generate CC_i are a basis of V. These basis vectors are then used as above. More precisely, as in §9.4.5, we define

$$e_I = \sum_{i\in I} \lambda_i^{(I)} u_i, \tag{9.82}$$

where the $\lambda_i^{(I)} > 0$ are given and are defined for every $I \subset [1,\dots,r]$ with $|I| \leqslant r-1$ and $i \in I$. These e_I can then be used to tessellate further each CC_i as in Exercise 9.29. In that fashion we finally obtain a tessellation of the whole space V into polyhedral convex conical domains.

Note that the above construction will be revisited in §10.3.

9.4.6 The choice of the ζ_I and the transversality condition §9.3.3

The choice of the ζ_I in §9.4.4 will now be made by choosing first some non-singular simplex $\sigma = [u_1,\dots,u_r] \subset V$ with $0 \in \operatorname{Int}\sigma$. We then set in (9.79) $\zeta_I^* = \sum_{i\in I} \lambda_i^{(I)} u_i = e_I$ of (9.82). The definition of f_V and the tessellation of §9.4.5 imply therefore that the transversality condition of §9.3.3 holds for the mapping $f = (f_N, f_V)$ of (9.75).

More precisely, using §9.4.5 we see that for every $P \in \partial_0 \square$, the mapping f maps each small neighbourhood of size 1/4 of P in $\partial \square$ in a piecewise affine and continuous manner on $\{P\} \times V \subset \mathbb{R}^r \times V$. If R is large enough, the transversality condition is therefore satisfied. To see this, note that if $(X, \partial_0 \square) > c$ for some c and $X \in \partial \square$, then $|f_V(X)|$, for $f_V(X) \in V$, is as large as we like provided that R is large enough. To see this last point we use the fact that f_V is affine on the simplexes of $\mathscr{S}$ and that for any such simplex that is defined by $\big(I_1 \subset \cdots \subset I_{r-1} \subset [1,2,\dots,r]\big)$ with $|I_j| = j$ as in (9.72), the convex hull $[\zeta_{I_1},\dots,\zeta_{I_r}]$ is far from 0. This function f_V so defined gives a one-to-one piecewise affine mapping between small neighbourhoods in $\partial \square$ of the vertices $P \in \partial_0 \square$ and a neighbourhood of $0 \in V$. These facts combined with the locally constant property (9.78) of f_N prove the transversality.

The dual simplex and an illustration Here briefly we shall explain an alternative way of defining the ζ_I. The reader who, like the author, does not enjoy piecewise linear combinatorial constructions, can ignore the next few lines. We shall start with $\sigma^* = [L_1,\dots,L_r] \subset V^*$ some non-singular simplex in the dual space of V, with 0 in the interior, for example the simplex determined by the roots as in §9.2.1. We can then define $\sigma = [e_1,\dots,e_r] \subset V$ some dual simplex that satisfies $L_j e_i > 0$ for all $i \neq j$ (see §10.3.2: this is also non-singular and 0 is in the interior). We can then give an alternative definition for the ζ_I in (9.79) by setting

$$\zeta_I^* = \sum_{j \notin I} e_j; \quad \emptyset \neq I \subset [1,\dots,r],\ |I| \leqslant r-1. \tag{9.83}$$

This can now be used as in §9.4.4–§9.4.6 to define first a tessellation of V and then the mapping f_V is defined by setting $\zeta_I = c_0(\log R)\zeta_I^*$ in (9.79).

Exercise 9.31 Prove this and prove that the transversality condition is satisfied again if R is large enough. We decompose the dual simplex σ into $(r-1)$-dimensional simplexes with 0 as common vertex and where the other vertices, $(r-|I|)^{-1}\zeta_I^*$, lie on the topological boundary. We then have a natural one-to-one correspondence between the above simplexes and those of $\mathscr{E}_{r-1}(P)$ (see §9.4.4). The details are left to the reader.

This approach, which is dual to the previous one, has the following advantage. It gives for free the additional condition

$$L_i\zeta_I \geqslant c_1 \log R; \quad i \in I, \tag{9.84}$$

for any $I \subset [1,2,\dots,r]$ with $|I| \leqslant r-1, \emptyset \neq I$, where the $c_1 > 0$ can be chosen as large as we like provided that $c_0 > 0$ in (9.79) is large enough.

Condition (9.84) plays a crucial role in proving the LL(R) property of the mapping (9.75). We shall examine property (9.84) from our previous point of view in the next subsection.

Proof of (9.84) **without using the dual simplex** It will be convenient now to assign $V = \mathbb{R}^{r-1}$ with its natural Euclidean inner product structure and identify it thus with its dual space V^*.

Let us fix $L_1,\dots,L_m \in V$ non-zero vectors that satisfy the NC-condition of §§2.2 and 2.3.1. Using the inner product, this means that there exists $u \in V$ such that

$$\langle L_j, u\rangle = L_j u > 0; \quad j = 1,\dots,m. \tag{9.85}$$

It is easy to show that it is possible to choose u in (9.85) to be of the form

$$u = \zeta = \sum_{j=1}^{m} \lambda_j L_j; \quad \lambda_j > 0. \tag{9.86}$$

Once this is proved we can go back to §9.4.5 and consider some simplex $\sigma = [L_1,\dots,L_r] \subset V$ with $0 \in \operatorname{Int}\sigma$. For every $I \subsetneqq [1,\dots,r]$, with $\emptyset \neq I$, we can then use (9.86) and define appropriate vectors

$$\zeta_I^* = \sum_{i\in I} \lambda_i^{(I)} L_i, \qquad \zeta_I = c_0(\log R)\zeta_I^* \tag{9.87}$$

for which (9.84) holds.

This is done now directly on σ and we do not need to use the dual simplex. As a result it is easier to keep track of the tessellation of V that is involved in the simplicial mapping f_V.

Exercise 9.32 Prove (9.86) (see Varopoulos, 2000b, §1.1.5). In §9.4.6 we avoid this but then we have to prove the existence of the dual simplex.

9.4.7 The $\mathrm{LL}(R)$ property of the mapping f in (9.75)–(9.76)

We already saw that f as defined in §9.4.4 is a Lipschitz mapping $f\colon \partial\square \to \mathbb{R}^r \times V$. Now consider $\Omega \subset \partial\square$ the open, everywhere-dense subset that consists of the union of all the interiors $\operatorname{Int}\sigma$ of the simplexes of maximal dimension $r-1$ in $\mathscr{E}_{r-1}(P)$ (cf. §9.4.2), where P runs through the set of all the vertices $P \in \partial_0\square$. For any $X \in \Omega$ we can then define the differential of f at X. We shall show that the norm of that differential satisfies

$$|df| \leqslant c\log R; \quad X \in \Omega,\ R > 10, \tag{9.88}$$

for some $c > 0$ provided that the constant c_0 in (9.79) has been chosen large enough. The norm of the differential in (9.88) is taken with respect to the Euclidean norm on $\square^r \subset \mathbb{R}^r$ and the group Riemannian structure of §7.2 on $G_r = \mathbb{R}^r \leftthreetimes V = N \leftthreetimes A$ (putting $V = A$) of §9.2.1. We have thus two Lipschitz properties. The first is global, but where the control of the Lipschitz constant is not specified. The second is on the set Ω; that is, a set of 'full measure' and here we have in addition the correct estimate (9.88). These two properties combined together imply the required property $f \in \mathrm{LL}(R)$. For more details on the correct Lipschitz constant see §7.1.1.

In the next few lines we shall prove (9.88) for

$$X \in \mathring{\tau} = \left[\tfrac{1}{2} > x_1 > x_2 > \cdots > x_{r-1} > 0;\ x_r = 0\right], \tag{9.89}$$

where $\tau \in \mathscr{E}_{r-1}(0)$, which is a typical interior of the simplexes that we are considering: see (9.70). Every other simplex can be renormalised by reordering the coordinates and replacing x_j by $1-x_j$ when necessary, bringing the simplex to the above case (9.89).

We shall first prove (9.88) with f replaced by f_V. This is easy because by the construction of f_V we have

$$|df_V| \leqslant c\log R; \quad R \geqslant 10, \tag{9.90}$$

for some c that depends on c_0 of (9.79). Inequality (9.90) holds because the values $f_V(\xi_F) = \zeta_I$ are $\approx \log R$ apart and $f_V \in V$ is an affine mapping on each simplex of $\mathscr{S}$ of §9.4.2.

Now we examine f_N and use the δ of the retract property of that mapping defined in §9.4.4. For $X = (x_1, x_2, \ldots, x_{r-1}, 0)$ as in (9.89) we shall determine the index k by

$$\tfrac{1}{2} > x_1 > \cdots > x_k > \delta > x_{k+1} > \cdots \tag{9.91}$$

and the possibility that some coordinate is δ can be discarded by reducing further the set Ω to $\Omega \cap (\text{no coordinate is equal to } \delta)$. By the construction of the mapping f it then follows that for X as in (9.91) we have

$$df_N\left(\frac{\partial}{\partial x_i}\right) \in N_k = \mathrm{Vec}\,(e_1,\dots,e_k) \subset N = \mathbb{R}^r; \quad i = 1,\dots,r, \tag{9.92}$$

where $e_1,\dots$ are the canonical basis vectors of $\mathbb{R}^r$. (For this observe two things. First, if $i > k$ then changing the ith coordinate x_i slightly does not change the value of $f_N(X)$. Second, we always have $f_N(F) \subset F$ for any face $F \subset \partial_p$ and for any dimension.)

To prove (9.88) for f_N the following observation about f_V will be needed. If the coordinates of $X_0 = (1/2,\dots,1/2,x_{k+1},\dots,x_{r-1},0)$ are all 1/2 up to the kth, and every other coordinate is $\leqslant\delta$, and if $I_j = [1,\dots,j]$, then

$$X_0 \in \text{ convex hull } \left[\xi_{I_k},\dots,\xi_{I_{r-1}}\right].$$

This follows from the alternative way the simplex τ in (9.89) is constructed (see §9.4.2) with 'successive' vertices $0, \xi_{I_1}, \xi_{I_2},\dots$ and the fact that the first k coordinates are $1/2$. From this and the definition of f_V in §9.4.4 it follows that

$$f_V(X_0)/\log R \in c_0 \text{ convex hull } \left[\zeta^*_{I_k},\dots,\zeta^*_{I_{r-1}}\right], \tag{9.93}$$

where the ζ^*_I are as in §9.4.6. It follows that if δ in (9.91) is close enough to 1/2 then, with X as in (9.91), $f_V(X)/\log R$ is close enough to the convex hull of (9.93) because of the affine definition of f_V. This in turn implies that on the Lie group $G_r = N \leftthreetimes V$ the $(\mathrm{Ad}(f_V(X)))^{-1}$ restricted on the subspace $N_k \subset N$ of (9.92) will have a norm $\lesssim R^{-C}$ for an arbitrarily large $C > 0$ provided that the c_0 in (9.79) is large enough. This last assertion is a consequence of the choice of the ζ^*_I in §9.4.6 and in (9.83) or (9.87). Indeed, that choice guarantees that (9.84) holds and this, in view of formula (9.33) for Ad in the group G_r, is exactly what is needed for our assertion.

Our assertion on estimate (9.88) for df_N then follows by (8.42) for the Riemannian structure on G_r.

The proof of the LL(R) property for f in (9.75)–(9.76) is thus complete.

9.4.8 An alternative description of the construction

What follows is a variant (see Varopoulos, 2000b) of the construction of $f = (f_N, f_V)$ of (9.75) that was done in the previous sections and it will give the reader another way of going through the details if they wish. We shall fix some $P \in \partial_0$ and some $\sigma \in \mathscr{E}_{r-1}(P)$ and define

$$f\colon \sigma \longrightarrow \sigma \times V \tag{9.94}$$

in a way such that those mappings for the different choices of the simplexes σ fit together to give a global mapping f on $\partial\,\square$, as required.

The following notation will be used. First of all let $\sigma \in \mathscr{E}_{r-1}(P)$ for some $P \in \partial_0$, and let $\{P\} = \sigma_0 \subset \cdots \subset \sigma_{r-1} = \sigma$ be as in (9.73) where $\xi_j = \xi_{F_j} \in \sigma_j$ are the centres of the faces F_j that contain σ_j as in (9.71). For every $\theta = (\theta_1, \ldots, \theta_{r-1})$ with $\theta_j \in [0,1]$ we can define inductively $x_j \in \sigma_j$ by

$$\big(1-\theta_{j+1}\big)\xi_{j+1} + \theta_{j+1}x_j = x_{j+1}; \quad 0 \leqslant j \leqslant r-2,\ x_0 = P. \tag{9.95}$$

Similarly, let $\zeta_j = \zeta_{I_j} \in V$ be as in §9.4.4, (9.79), for the subsets $I_j = I_{F_j}$ that determine the faces F_j defined in (9.69). We can again define inductively $y_j \in V$ by

$$\big(1-\theta_{j+1}\big)\zeta_{j+1} + \theta_{j+1}y_j = y_{j+1} \in V, \qquad y_0 = 0. \tag{9.96}$$

We combine these to define

$$\begin{gathered} \Phi(\theta_1, \ldots, \theta_{r-1}) = (x_{r-1}, y_{r-1}) \in \sigma \times V, \\ \Phi = (\Phi_N, \Phi_V) : [0,1]^{r-1} \longrightarrow \mathbb{R}^r \times V = N \leftthreetimes V = G_r. \end{gathered} \tag{9.97}$$

Observe that if $\theta_j = 0$ for some $j \geqslant 1$, then $x_j = \xi_j, y_j = \zeta_j$ and from then onwards only $\theta_{j+1}, \theta_{j+2}$ come into play. One could say that the 'memory before j' has been erased and $\Phi(\theta)$ is independent of θ_i for $i < j$.

The mapping $\Phi_N(\theta_1, \ldots, \theta_{r-1}) = x_{r-1} = x \in \sigma$ admits an inverse

$$\Phi_N^{-1} \colon \operatorname{Int}\sigma \longrightarrow [\theta_j;\ 0 < \theta_j < 1,\ 1 \leqslant j \leqslant r-1], \tag{9.98}$$

which is also defined inductively by (9.95). To wit, we start from $x_{r-1} = x \in \operatorname{Int}\sigma$ and this in (9.95) determines uniquely (x_{r-2}, θ_{r-1}) with $x_{r-2} \in \sigma_{r-2}$ and $0 < \theta_{r-1} < 1$ and then repeat with x_{r-2} and work our way all the way down to (x_0, θ_1) with $x_0 = P$.

We shall now use some function $\alpha \colon [0,1] \to [0,1]$ and, abusing notation, also denote $\alpha \colon [0,1]^{r-1} \to [0,1]^{r-1}$ by $\alpha(\theta_1, \ldots, \theta_{r-1}) = (\alpha(\theta_1), \ldots, \alpha(\theta_{r-1}))$. Now consider the composed mapping

$$f_N = \Phi_N \circ \alpha \circ \Phi_N^{-1} \colon \operatorname{Int}\sigma \longrightarrow \sigma. \tag{9.99}$$

The function α will be non-decreasing, smooth and equal to 0 in some neighbourhood of 0, and equal to 1 in some neighbourhood of 1. See Figure 9.10.

The function f_N in (9.99) is then Lipschitz and extends to the whole of σ. To verify this Lipschitz property we shall fix some $\varepsilon > 0$, that will be determined later and will depend on the function α, and consider the following cases.

Case (i) $x \in \operatorname{Int}\sigma$ is such that the coordinates $(\theta_1(x), \dots, \theta_{r-1}(x))$ of $\Phi_N^{-1}x$ are all $\theta_j > \varepsilon$. In that case the Lipschitz condition for $\Phi_N^{-1}x$ holds in some neighbourhood of x with a constant that depends on ε.

This is seen by the inductive construction of the $\theta_{r-1}, \theta_{r-2}, \dots$. To wit, $\theta_{r-1}(x)$ is always a Lipschitz function of x. But if $\theta_{r-1}(x) \geqslant \varepsilon$ we see from (9.95) (with $x_{j+1} = x_{r-1} = x$) that $x_j = x_{r-2}$ is also a Lipschitz function of x. We then repeat the argument using (9.95) with $x_{j+1} = x_{r-2}$, $\theta_{r-2}(x) \geqslant \varepsilon$, and so on. The function f_N in (9.99) is therefore Lipschitz in that range with a constant that depends on ε.

Case (ii) If case (i) fails it could be that $\theta_{r-1}(x) \leqslant \varepsilon$. On the other hand, if case (i) fails and $\theta_{r-1}(x) > \varepsilon$ let k, with $1 \leqslant k \leqslant r-2$, be the first integer for which $\theta_{r-k-1} \leqslant \varepsilon$. In that case,

$$\theta_{r-1}(x) > \varepsilon, \dots, \theta_{r-k}(x) > \varepsilon; \quad \theta_{r-k-1}(x) < 2\varepsilon. \tag{9.100}$$

Incidentally, case (i) can be interpreted as (9.100) with $k = r-1$. Similarly, (9.100) with $k = 0$ is interpreted as $\theta_{r-1} < 2\varepsilon$. At any rate, in case (ii) when (9.100) holds, even when $k = 0$, the functions $\theta_{r-1}, \dots, \theta_{r-k-1}$ are Lipschitz in a neighbourhood of x with a Lipschitz constant that depends only on ε for the same reasons as before.

To determine the value $f_N(x) = \Phi_N \circ \alpha \circ \Phi_N^{-1}(x)$, we must use the inductive hypothesis construction (9.95). It then follows that if for some x and j we have $\alpha\theta_j(x) = 0$ (an abuse of notation for $\alpha \circ \theta_j$) then, as observed just after (9.97), $f_N(x)$ depends only on $\alpha\theta_{r-1}(x), \dots, \alpha\theta_j(x)$ and is independent of $\alpha\theta_i(x)$ for $i < j$. This observation is now combined with condition (9.100), where we make sure that the choice of ε is such that $\theta_{r-k-1}(x) < 2\varepsilon$ implies that $\alpha\theta_{r-k-1}(x) = 0$. This, and our previous remark about the Lipschitz nature of $\theta_{r-1}, \dots, \theta_{r-k-1}$, completes the proof that f_N is Lipschitz.

Definition 9.33 (Definition of f_V) We can now use another non-decreasing smooth function $\beta\colon [0,1] \to [0,1]$ with the same property, that $\beta(0)$ equals 0 in some neighbourhood of 0 and equals 1 in some neighbourhood of 1 (see Figure 9.10), and compose again

$$f_V = \Phi_V \circ \beta \circ \Phi_N^{-1}\colon \operatorname{Int}\sigma \longrightarrow V. \tag{9.101}$$

For the same reason as before, this is again Lipschitz and it extends to σ and, considering once more the above two cases, we see that the Lipschitz norm satisfies $\operatorname{Lip} f_V = O(\sup|\zeta_j|^C)$. If we combine (9.99) and (9.101), we finally obtain the required mapping:

$$f = (f_N, f_V) : \sigma \longrightarrow \sigma \times V. \tag{9.102}$$

Two points have to be verified.

Point 1 The functions f defined in (9.94) for the simplexes σ of maximal dimension ‘fit together’. More precisely, they extend to some unique global function,

$$f\colon \partial\square \longrightarrow \partial\square\times V, \tag{9.103}$$

that is Lipschitz. Furthermore, for the choice in (9.96) of the ζ_I, as in (9.79) and for an appropriate choice of the shape the ‘stretching functions’ α, β (see Figure 9.10), the above function f satisfies the transversality condition of §9.3.3.

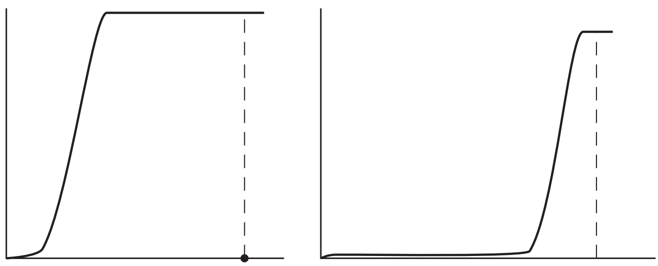


Figure 9.10 The two functions used in (9.99) and (9.101).

Point 1 is easy to verify. Probably the best way to see this is to define directly the global mapping

$$f_N\colon \partial\square \longrightarrow \partial\square,$$
$$f_N\colon \partial_j \longrightarrow \partial_j; \quad j=0,1,\ldots,r-1,$$

inductively on j: assuming that f_N is defined on ∂_j we consider $F\subset\partial_{j+1}$, some face, and for all $x\in F$ different from ξ_F we can define x', θ by

$$x=(1-\theta)\xi_F+\theta x'; \quad \theta\in]0,1],\ x'\in F\cap\partial_j. \tag{9.104}$$

The extension of f_j to F is then given by

$$f_{j+1}(x)=(1-\alpha(\theta))\,\xi_F+\alpha(\theta)f_j(x'). \tag{9.105}$$

If we set $f_{j+1}(\xi_F)=\xi_F$ we have a definition on the whole of F and obtain thus the required extension of f_N to ∂_{j+1}. This proves the required inductive step of the construction and completes the definition of

$$f_N\colon \partial\square \longrightarrow \partial\square. \tag{9.106}$$

Once θ has been determined as above, the construction can be made for

$f_V : \partial\square \to V$ by (9.96). The two put together give the construction of the function in (9.103). The fact that this function f restricted to each simplex σ of maximal dimension is our original function in (9.94), (9.102) can again be seen by induction by considering $\sigma_0 \subset \cdots \subset \sigma_{r-1} = \sigma$ as in (9.73). To wit, we consider the successive restrictions of f_N to the lower-dimensional subsimplexes σ_j, defined either as in (9.94) or as in (9.103) and show that they coincide, and similarly for f_V.

The transversality condition is a consequence of the shape of the two functions α, β that will be chosen to look as in Figure 9.10, where (thinking of time as moving backwards from 1 to 0) α does not move away from 1 until β has completed its 'life span'. With the functions α, β as above we prove the transversality condition with the same tessellation argument that was used in §9.4.6. The details are identical and will not be repeated. More about the verification of transversality can be found in the original presentation of this construction in Varopoulos (2000b). Notice also that, in the above construction, the mapping f, (9.103), is constant in small neighbourhoods of the vertices. This can easily be rectified and the mapping can be made one-to-one there by a slight modification of β near 1.

Rescaling Let us now go back to the mapping Φ defined in (9.97). The parameter $R \geqslant 1$ is built into the mapping Φ_V because the choice of the ζ_I in (9.96) depends on R. We shall now modify f in (9.102) and, using for simplicity the same notation, set

$$f = (R\Phi_N \circ \alpha,\ \Phi_V \circ \beta) \circ \Phi_N^{-1} \in \mathbb{R}^{r-1} \times V, \tag{9.107}$$

where we use scalar multiplication by R in $\mathbb{R}^r$ and the composition with α and β. For this mapping and the Riemannian group structure on G_r (and the Euclidean structure on $[0,1]^{r-1}$) we shall prove point 2.

Point 2 If c_0 in (9.79) is large enough, the function f of (9.107) satisfies $f \in \mathrm{LL}(R)$. Given that f is already known to be Lipschitz (as was verified when $R = 1$ in (9.99)), this follows from the fact that

$$|df| \leqslant C \log R; \quad R \geqslant 10, \tag{9.108}$$

for some $C > 0$ for the same reason as before (explained in §7.1.1).

We now come to the proof of (9.108). By the definition of the ζ_I and the Lipschitz norm of f_V that we estimated in (9.101), it is clear that only the N-coordinate of f in (9.107) is an issue. We must therefore estimate the differential of the function $\Phi_N \circ \alpha \circ \Phi_N^{-1} = \Psi$. This will be done in the next few lines.

By the properties of α and β the function Ψ is constant in some neighbourhood of a point x for which $\beta_j = \beta \circ \theta_j(x) \neq 0$, with $1 \leqslant j \leqslant r-1$. We may therefore assume that there exists some $1 \leqslant p \leqslant r-1$ such that $\beta_p = 0$ and for such a p we have

$$f_V(x) = y_{r-1} = \text{convex hull } [\zeta_{r-1}, \dots, \zeta_p] \tag{9.109}$$

because of the inductive definition in (9.96) (this should be compared with (9.93) and with the comments after (9.97)). Let p be the largest possible integer for which $\beta_p = 0$, so that either $p = r-1$ or

$$\beta_p = 0, \qquad \beta_{p+1} \neq 0, \dots, \beta_{r-1} \neq 0. \tag{9.110}$$

But by the choice of the functions α, β from (9.109) it follows that

$$\alpha_{p+1} \circ \Phi_N^{-1} = \cdots = \alpha_{r-1} \circ \Phi_N^{-1} = 1 \tag{9.111}$$

in some neighbourhood of x. By (9.97), (9.99) (see the inductive procedure (9.95) that gives the definition), this means that $\Psi = \Phi_N \circ \alpha \circ \Phi_N^{-1} \in \sigma_p$ (defined in (9.73)). Let $F = F(I_p, P) \in \partial_p$ be the p-dimensional face (9.69) that contains $\sigma_p \subset F$. Then the differential of this mapping Ψ takes its values in the vector subspace

$$\text{Vec}\left[\frac{\partial}{\partial x_i};\ i \in I_p\right] = N_p \subset TN, \tag{9.112}$$

the tangent subspace generated by the coordinates in I_p. Of course, (9.112) also holds trivially when $p = r-1$.

Now, in view of (8.42), where we specified the Riemannian structure on G_r, what needs to be estimated is $\text{Ad}(f_V(x))^{-1}$, and more precisely, because of (9.112), it is the restriction of that action on N_p that counts. The argument used for this estimate is exactly as at the end of §9.4.7 (especially the use of (9.93), which now is replaced by the use of (9.109)) and we shall not repeat the details. The crux is that because of (9.84) in the choice of the ζ_I, the norm of that restriction is $\text{O}(R^{-c})$ for any preassigned c provided that the c_0 in (9.79) is large enough. We use formula (9.33) again here for Ad in the group G_r.

9.5 The Second Basic Construction

9.5.1 The general SSG that are C-groups

We shall retain our previous notation and if G is some general SSG that is a C-group, then, by §9.1.7, we have two alternatives.

The Abelian alternative The C-group G is of the form $G = N \leftthreetimes (A' \oplus A)$ where N, A', A are the simply connected groups that correspond to $\mathfrak{n}$, $\mathfrak{a}'$, $\mathfrak{a}_0$ of (9.32). We have dropped the index 0 here because no confusion will arise. The groups A', A are Euclidean spaces: $A' \cong \mathbb{R}^{r-1}$, $A \cong \mathbb{R}^s$, with $r \geqslant 2$, $s \geqslant 0$. The group N contains $N' \subset N$ some subgroup that is Abelian and is the subgroup that corresponds to the subalgebra $\mathfrak{n}'$: $N' \cong \mathbb{R}^r$ and the elements of N' commute with the elements of A. We shall write

$$G' = N' \leftthreetimes (A' \oplus A) = (N' \leftthreetimes A') \oplus A, \qquad G = N \leftthreetimes (A' \oplus A). \tag{9.113}$$

Furthermore, $N' \leftthreetimes A'$ is a G_r group as in §9.2.1 and the action of A' on N' is given by roots $L_1, \ldots, L_r \in (A')^*$ (the dual space). These roots satisfy the C-condition as in §9.2.1.

The Heisenberg alternative We preserve the same notation for $G = N \leftthreetimes (A' \oplus A)$ but now $\dim A' = 1$ and $N' \subset N$ is a Heisenberg group with Lie algebra $\mathfrak{n}'$ as in Definition 9.5, and the Lie algebras of A', A are $\mathfrak{a}'$, $\mathfrak{a}_0$, respectively, as in (9.32). The action of A', A on N, N' is induced by the corresponding actions in (9.32) and for the group

$$G' = N' \leftthreetimes (A' \oplus A) = (N' \leftthreetimes A') \oplus A, \tag{9.114}$$

the action of A on N' is trivial and $N' \leftthreetimes A'$ is then the group K that we considered in §9.2.5.

When $A = \{0\}$, the constructions that we made in §9.2.5 (for the Heisenberg case) and in §§9.2–9.4 (for the Abelian case) suffice for the proofs that we shall give in the next chapter. The new problem that arises in the general situation for both the above alternatives is the presence of the additional group A. Indeed, while A splits off as a direct factor in the group G', its action on N is in general non-trivial. This radically changes the whole picture and new ideas are needed. *For the rest of this section we shall assume that $A \neq 0$ and that $s \geqslant 1$.*

9.5.2 The new tools. A special case

Here we shall introduce the new idea that is needed to handle the general SSG of the previous sections. We shall do so first for the Abelian case in the special situation $s = 1$ in §9.5.1 and, with the notation from there, what we have is

$$G' = G_r \oplus A, \quad G = N \leftthreetimes (A' \oplus A), \quad A = \mathbb{R}. \tag{9.115}$$

We shall show that the following result.

Proposition 9.34 *One or both of properties $\mathscr{F}_r$, $\mathscr{F}_{r+1}$ fail to hold on G.*

This proposition is a special case of the following more general one.

Proposition 9.35 *Let $G = N \leftthreetimes (A' \oplus A)$ with $G_r = N' \leftthreetimes A'$ and $A = \mathbb{R}^s$ as in §9.5.1. Then at least one of properties $\mathscr{F}_r, \mathscr{F}_{r+1}, \ldots, \mathscr{F}_{r+s}$ fails to hold on G.*

The new idea involved is that while we shall not be able to pinpoint the exact index j for which $\mathscr{F}_j$ fails, we shall at least show it fails for some j. This of course is good enough to show that these groups are not polynomially retractable as in Theorem 7.11.

We shall argue by contradiction and for a group G as in Proposition 9.34, we shall assume that both properties $\mathscr{F}_r$, $\mathscr{F}_{r+1}$ hold. A contradiction will be obtained. Towards that we shall assume that condition $\mathscr{F}_r$ holds on G and proceed to make the following construction in G. That construction will be shown to be contradictory with $\mathscr{F}_{r+1}$.

9.5.3 The construction needed for Proposition 9.34 under $\mathscr{F}_r$ for $s = 1$

The notation is as above. Our first step is to construct $S \subset G_r$ that satisfies properties (9.61), (9.63), that is,

$$S \text{ is an } \mathrm{LL}(R) - \partial \square^r \subset G_r \tag{9.116}$$

and the transversality condition is satisfied.

The underlying difficulty is the following. We showed in §9.3.4 that this construction contradicts property $\mathscr{F}_r$ in G_r. The problem is that property $\mathscr{F}_r$ could still hold in the larger group G. In other words, although we may not be able to fill in S by some S' that is $\mathrm{LL}(R) - \square^r \subset G'$ (where the terminology is as in §9.3.2), this may still be possible if we allow S' to spill out of G' and be some $\hat{S}$ that is $\mathrm{LL}(R) - \square^r \subset G$ and such that $\partial \hat{S} = S$.

The difficulty does not lie in the fact that $N' \subsetneq N$ because, as long as $A = \{0\}$, using the same ideas as §9.3.4 we can easily adapt that proof here. (This will be one of the things that will be done in the general proof in Chapter 10. Adapting §9.3.4 in this case is an interesting exercise that the reader may like to think about.) The problem is rather the presence of A that has a non-trivial action on N.

Be that as it may, as we said, the assumption is that indeed $\mathscr{F}_r$ does hold for the group G and we shall proceed with our construction.

The set $S \subset G'$ can be translated to $S_a = S + a \subset G$ ($a \in A$) and since A lies in the centre of G' we can of course use either the left or right group action. Additive notation has deliberately been used to stress this point. Using the left action we see that

$$S_a \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset G \tag{9.117}$$

uniformly in $a \in A$. Since property $\mathscr{F}_r$ has been assumed in G, we can fill in S_a in G and for each $a \in A$ construct

$$\begin{gathered}\hat{S}_a \text{ is an } \mathrm{LL}(R) - \square^r \subset G,\\ \partial\hat{S}_a = S_a\end{gathered} \tag{9.118}$$

and the $\mathrm{LL}(R)$ constants in (9.118) are uniform in $a \in A$. We shall denote by Φ_a the function that does the filling: $\hat{S}_a = \Phi_a(\square^r)$.

Let us now consider the set $X = S \times [-a_0, a_0]$ for some $a_0 \in A = \mathbb{R}$ with $a_0 = c_0(\log R)^{c_0}$ for some large c_0 that will be chosen later. What counts is that

$$X \text{ is an } \mathrm{LL}(R) - (\partial\,\square^r) \times [0,1] \subset G' = G_r \times A. \tag{9.119}$$

This is a consequence of the direct product structure and the choice of a_0. We shall denote by Φ_X the function that does the mapping in (9.119). The $\mathrm{LL}(R)$ constants in (9.119) depend on c_0 and the $\mathrm{LL}(R)$ constants of (9.117). We can now consider the set

$$Y = \hat{S}_{-a_0} \cup X \cup \hat{S}_{a_0} \text{ is an } \mathrm{LL}(R) - \partial\,\square^{r+1} \subset G. \tag{9.120}$$

This set Y looks like an empty food can. The lateral boundary X sits in G' but the top and bottom spill out to G. The $\mathrm{LL}(R)$ property follows by glueing together two $\mathrm{LL}(R)$ functions

$$\begin{gathered}\Phi_{\pm a_0} \colon \square^r \longrightarrow G,\\ \Phi_X \colon \partial\,\square^r \times [0,1] \longrightarrow X\end{gathered} \tag{9.121}$$

that coincide on their common range and construct the function needed in (9.120).

The additional crucial geometric property that the set Y has is described after the following exercise.

Exercise One thing that we are *not allowed* to do for the construction of the $\hat{S}_a$ is to fix one, say $\hat{S}_0$ (i.e. $a = 0$) and then 'translate it around' to $a + \hat{S}_0$ in G. Why not? In the construction, the fact that A lies in the centre of G' plays a crucial role.

Transversality We shall denote by $(n, a', a) \in G$, with $n \in N$, $a' \in A'$, $a \in A$, the coordinates in G induced by (9.113). We can make the choice of c_0, in the choice of a_0, large enough depending on the $\mathrm{LL}(R)$ constants of $\hat{S}_{\pm a_0}$ in (9.118) (that are uniform in a_0) in such a way that $[|a| < 1] \cap \hat{S}_{\pm a_0} = \emptyset$. This holds because for c_0 sufficiently large the Lipschitz condition (9.118) on $\hat{S}_{\pm a_0}$ does not allow it to reach anywhere near $a = 0$. It follows that

$$Y \cap [|a| < 1] = X \cap [|a| < 1], \tag{9.122}$$

$$Y \cap \left[|a'| < 1,\ |a| < 1\right] = X \cap \left[|a'| < 1\right] \cap [|a| < 1]. \tag{9.123}$$

Since X is the product set $S \times [-a_0 < a < a_0]$, the set in (9.123) is simply

$$\left(S \cap \left[|a'| < 1\right]\right) \times (|a| < 1). \tag{9.124}$$

The first factor in (9.124) is the 'old friend' from the transversality condition (9.63), which applies so that the set (9.124) has exactly the same shape as (9.63).

The same argument as in §9.3.4 can therefore be applied and shows that (9.120) and the transversality property of (9.124) imply that G cannot have the $\mathscr{F}_{r+1}$ property and Y cannot be filled in with some

$$\hat{Y} \text{ that is an } \mathrm{LL}(R) - \square^{r+1} \subset G.$$

Hence the required contradiction.

We shall not give more details here because all of this will be treated in detail and in full generality in Chapter 10.

A modification in the construction of (9.120) Instead of allowing the size of the interval $[-a_0, a_0] \approx c_0(\log R)^{c_0}$ we shall exploit the uniformity in the parameter a of (9.117) and apply the above argument to $X_n = S \times [n, n+1]$. Then (9.119) holds for X_n with constants that are uniform in n. We define $Y_n = \hat{S}_n \cup X_n \cup \hat{S}_{n+1}$ and then (9.120) holds for Y_n with constants that are uniform in n. By glueing together and erasing the intermediate boundaries we obtain

$$\Omega = \bigcup_{-N \leqslant n \leqslant N-1} Y_n \text{ which is an } \mathrm{LL}(R) - \partial\,(\square^r \times [-N, N]) \subset G \tag{9.125}$$

and the $\mathrm{LL}(R)$ constant is uniform in N. For the same reason as before, if N is large enough so that $\hat{S}_{-N}, \hat{S}_N$ are far enough, (9.122)–(9.123) hold for this Ω.

To finish the argument we must get rid of the N in (9.125). But this is easy because if $\Phi\colon \partial\,(\square^r \times [-N, N]) \to G$ is Lip $\left(c(\log R)^C\right)$ then the rescaled function on the last coordinate gives a mapping $\partial\,\square^{r+1} \to G$ that, being a composition of two Lipschitz functions, is Lip $(cN(\log R)^c)$. A choice of $N \approx (\log R)^{c_0}$ for c_0 large enough will therefore do the trick. Arguably, nothing has changed in this 'new proof'. However, this is the point of view that has to be adopted in the general case that we shall give in the next subsection for $s \geqslant 1$.

9.5.4 The construction needed for Proposition 9.35 with $s \geqslant 1$

The strategy for the proof of Proposition 9.35 is identical. We shall assume that $\mathscr{F}_r, \mathscr{F}_{r+1}, \ldots, \mathscr{F}_{r+s-1}$ hold and proceed to perform an analogous construction. That construction at the end will contradict property $\mathscr{F}_{r+s}$ and this will provide the required contradiction.

A priori this extension seems 'innocent'. We construct $S \subset G'$ as in §9.5.3 and one should be able to repeat the previous construction by generalising $X = S \times [-a_0, a_0]^s$ of (9.119), one would think. This approach can be made to work but there is an unexpected snag. The point is that, exactly as in (9.120), X is only part of the boundary $\partial \square^{r+s}$: the other part 'looks like' $\hat{S} \times \partial([-a_0, a_0]^s)$ for some $\hat{S} \subset G$ that has been used to fill in S. In doing that one can proceed inductively and consider successively $\hat{S} \times \partial_{s_1}([-a_0, a_0]^s)$ (notation of §9.4.1, $s_1 = 1, 2, \ldots$) and this is exactly what we shall do below. However, at the very end we are stopped by the choice of c_0 in $a_0 = c_0(\log R)^{c_0}$ that is needed for the transversality §9.3.3. If we try to carry out the details we see that 'things do not add up'. Therefore this approach has to be carried out with care. Already the case $s = 2$ is typical. We shall therefore start with the following.

Construction for $s = 2$ under $\mathscr{F}_r, \mathscr{F}_{r+1}$ The way we shall organise the notation for $s = 2$ is as follows. We shall write

$$L = [a_{1,1}, a_{1,2}, a_{2,2}, a_{2,1}] \subset A, \tag{9.126}$$

a rectangle with sides parallel to the axes, and denote the two sides by

$$[a_{i,1}, a_{i,2}] = l_i^1, \qquad [a_{1,i}, a_{2,i}] = l_i^2. \tag{9.127}$$

The size of L is $|l_i^1|, |l_i^2| \leqslant 1$ but all the constructions below will be uniform with respect to the initial position $a_{1,1} \in A$. The fact that G' and A commute as before implies that for $i, j = 1, 2$,

$$S_{i,j} = S \times \{a_{ij}\} \text{ is an } \mathrm{LL}(R) - \partial \square^r \subset G' \tag{9.128}$$

and similarly, also,

$$S \times l_i^1,\ S \times l_i^2 \text{ are } \mathrm{LL}(R) - (\partial \square^r) \times [0,1] \subset G'. \tag{9.129}$$

Using (9.128) and the $\mathscr{F}_r$-property in G that is assumed to hold, we can construct $B_{i,j}$ such that

$$\begin{aligned} &B_{i,j} \text{ is an } \mathrm{LL}(R) - \square^r \subset G, \\ &\partial B_{i,j} = S_{i,j}; \quad i, j = 1, 2. \end{aligned} \tag{9.130}$$

But then by the same argument as in (9.121), the two sets

$$X_i^1 = B_{i,1} \cup B_{i,2} \cup \left(S \times l_i^1\right),$$
$$X_i^2 = B_{1,i} \cup B_{2,i} \cup \left(S \times l_i^2\right) \qquad (9.131)$$

are both

$$X_i^k \text{ are } \mathrm{LL}(R) - \partial\,\square^{r+1} \subset G; \quad i,k = 1,2 \qquad (9.132)$$

and the constants are uniform with respect to the initial position $a_{1,1}$. We can therefore use the $\mathscr{F}_{r+1}$-property in G and fill in the two sets (9.131) so that for $i,k = 1,2$,

$$B_i^k \text{ is an } \mathrm{LL} - \square^{r+1} = \square^r \times [0,1] \subset G, \qquad (9.133)$$

$$\partial B_i^k = X_i^k. \qquad (9.134)$$

Now, for the same reason as in (9.119), (9.129),

$$S \times L \text{ is an } \mathrm{LL}(R) - (\partial\,\square^r) \times \square^2 \subset G'. \qquad (9.135)$$

So we can glue things together and we see that

$$\Delta = \bigcup_{i,k} B_i^k \cup (S \times L) \text{ is an } \mathrm{LL}(R) - \partial\,\square^{r+2} \subset G. \qquad (9.136)$$

The set in (9.136) is given by five pieces and, as in (9.121), it is clear that each piece is determined by an $\mathrm{LL}(R)$ function. The thing that might be slightly challenging for the readers who (like the author of this book) do not have much background in combinatorial topology, is to verify that the pieces do indeed fit together to make up a boundary of a $\square^{r+2}$.

Exercise 9.36 One construction that may clarify matters is the boundary $\partial(\square^r \times \square^2)$ that consists of $(\partial\,\square^r) \times \square^2$, which looks like $S \times L$ in (9.136), and $\square^r \times \partial\,\square^2$. In this second component of the boundary, the four sides in $\partial\,\square^2$ give terms that in view of (9.133) look like B_i^k.

The transversality We now return to the argument where we were left: on the set Δ at (9.136). We start by subdividing the large square $[-N,N]^2 \subset \mathbb{R}^2 = A$ into unit subsquares $L_{n,m} = [n, n+1] \times [m, m+1]$ as in §9.5.3 and we perform the above construction with each $L_{n,m}$. This, however, is done in such a way that the $B_{i,j}(n,m)$, with $i,j = 1,2$, that correspond to the four vertices of $L_{n,m}$ depend only on the position of that vertex and not on which kind (left or right, top or bottom) of vertex it is. This amounts of course to specifying some filling for each $S \times (n,m)$ that satisfies properties (9.128). Similarly, the B_i^k of (9.133) are made to depend only on the position of the edge $l_{n,m}^{(1)} = (n, [m, m+1])$ or

$l_{n,m}^{(2)} = ([m, m+1], n)$ as the case might be. For every $L_{n,m}$, we have constructed thus some $\Delta_{n,m}$ as in (9.136). The union $\Omega = \bigcup \Delta_{n,m}$ is like an empty honeycomb and the walls of two neighbouring cells $\Delta_{n,m}$ are the same. The inner walls 'cancel' and only the outer walls of Ω remain. More precisely,

$$\Omega = \left(S \times [-N, N]^2\right) \cup \text{ outer walls.} \tag{9.137}$$

What we mean by walls in the above statements is the part of the boundary of $\Delta_{n,m}$ that corresponds to $\bigcup B_i^k$ in (9.136). The outer walls correspond to $\Delta_{n,m}$ with n or $m = \pm N$ that are constructed by some edge $l_{n,m}^{(1)}$ or $l_{n,m}^{(2)}$ with $n = \pm N$. By the LL(R) properties of these walls, which are uniform with respect to the base point $(n, m) \in A$, the outer walls in (9.137) are far from $a = 0$ in the coordinates (x, a', a) of §9.5.3. The transversality property, (9.122)–(9.124), for Ω therefore follows for $N = C(\log R)^C$ and C large enough depending on the constants of the $\mathscr{F}_r$, $\mathscr{F}_{r+1}$ properties that have been assumed to hold on G. As in (9.125),

$$\Omega \text{ is an } \mathrm{LL}(R) - \partial \left(\square^r \times [-N, N]^2\right) \subset G \tag{9.138}$$

and therefore, by the same rescaling as in §9.5.3,

$$\Omega \text{ is an } \mathrm{LL}(R) - \partial \square^{r+2} \subset G \text{ if } N = c_1 (\log R)^{c_1}. \tag{9.139}$$

From Ω the contradiction with property $\mathscr{F}_{r+2}$ that is assumed to be satisfied on G follows again as in §9.3.4. But as said before, this contradiction will be explained in detail in Chapter 10.

Remark If, instead of the $\sim N^2$ different cells in the above honeycomb configuration, we had worked with just one cell of size N, we would have four walls whose size would also have depended on N. And here lies the problem: by choosing N large enough we would not have been able to conclude that, for $a \in A$, the a-coordinate on these walls does not get close to zero. It follows that the transversality property is not provable that way.

9.5.5 The general case $s \geqslant 2$

As in (9.126) we shall start with some cube,

$$L = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_s, b_s] \subset A \tag{9.140}$$

of size $|a_i - b_i| \leqslant 1$ and proceed exactly as in the special cases of §§9.5.3–9.5.4. To make the notation manageable we shall denote by $P_0, P_0', P_0'', \ldots \in \partial_0 L = \partial_0$ the vertices of the cube L; further, we denote $P_1, P_1', \ldots \in \partial_1 L = \partial_1$ and $P_2, P_2', \ldots \in \partial_2 L$, and so on. We generalise (9.128), (9.129) and, for $j = 0, 1, \ldots,$

we use the notation $S_j = S \times P_j$, $S'_j = S \times P'_j, \dots$ for $(P_j, P'_j, \dots \in \partial_j)$, so that $S_0, S'_0, \dots S_1, S'_1$ and are the generalisations of (9.128) and (9.129) respectively. We now generalise (9.130) and use the $\mathscr{F}_r$-property of G to define $B_0, B'_0, \dots,$ the fillings of $S_0, \dots$ that satisfy

$$B_0 \text{ is an } \mathrm{LL}(R) - \square^r \subset G; \quad \partial B_0 = S_0. \tag{9.141}$$

We define $B'_0, \dots$ similarly. Now let $P_1 \in \partial_1$ and let P_0, P'_0 be its two boundary points; then

$$X_1 = S_1 \cup B_0 \cup B'_0 \text{ is an } \mathrm{LL}(R) - \partial\, \square^{r+1} \subset G. \tag{9.142}$$

This generalises (9.131) and is proved in an identical manner. The analogous $X'_1, \dots$ are constructed. Property $\mathscr{F}_{r+1}$ can then be used and we construct

$$B_1 \text{ is an } \mathrm{LL}(R) - \square^{r+1} \subset G; \quad \partial B_1 = X_1. \tag{9.143}$$

This is what generalises (9.133)–(9.134). To carry out one more step we define

$$X_2 = S_2 \cup B_1 \cup B'_1 \cup B''_1 \cup B'''_1 \tag{9.144}$$

for some $P_2 \in \partial_2$ and where the B_1's are constructed in (9.143), (9.144) from the four sides $P_1, P'_1, \dots \in \partial_1$ of the square P_2. This is the analogue of the Δ in (9.136). Property $\mathscr{F}_{r+2}$ is then used to fill in X_2 and obtain B_2, and so on.

At the end, when we have exhausted all dimensions $0, 1, \dots, s$, we obtain a number, $(2N)^s$, of cells B_s that corresponds to the honeycomb of §9.5.4 and these cells have common walls. The union is

$$\Omega = (S \times [-N, N]^s) \cup \text{ outer walls.} \tag{9.145}$$

The walls of the cells B_s are $B_{s-1} \cup B'_{s-1} \cup \cdots$ in the last step of the construction

$$X_s = S_s \cup B_{s-1} \cup B'_{s-1} \cup \cdots \tag{9.146}$$

and by outer walls we mean the B_{s-1} that are constructed by the $P_{s-1} \in \partial_{s-1}$ that are contained in $\partial[-N, N]^s$. The $\mathrm{LL}(R)$ properties of the walls of the cells are uniform with respect to all the cells and therefore the outer walls are far from any point whose A-coordinate $a = 0$ (for the coordinates we use the notation of §9.5.3) as long as $N = c_1 (\log R)^{c_1}$ is chosen with c_1 large enough. The transversality property for Ω as in §9.5.3 follows and we have the analogue of (9.122):

$$\Omega \cap [|a| < 1] = (S \times [-N, N]^s) \cap [|a| < 1]. \tag{9.147}$$

On the other hand, by our construction, Ω is an $\mathrm{LL}(R) - \partial\, (\square^r \times [-N, N]^s) \subset G$. By the choice of N and by the same rescaling as in §9.5.3, it follows that

$$\Omega \text{ is an } \mathrm{LL}(R) - \partial\, \square^{r+s} \subset G. \tag{9.148}$$

As in §9.3.4, the two properties can now be used to contradict condition $\mathscr{F}_{r+s}$ that is assumed to hold on G. This contradiction will be explained in detail in Chapter 10.

Note The above description is informal and we feel that this is the best way to present things. Note that, for technical reasons, we shall have to come back to the above constructions again in §10.3.7. Note also that a more formal treatment of the same problem is given in §13.6. In fact, in Chapter 13 the construction will be formalised for the use of homology theory with the systematic use of currents. This allows us to present everything that we did in this section in algebraic terms. The reader could anticipate matters by looking ahead if so wished.

9.5.6 The Heisenberg alternative

We now move to the Heisenberg alternative of §9.5.1 and if we use the notation there we have $N' \leftthreetimes A' = K$ where K is as in §9.2.5. As before we denote $s = \dim A$. We can then construct an embedded circle $S^1 \cong S \subset K$ that goes through the segments of the word (9.44). This is the only change that we need to make to prove the analogue of Proposition 9.35.

Proposition 9.37 *Let G and $G' = K \times A$ be as above in the Heisenberg alternative. Then the group G cannot satisfy all of the properties $\mathscr{F}_2, \mathscr{F}_3, \ldots, \mathscr{F}_{s+2}$.*

In other words, at least one of the properties $\mathscr{F}_2, \ldots, \mathscr{F}_{s+2}$ fails.

This proposition is again proved by assuming that properties $\mathscr{F}_2, \ldots, \mathscr{F}_{s+1}$ hold for the group G and making a construction that, combined with $\mathscr{F}_{s+2}$, will lead to a contradiction. This construction is nearly (verbatim) identical to the constructions of §9.5.5.

The only difference is that now we start from this new $S^1 \cong S \subset K$ instead of the previous $S^{r-1} \cong S$. Then, if $s = 1$, we proceed as in §9.5.3 and the transversality condition (i.e. the analogue of Theorem 9.25) holds. For this construction $\mathscr{F}_2$ is used. If $s \geqslant 2$ the construction is done step by step as in §9.5.4–9.5.5 until we reach the top dimension. We construct then the analogue of Ω in (9.145) and the corresponding transversality property §9.3.3 holds for this Ω. For this construction $\mathscr{F}_2, \ldots, \mathscr{F}_{s+1}$ is used. Rewriting the details of these constructions does not seem to be necessary.

The final contradiction with $\mathscr{F}_{s+2}$ is then obtained by the method of §9.2.6; details will again be given in Chapter 10.

9.5.7 A recapitulation

We shall collect together the essential features of the constructions of this chapter. We shall use the notation $G = N \leftthreetimes V$, $G' = N' \leftthreetimes V$ for the SSG of §9.5.1, where N, N' are nilpotent, A, A' are Abelian and where we write $V = A \oplus A'$. This notation is the same for both the Abelian and the Heisenberg cases.

The special case $A = \{0\}$ was treated in §9.2–9.4. In the Abelian case of §9.5.1, $A' = \mathbb{R}^{r-1}$, with $r \geqslant 2$. In that case we constructed $S = \Phi(\partial \square^r)$ an $\mathrm{LL}(R)$ subset of $N' \leftthreetimes A'$, with $R \geqslant 1$. The set $N \cap S = E$ consists of 2^r distinct points $\Phi(\partial_0 \square^r)$. There exist constants such that

$$d(e,e') \geqslant CR^c \qquad e,e' \in E \qquad e \neq e' \qquad S \cap N = E, \tag{9.149}$$

where we can measure the distance in N because of the polynomial distortion of distances between N and N' (see §2.14).

In the Heisenberg case of §9.2.5, $A' = \mathbb{R}$, $r = 2$ and $A = \{0\}$. We then constructed $S = \Phi(\partial \square^r)$ an $\mathrm{LL}(R)$ subset of G' and now $E = N \cap S = (a, b, \ldots, f)$ are the six segments of the word (9.44) (cf. Figure 9.3). There exist again constants for which (9.149) holds.

In the general situation of §§9.5.5, 9.5.6, we placed ourselves in the case $A = \mathbb{R}^s$, with $s \geqslant 1$ and we assumed that the conditions $\mathscr{F}_r, \ldots, \mathscr{F}_{r+s-1}$ hold in G. In both the Abelian and the Heisenberg cases we started from the previous construction of $S \subset N' \leftthreetimes A'$ and E the same set as above. Then we proceed to construct some

$$\Omega \text{ is an } \mathrm{LL}(R) - \partial \square^{r+s} \subset G; \quad \Omega \supset S,\ \Omega \cap N = E. \tag{9.150}$$

Furthermore, if we denote by $\pi\colon G \to V$ the canonical projection and $V_1 = [v \in V;\ |v| < 1]$ the unit ball of V, then we have the transversality condition

$$\Omega \cap \pi^{-1}(V_1) = E \times V_1. \tag{9.151}$$

10

The Endgame in the C-Theorem

In this chapter we put together what was done in Chapters 7–9 to complete the proof of the C-theorem (Theorem 7.11). What is difficult in this proof is that it uses the different components that were constructed in the course of these previous chapters. Collecting together and describing these components without going through the details of the constructions again is what we have tried to do in the first section of this chapter. The reader is invited to spend time there and to make sure that they understand what is involved, by cross-checking with the previous three chapters for the definitions and notation that are used. Once this is done we give two different proofs of the C-theorem. One is based on ideas from differential topology (see de Rham, 1960; Hirsch, 1976), the other on ideas from geometric measure theory (see Federer, 1969). More precisely, the proof of the C-theorem is reduced to the proof of Proposition 10.5 and, as explained in §10.1.7, two different proofs of this proposition will be given in §§10.2 and 10.4 respectively. Of the proofs, the former is by far the shorter of the two, but to make it self-contained, familiarity with the notion of currents is needed. The purpose of the second proof is to avoid this notion. More will be said on these two proofs in §§10.1.7–10.1.8.

The reader is encouraged to navigate freely in this chapter and not to try to read it rigidly from beginning to end. Better still, once the reader has captured the general idea, they are urged to move to the next chapter without worrying unduly over the details. This general idea, both here and in the earlier §9.4, is nothing other than extending to higher dimensions what was done in §9.2 and seen very clearly in Figures 9.1 and 9.2, where everything is very easy indeed. That we can extend this simple construction to higher dimensions is not surprising: the opposite would have been! These extensions are never difficult to visualise (see §9.2.7), but they take pages and pages of new notation....

Added in proof The footnote on page 294 very much applies to the presentation of this chapter which simplifies considerably in this new approach. On the other hand, readers familiar with topological methods may well find other ways to simplify this endgame.

10.1 An Overview and a Guide for the Reader

10.1.1 Notation and the previous constructions

We shall consider throughout this chapter a simply connected soluble Lie group that satisfies the C-condition and is of the form $Q = N \leftthreetimes V$, where N is nilpotent and $V = \mathbb{R}^{d-1}$ is a Euclidean space with $d \geqslant 2$ and $\dim N = n$, $\dim Q = n + d - 1 = p$. The left-invariant Riemannian metric will be considered throughout. In Chapter 9 we constructed mappings

$$f\colon \partial \square^d \longrightarrow S = \operatorname{Image} f \subset Q \tag{10.1}$$

and we shall recapitulate here some of the key properties of these constructions.

The function f is Lipschitz,

$$f \in \operatorname{Lip}(C(\log R)^c); \quad R > C, \tag{10.2}$$

for fixed constants and a parameter R that will be made to tend to ∞. With the notation of §§7.5.1 and 9.3.2 we can then write

$$S \text{ is an } \operatorname{LL}(R) - \partial \square^d \subset Q. \tag{10.3}$$

We shall use the translation of Q and assume, as we may, that e, the identity of Q, lies on S, and this implies that S lies in a ball of Q centred at e with radius $C(\log R)^c$. Furthermore, since $\partial \square^d$ is bi-Lipschitz homeomorphic with the unit sphere $S^{d-1} \subset \mathbb{R}^d$, whenever this is convenient the mapping (10.1) will be replaced by

$$f\colon S^{d-1} \longrightarrow S \subset Q; \quad f \in \operatorname{Lip}(C(\log R)^c). \tag{10.4}$$

For the critical transversality condition of this mapping we refer back to §9.3.3, and forward to §10.1.4.

Example 10.1 Here $Q = G_r = \mathbb{R}^r \leftthreetimes \mathbb{R}^{r-1}$ with $d = r \geqslant 2$, $p = 2r - 1$. This is as in §§9.2–9.4.

Example 10.2 Here $Q = \mathbb{H} \leftthreetimes \mathbb{R}$ where $\mathbb{H}$ is the Heisenberg group as in §9.2.5. We then have $d = 2$ and $p = 4$.

For the SSG of §9.5, $Q = N \leftthreetimes (A' \oplus A)$ and $A' = \mathbb{R}^{r-1}$, $A = \mathbb{R}^s$, $V = A' \oplus A$ and therefore $d = r + s$, $p = n + d - 1$. When $A = \{0\}$ and $s = 0$ we can use the constructions for the above two examples and then f takes its values in the subgroup $N' \leftthreetimes A'$ (with the notation of §9.5).

Example 10.3 We take the Abelian alternative in §9.5 with $s \geqslant 1$; the construction of f and S was carried out under the assumptions $\mathscr{F}_r, \ldots, \mathscr{F}_{r+s-1}$ on Q.

Example 10.4 We take the Heisenberg alternative in §9.5.6; then $r = 2, s \geqslant 1$ and the construction was carried out under the assumptions $\mathscr{F}_2, \ldots, \mathscr{F}_{s+1}$ on Q.

10.1.2 General soluble simply connected C-groups and their 'rank'

The importance of Examples 10.1–10.4 lies in the fact that up to polynomial homotopy equivalence they cover the general case. More precisely, let G be an arbitrary simply connected soluble C-group and let $N \lhd G$ be its nilradical. We set

$$\rho = \dim \left(\frac{G}{N} \right) = \text{rank } (G). \tag{10.5}$$

(This terminology for the rank is used mostly in the realisation of symmetric spaces as soluble Lie groups; see Helgason, 1978, Chapter 6.) We saw in §§8.3.4 and 9.5 that there then exists an SSG Q as in the above examples for which $G \simeq Q$ is polynomially homotopically equivalent. Furthermore, with the notation of §10.1.1 and (10.5) we can choose Q so that $d - 1 = \dim V = \text{rank}(G)$ (see §8.5.1.3).

10.1.3 Coordinates, distances and Riemannian metrics

We shall denote by $\pi \colon Q \to V$ the canonical projection and by

$$g = (x_1, \ldots, x_n, y_1, \ldots, y_{d-1}) \in \mathbb{R}^p; \quad g \in Q \tag{10.6}$$

the exponential coordinates of the second kind (see §7.3.1) where the x_i are the coordinates on N and the y_i are the Euclidean coordinates of V. With these coordinates $\pi(g) = (y_1, \ldots, y_{d-1})$, and we shall denote

$$U = U_a = \pi^{-1}(V_a), \quad V_a = [v \in V;\ |v| < a]; \quad 0 < a < 10^{-10}. \tag{10.7}$$

We can identify U with

$$U = U_a \simeq N \times V_a. \tag{10.8}$$

We shall consider the following distances and Riemannian structures on U.

(i) The Riemannian norm $|\ |_1$ on TU induced the Riemannian norm of Q.
(ii) The product Riemannian norm of $|\ |_N$, the left-invariant Riemannian norm on N and the Euclidean norm on V. This product norm is denoted by $|\ |_2$.
(iii) We have $|\ |_1 \approx |\ |_2$ on U. This follows from Exercise 8.6.

Given that N is nilpotent it also follows (see §§7.3 and 2.14) that on U the Riemannian distance d_1 induced by the above Riemannian norms is polynomially distorted with respect to the Euclidean distance d induced by the coordinates (10.6); that is, $d_1 \leqslant C(1+d)^C$ and vice versa.

10.1.4 The transversality condition on the mapping (10.1), (10.4)

Here a in (10.8) is assumed to be sufficiently small and then

$$U_a \cap S = D_1 \cup \cdots \cup D_m \tag{10.9}$$

breaks up into m disjointed pieces as explained in the transversality condition of §9.3.3. In particular, $\pi\colon D_j \to V_a$ is bijective and for the distance d we have

$$d(D_i, D_j) \geqslant R^c; \quad R \geqslant C, \tag{10.10}$$

for appropriate constants. The exact value of m is 2^r in the Abelian case (where r is as in Examples 10.1 or 10.3; see also §9.5.7) and $m = 6$ in the Heisenberg case, but for our purposes all that counts is that $m \geqslant 2$. Furthermore, for $j = 1, \ldots, m$, the mappings

$$f\colon A_j \longrightarrow D_j; \quad A_j = f^{-1}(D_j) \subset \partial \square^d \text{ or } S^{d-1} \\ \text{as may be the case in (10.1) or (10.4),} \tag{10.11}$$

are bijective and Lipschitz with Lipschitz constants bounded by $C(\log R)^c$ for fixed constants. (For the Lipschitz condition we use the Euclidean structure on $\square^d$ and the Riemannian structures of §10.1.3.) By its construction the mapping (10.1), (10.4) has considerably more structure and is piecewise affine (see §§9.4 and 9.5). This additional structure will be exploited in an essential way in the constructions below in §10.2 and §10.3.

10.1.5 The filling function and the key proposition

To fix ideas we shall adopt here the version of f given in (10.4) with f defined in the unit sphere. As in §7.5, a filling function will then be some Lipschitz function

$$F\colon B^d \longrightarrow Q; \quad B^d = \left[x \in \mathbb{R}^d;\ |x| < 1\right],\ F|_{S^{d-1}} = f,\ S^{d-1} = \partial B^d. \tag{10.12}$$

Several such extensions clearly exist.

For the next proposition we recall that the Lipschitz norm of the functions F, f is defined to be

$$\|F\|_{\text{Lip}} = \inf A \text{ such that } d\,(F(x), F(y)) \leqslant A|x-y|, \tag{10.13}$$

where the distance is given by the left-invariant metric on Q. The aim of this chapter is to prove the following result.

Proposition 10.5 *Let Q, f, F be as above. Then the Lipschitz norm of F satisfies*

$$\|F\|_{\text{Lip}} \geqslant cR^c; \quad R \geqslant C, \tag{10.14}$$

for some constants $C, c > 0$ that depend only on Q.

By property (10.2) on f it follows from (10.14) that as long as it is possible to make a construction of f as above then property $\mathscr{F}_d$ breaks down in the group Q (see §7.5).

10.1.6 Deducing the C-theorem (Theorem 7.11)

We shall apply the proposition to Examples 10.1–10.4 considered earlier, where we recall that $d-1 = \dim V$ and that for the construction in Examples 10.3 and 10.4 we had assumed that properties $\mathscr{F}_2, \ldots, \mathscr{F}_{d-1}$ hold. The conclusion from the proposition is therefore that in all these examples $\mathscr{F}_d$ breaks down. Therefore, finally, in all four examples it is not possible that $\mathscr{F}_2, \ldots, \mathscr{F}_d$ simultaneously hold.

The C-theorem (Theorem 7.11) follows from this because of the polynomial homotopy equivalence of §8.5.1.3. The following more precise result in fact follows.

Theorem 10.6 (Theorem C) *If Q is some soluble simply connected* C-*group then at least one of the properties*

$$\mathscr{F}_2, \mathscr{F}_3, \ldots, \mathscr{F}_{\rho+1}; \quad \textit{and} \quad \rho = \text{rank}(Q) \tag{10.15}$$

breaks down.

Open Problem 10.7 Given Q as in the theorem it remains an interesting open problem to specify exactly those indices j for which $\mathscr{F}_j$ breaks down. This problem was seen in §9.5.2, though there it was not described exactly which (or possibly both) of the two properties $\mathscr{F}_r$ or $\mathscr{F}_{r+1}$ breaks down.[1]

[1] **Added in proof:** This type of problem, or rather the corresponding version in homology theory, is what the Epilogue at the end of the book is all about.

10.1.7 Guide for the reader

Two different proofs will be given for Proposition 10.5. The first, in §10.2, is essentially an application of Stokes' theorem combined with the transversality of §10.1.4. This proof will be given in §10.2.1 under the additional condition that f, F are smooth (in fact they will be assumed to be embeddings). This proof is straightforward and the reader should definitely start with it. The best way to get round the smoothness is to use the notion of currents where Stokes' theorem amounts to the definition of the boundary operator on currents. This is done in §10.2.2 and for the convenience of the reader we shall there recall in an informal way the definitions and properties of currents that are needed. The advantage of this way of using currents is that it is light, but like every informal approach it leaves something to be desired. For readers who wish to have more information on currents we give a guide to the literature on the subject in §10.2.6.

At any rate the reader who is not happy with this attitude can fall back to the second proof of the proposition that is based on ideas from differential topology. This second proof is carried out in §§10.3–10.4 and we give a brief description in the next subsection.

10.1.8 The second approach based on smoothing

When $d = 2$ in (10.4) the smoothing is exactly what we see in Figures 9.1 and 9.2, where a circle $S \subset \mathbb{R}^3$ is smoothly placed in $\mathbb{R}^3$, albeit in a 'twisted' manner. The aim is to do the same thing with higher-dimensional spheres $S^{d-1} \subset \mathbb{R}^{2d-1}$ by smoothing out the constructions of Chapter 9. This smoothing uses standard tools from differential topology and is not difficult, although in places it becomes elaborate to describe. This is very definitely stuff that the reader should skip in a first reading.

The notation is as in the previous subsection and f is as in (10.4) and F as in (10.12). By modifying a posteriori the constructions of f and F we shall show that together with all the previous properties in §§10.1.1–10.1.4 they also satisfy the following:

(i) F, f are C^∞,
(ii) $f\colon S^{d-1} \to Q$ and $F\colon B^d \to Q$

are embeddings of manifolds (the second is a manifold with boundary).

This means that these mappings are one-to-one and that their differentials are injective (see Hirsch, 1976 for more details).

(iii) For all $x \in A_j$ of (10.11), with $1 \leqslant j \leqslant m$, the differential $d\pi \circ df$ of the composed mapping $\pi \circ f$ is surjective (i.e. onto) at x.

The regularisation needed to achieve (i) and (iii) is elementary if slightly technical. This is done in §10.3.

The procedure for achieving (ii) is much deeper and it uses the Whitney approximation theorem (see Hirsch, 1976, Chapter 2) and can only be carried out if we assume in addition that

$$p \geqslant 2d + 1. \tag{10.16}$$

Condition (10.16) necessitates in Examples 10.3 and 10.4 large values of n and is for instance never satisfied in Examples 10.1 and 10.2. In the general case, to get round (10.16) an ad hoc device of jacking the dimension p of Q has to be used. This will be done in §10.4 where we shall also complete the proof with the use of Sard's theorem and the use of transversality from differential topology.

It is also clear that once we have achieved the smoothness of the mappings we can fall back on Stokes' theorem in its smooth version in §10.2.1 and we can thus avoid the use of currents.

10.2 The Use of Stokes' Theorem

10.2.1 The smooth case

In this subsection we shall give a proof of Proposition 10.5 under the assumption that F gives a C^∞ embedding of the manifold with boundary B in Q. We recall that this means that F is one-to-one and an immersion (i.e. dF is one-to-one; see Hirsch, 1976 and §10.1.5).

We shall use, systematically, the notation of §§10.1.1–10.1.5 and the coordinates (10.6) for $Q = N \leftthreetimes V$ where $X = (x_1, \dots, x_n)$ are the coordinates of N and $Y = (y_1, \dots, y_{d-1})$ are the coordinates of V. We thus identify Q with $\mathbb{R}^p$ and we shall consider a smooth differential form with support in $U_{a/2}$ in (10.7) with $a > 0$ small enough. With the above coordinates this differential form will be of the form

$$\omega = \varphi(g)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{d-1}; \quad g = (X, Y) \in Q. \tag{10.17}$$

Stokes' theorem will then be used and with the notation of (10.9) and $B = F(B^d)$ we have

$$\int_B d\omega = \int_S \omega = \int_{S \cap U_a} \omega = \sum_j \int_{D_j} \omega. \tag{10.18}$$

The projection $\pi\colon Q \to V$ can then be used to identify each D_j with the V_a of (10.8). An orientation on $\mathbb{R}^d$ and on V will be fixed throughout. From the orientation on $B^d \subset \mathbb{R}^d$ we obtain an orientation on S^{d-1} as the ∂B^d and therefore an orientation of each D_j. We shall set $\varepsilon = +1$ if π preserves the above orientations and $\varepsilon = -1$ if it reverses these orientations.

We shall now fix $0 \leqslant \varphi_0 \in C_0^\infty(V_{a/2})$ and define φ in (10.17) in such a way that

$$\varphi(g) = \varepsilon\varphi_0(Y); \quad g = (X,Y) \in D_j. \tag{10.19}$$

For such a choice, provided that φ_0 has been chosen appropriately, we clearly have

$$\left|\int_S \omega\right| = m\left|\int_{V_a} \varphi_0\, dy\right| \geqslant 1. \tag{10.20}$$

Definition (10.19) specifies the values of the C_0^∞ function φ in (10.17) on the sets D_j. It is possible to extend this definition to some $C_0^\infty(Q)$ function, and if we take (10.10) into account, this extension can be done in such a way that the corresponding extension of the form (10.17) satisfies

$$|\mathrm{d}\omega|_1 \sim |\mathrm{d}\omega|_2 \leqslant CR^{-c}; \qquad R \geqslant C, \tag{10.21}$$

for some appropriate choice of the constants. Here the Riemannian norms $|\ |$ of §10.1.3 are used to give norms $|\mathrm{d}\omega|$ on differential forms. Towards that, we shall choose $\varphi(g) = \varphi_1(n)\varphi_2(v)$, for $g = nv, n \in N, v \in V$, with $|\mathrm{d}\varphi_1|_N \leqslant CR^{-c}$ which is possible by (10.10), and we then have

$$\|\mathrm{d}\omega\|_\infty = \sup_{g\in Q}|\mathrm{d}\omega(g)|_1 \leqslant CR^{-c}, \tag{10.22}$$

for some new constants $C, c > 0$, provided that $\operatorname{supp}\varphi_2 \subset V_a$.

From this we conclude that for an appropriate choice of constants we have

$$\left|\int_B \mathrm{d}\omega\right| \leqslant \|\mathrm{d}\omega\|_\infty \operatorname{Vol}_d B \leqslant CR^{-c} \operatorname{Vol}_d B; \quad R \geqslant C, \tag{10.23}$$

where here the d-dimensional volume for the smooth submanifold B is measured with respect to the Riemannian norm $|\ |_1$ on Q as in §10.1.3(i). Combining (10.18), (10.20) and (10.23), we conclude that

$$\operatorname{Vol}_d B \geqslant CR^c; \quad R \geqslant C, \tag{10.24}$$

for appropriate constants. This clearly implies that the Lipschitz constant of

$$F\colon B^d \longrightarrow B \subset Q \text{ satisfies } \operatorname{Lip} F \geqslant CR^c; \quad R \geqslant C, \tag{10.25}$$

for some constants and we have a proof of (10.15) and of the proposition in this special smooth case.

10.2.2 Stokes' theorem for the general case. The use of currents

The notation is as in §§10.1.1–10.1.4 but now it is preferable to use the definition in (10.1) and consider $S = f(\partial \square^d)$ where f is assumed to be Lipschitz and satisfying the conditions of §§10.1.1–10.1.5.

For F an extension of f to $\square^d$ as in §10.1.5 and $B = F(\square^d)$, the proof of Proposition 10.5 will follow identical lines. We shall construct some appropriate smooth differential form ω for which

$$\int_S \omega \geqslant 1, \qquad \int_S \omega = \int_B \mathrm{d}\omega. \tag{10.26}$$

Here B, S are not hypersurfaces and the above integrals, as well as the fact that the boundary of B is S, have to be interpreted in the sense of *currents*. This we shall explain in the next few lines.

The space of currents $\bigwedge^*$ is by definition the dual space of $\bigwedge T^*Q$, the space of compactly supported smooth differential forms assigned with the C^∞ topology on compact sets. (This is an inductive limit topology; see the guide to the literature in §10.2.6 below.)

Since f, F are Lipschitz we can pull back ω and we obtain $f^*\omega$, $F^*\omega$, which are differential forms with L^∞ coefficients. Being slightly pedantic here, in the definition of $F^*\omega$ observe that F extends to some Lipschitz function in some neighbourhood of $\square^d$ (see Federer, 1969, §2.10.43). Two linear forms on $\bigwedge T^*Q$ can thus be defined by

$$\langle [S], \omega \rangle = \int_{\partial \square^d} f^*\omega, \quad \langle [B], \theta \rangle = \int_{\square^d} F^*\theta; \quad \omega, \theta \in \wedge T^*Q. \tag{10.27}$$

This is the interpretation of the two integrals $\int_S$ and $\int_B$ in (10.26).

The above currents are special: they are integration currents. These are linear forms on the space of compactly supported differential forms with continuous coefficients. For any current T we can define (possibly $+\infty$)

$$\|T\| = \sup\left[|\langle T, \theta \rangle|;\ \theta \in \textstyle\bigwedge T^*Q,\ \|\theta\|_\infty \leqslant 1\right]. \tag{10.28}$$

In (10.28) we set, as before,

$$\|\theta\|_\infty = \sup\left[|\theta(g)|_1;\ g \in Q\right] \tag{10.29}$$

for the Riemannian norm on T^*Q.

If the currents in Q are written as differential forms with coefficients that are distributions, then the integration currents are the ones for which the coefficients are measures: see the guide to the literature in §10.2.6 below; note that in Federer (1969, §§4.1.5–4.1.7), these are called 'currents representable by integration', and that the norm (10.28) is denoted by M(T). Note also that we have to be careful when we say that currents are 'differential forms with coefficients that are distributions'; this is standard practice in the literature but is an abuse of terminology. For the formally correct way of seeing things see de Rham (1960, §8).

10.2.3 The boundary operator on currents

The boundary operator b on $\bigwedge^*$ is defined by (see de Rham, 1960)

$$\langle bT, f\rangle = \langle T, \mathrm{d}f\rangle; \quad T \in \bigwedge^*, \ f \in \bigwedge T^*Q. \tag{10.30}$$

If T is represented as a differential form with distributional coefficients, and at least when T is 'homogeneous', we have $bT = \pm\mathrm{d}T$ where the derivatives are taken in the sense of distributions and the $\pm$ depends on the degree of T (see the literature survey in §10.2.6), then, with that terminology, the second formula in (10.26) says that

$$b[B] = [S]. \tag{10.31}$$

The analogous formula $b[\square^d] = [\partial\,\square^d]$ in $\mathbb{R}^d$ is the classical Stokes' theorem.

The dual mappings F_*, f_* of the pullback mappings F^*, f^* give the direct images of currents and the definition (10.27) can be rephrased as

$$F_*\left[\square^d\right] = [B], \qquad f_*\left[\partial\,\square^d\right] = [S]. \tag{10.32}$$

Equation (10.31) therefore says that for the *normal* current $T = [\square^d]$ on $\mathbb{R}^d$ and the Lipschitz function φ we have

$$\varphi_* bT = b\varphi_* T. \tag{10.33}$$

An integration current with compact support is called *normal* if bT is also an integration current (see §10.2.6 for more information).

Notice that additional conditions are needed for (10.33) to hold. Observe for instance that if T is a general integration current, $\varphi_* T$ cannot a priori be defined unless the coefficients of T as a differential form are L^1 functions. This is because for the pullback of a smooth differential form ω the coefficients of $\varphi^*\omega$ are a priori, L^∞ functions (see Federer, 1969, as explained in §10.2.6 below).

10.2.4 The general proof of Proposition 10.5

We shall apply (10.26) to the differential form defined in (10.17), (10.19),

$$\omega = \varphi(X,Y)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{d-1}; \quad \varphi = \varepsilon\varphi_0(Y), \tag{10.34}$$

but here some care is needed in the definition of the orientation ε and φ_0 because since f is only Lipschitz it is not a priori clear how it transports the orientations from $\partial \square^d$ to V.

The easiest way to handle this difficulty is to suppose that the support of φ_0 lies in Ω which is some very small ball very close to 0. By the piecewise linear structure of the mapping f_V in the definition of f (see §9.4), this Ω can be chosen in such a way that in the correspondence between $\partial \square^d$ and V by the two mappings f and π of §10.1.4, the sets Ω_j that correspond to Ω lie in the interior of simplexes σ of dimension $(d-1)$ of the simplicial decomposition of the boundary $\partial \square^d$ (see §9.4.1). To clarify this further, consider Example 10.1. Then the simplex σ is one of the simplexes of $\mathscr{E}_{r-1}(P)$, for some vertex P, of §9.4.2. What we are then saying is that $(\pi \circ f)^{-1}(\Omega) \subset \sigma$. An analogous definition can be given for all the other examples of §10.1.1. The mapping $\pi \circ f$ can thus pull back the differential form $\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{d-1}$ and the orientation from V to $\partial \square^d$. Therefore the first formula in (10.26) is guaranteed exactly as in (10.20).

We can, on the other hand, extend the definition of ω as in §10.2.1 with $\|\mathrm{d}\omega\|_\infty \leqslant CR^{-c}$ for the same reasons and what replaces (10.23) is that for appropriate constants the following estimate holds (see (10.28)):

$$\left| \int_B \mathrm{d}\omega \right| \leqslant \| \, [B] \, \| \; \|\mathrm{d}\omega\|_\infty \leqslant CR^{-c} \| \, [B] \, \|. \tag{10.35}$$

The estimate

$$\| \, [B] \, \| \geqslant CR^c \tag{10.36}$$

follows for appropriate constants. On the other hand, by the definition of $[B]$ we have

$$\| \, [B] \, \| \leqslant C(\operatorname{Lip} F)^d \tag{10.37}$$

for some constant and the Lipschitz norm of F in §10.1.5. The conclusion from (10.36), (10.37) is

$$\operatorname{Lip} F \geqslant CR^c; \tag{10.38}$$

hence the proof of the proposition.

Remark 10.8 We have been unduly careful with choosing the small ball Ω and the $\operatorname{supp} \varphi_0 \subset \Omega$. In fact any $\varphi_0 \geqslant 0$, with support in $V_{a/2}$ as in the smooth case, would have done just as well for the first formula in (10.26). This follows from the following.

Exercise 10.9 For simplicity let us stick with Example 10.1. In the definition of $f_V : \partial \square^d \to V$ in §§9.4.5–9.4.6, this mapping is affine when restricted to the simplexes of dimension $(d-1)$ as long as we stay close to P, where P is any vertex of $\partial_0 \square^d$. Now let us fix some orientation of $\mathbb{R}^d$. This induces an orientation on $\partial \square^d$ and therefore on all the above simplexes. We can also fix some orientation of V and so what has to be verified is that f_V either respects or reverses these orientations (simultaneously) for all these simplexes that contain P, and which of the two happens depends only on P. The reader is urged to think about this. Writing down the details of this is long but, fortunately, as we saw, we do not *have* to use this exercise. From this exercise it clearly follows that the $\varepsilon = +1$ or -1 and $\varphi = \varepsilon \varphi_0$ can be chosen as in (10.19).

10.2.5 Slicing of currents and yet one more proof of Proposition 10.5

The notion of slicing of currents is subtler and far less standard (cf. §10.2.6). However, if one is willing to use it, one can present an alternative way of formulating the proof given in §10.2.4. This approach is interesting and in particular it avoids the questions of orientation that we had to consider in §§10.2.2 and 10.2.4.

We shall give only an informal (in fact heuristic!) description of this approach because we feel that very few readers know or are willing to learn what the slicing of currents really means. The reader who wishes to pursue the matter further should study first the literature given in §10.2.6.

Heuristic description of slicing When $\pi \colon U = \mathbb{R}^p \to \mathbb{R}^{d-1} = V$ in the canonical projection every absolutely continuous measure $d\mu = F\,dx$ on U can be disintegrated by Fubini as an integral on V of $L^1(\pi^{-1}(v))$ functions on the fibres of π, for $v \in V$. The same thing is possible for currents of dimension m (i.e. the dual of $\bigwedge_m T^*U$ the differential forms of degree m) provided that $m \geqslant \dim V = d-1 = n$. Here the example that is relevant is the current $[B]$ of §10.2.3 where $m = d$.

The currents for which this can be done are the integration currents T for which ∂T is also an integration current. These currents are called 'normal'. In that case, for almost all $v \in V$ we can define a new current on U, called the *slice* on $\pi^{-1}(v)$, which we denote by $\langle T, \pi, v\rangle$. This is also an integration current that is supported on $\pi^{-1}(v)$ and has dimension $(m-n)$. For this the following Fubini-type of disintegration holds:

$$T = \int_V \langle T, \pi, v\rangle \, dv. \tag{10.39}$$

A correct interpretation is needed for this formula to make sense. Observe for instance that in a naive interpretation the degrees of the two sides of (10.39) do not match! What counts, and this time the formulas are formally correct, is that when $m > n$ we have

$$\|T\| \geqslant \int \|\langle T, \pi, v\rangle\| \, dv, \qquad \partial\langle T, \pi, v\rangle = (-1)^n \langle \partial T, \pi, v\rangle. \tag{10.40}$$

The informal proof of Proposition 10.5 that uses the slicing We use the notation of §10.2.2, apply the slicing (10.39) and (10.40) to $T = [B]$, $\partial T = [S]$, and write $B_v = \langle [B], \pi, v\rangle$, $S_v = \langle [S], \pi, v\rangle$ with $\pi\colon Q \to V$. Here $\dim S_v = 0$, and for $v \in V$ close enough to zero, S_v consists of m points at distances $\approx R^c$ apart (m and R are as in §10.1.4 and (10.9)).

This means that we can argue on each separate slice $\pi^{-1}(v) = N_v$ separately and construct $\varphi \in C_0^\infty(N_v)$ which is a scalar function this time for which

$$\|d\varphi\|_\infty = \mathrm{O}(R^{-c}), \qquad \langle S_v, \varphi\rangle \geqslant 1. \tag{10.41}$$

On the other hand, since by (10.40)

$$\langle B_v, d\varphi\rangle = \langle S_v, \varphi\rangle, \tag{10.42}$$

we deduce that $\|B_v\| \geqslant CR^c$ for appropriate constants. By (10.40) this implies that $\|[B]\| \geqslant CR^c$, which is the required conclusion in (10.36). This finishes the outline of this approach.

One reason why this formulation of the proof is interesting is that it runs in complete analogy with the endgame that we shall play in §10.4 below. In §10.4.2 we use Sard's theorem and transversality from differential topology, while here we use the slicing of currents from geometric measure theory.

10.2.6 Guide to the literature on currents

The classical book on currents is de Rham (1960). One only needs to browse through the first part of the book to get a clear picture. It is very readable and the old-fashioned style in which it is written is I feel an advantage.

To get more precise information on the topological vector spaces aspect, one would have to consult Schwartz (1957) also. In Chapter 12 we shall need quite a bit of information on those topologies and then more specialised literature is needed. We shall use Bourbaki (1953) and Grothendieck (1958) for this.

Coming back to currents, unfortunately what one finds in de Rham (1960) is not the end of the story for us. The integration currents that are our main object of interest can be found in Federer (1969) and are part of geometric measure theory. The subject is technical and even pinpointing the exact references that are needed in this book is not an easy task. Starting to read it from scratch to find each reference as needed is very time consuming; in the next few lines a guide will be given as to what to look for.

Federer (1969, Chapter 4) is the relevant chapter for us and in fact only §4.1.1–§4.1.19 are really needed. The definitions of *normal* currents are in §4.1.7. Our formula (10.33) can be found in §4.1.14. This formula is explicitly stated in the second half of that section. The important point that has to be understood is that we can take *direct images of normal currents by Lipschitz functions*. This is explained in the first part of §4.1.14. For this the notion of *flat seminorms* (see §4.1.12) and a specific homotopy formula (see §4.1.13) have to be used.

For the slicing and the relevant formulas that we use in §10.2.5 one has to look at §4.3.1 and §4.3.2 and here §4.1.18 is needed.

The hope is that the above indications will help the reader to inform themselves on these important notions.

10.3 The Smoothing of the Mapping $f : \partial\,\Box^d \to Q$

10.3.1 An overview

As already explained in §10.1.8, for the second approach to the C-theorem we shall need that the mapping given in (10.1) and (10.4) is smooth. The purpose of this section is to show how we can start from the Lipschitz mappings as defined in §10.1.1 and then modify them and smooth them out while preserving their essential properties.

As usual, the smoothing is done by convolution with a mollifier but the problem here is that one usually loses the bijectivity as soon as convolution is used. These mappings therefore have to be modified first and be 'linearised' near the vertices of the cube $\partial_0\,\Box^d$. This is the reason for the combinatorial considerations that we start with in §§10.3.2–10.3.4. For the first basic construction of §§9.3–9.4 the linearization is done in §10.3.5. For the second basic construction of §9.5 the linearization is done in §10.3.7. Once this is achieved the rest of the smoothing procedure is straightforward. This linearization in the first part of the smoothing consists of an ad hoc elementary argument that is, however, long to write down. The reader should first concentrate on understanding what is happening rather than worrying too much about the details (see the informal recapitulation at the end of §§10.3.5 and 10.3.6). In the exercise of §10.3.6 the need for that linearization will become clear.

Basic non-trivial ideas from differential topology that are not elementary will be needed to guarantee that f is in addition an embedding $f\colon S^{d-1} \to Q$

(see §10.1.8). This, however, will be postponed until §10.4 where the final endgame is also played out.

10.3.2 Simplexes revisited. Their canonical position in affine space

We shall denote, throughout, $V = \mathbb{R}^n$ a real finite-dimensional vector space. We shall consider $\sigma = [v_0, \dots, v_n] \subset V$ a non-singular simplex, that is, we assume that $\operatorname{Int}\sigma = \mathring{\sigma}$ (see §9.1.3) and that $0 \in \operatorname{Int}\sigma$. We shall denote the basis vectors of V by $e_j = (0, \dots, 0, 1, 0, \dots, 0)$, with $1 \leqslant j \leqslant n$, and define the linear transformation $A \in \mathrm{GL}(V)$ by $Av_j = e_j$. Since $0 \in (A\sigma)^\circ$ lies in the interior of the image, the first vector Av_0 has to lie in the negative quadrant and there exist scalars λ_j such that

$$Av_0 = e_0 = (\lambda_1, \dots, \lambda_n) \in V; \quad \lambda_j < 0,\ j = 1, \dots, n. \tag{10.43}$$

In what follows, the image simplex $A\sigma = [e_0, \dots, e_n]$ will be called the canonical position of σ. The A is clearly unique if, as we always do, we fix the order of the vertices of the simplex.

We shall identify, in what follows, V with its dual space V^* by the scalar product induced by the orthonormal basis $e_1, e_2, \dots, e_n$. We shall illustrate the canonical position of σ by constructing the dual simplex $\sigma^* = [\sigma_0^*, \dots, \sigma_n^*] \subset V^*$ that we used in §9.4.6. We recall that the non-singular simplex $\sigma^* \subset V^*$ is called a dual simplex if $0 \in \operatorname{Int}\sigma^*$ and if

$$\langle \sigma_i, \sigma_j^* \rangle > 0; \quad i, j = 0, 1, \dots, n,\ i \neq j. \tag{10.44}$$

Example 10.10 The simplex $\sigma = [\sigma_0, \sigma_1, \sigma_2,] \subset \mathbb{R}^2$ is a triangle in the plane and, if we denote by σ_j^* the orthogonal projection of σ_j on the face opposite σ_j, we have a dual simplex.

Rather than trying to generalise this example in higher dimensions we shall assume that $\overline{\sigma} = [e_0, e_1, \dots, e_n]$ is in the canonical position (10.43). Then the simplex $\overline{\sigma}^* = [e_0^*, \dots, e_n^*]$ given by

$$e_0^* = (1, 1 \cdots 1), \quad e_j^* = -e_j + ae_0^*;\ j \neq 0 \tag{10.45}$$

is a dual simplex provided that $a > 0$ is small enough. The verification of (10.44) for this definition is immediate.

For a general simplex σ as above let $A \in \mathrm{GL}(V)$ be the linear transformation that brings it to its canonical position $\overline{\sigma}$. Then $\sigma^* = A^* \overline{\sigma}^*$ for the adjoint transformation is a dual simplex.

(Verification: $\langle \sigma_j, \sigma_k^* \rangle = \langle A^{-1} e_j, A^* e_k^* \rangle = \langle e_j, e_k^* \rangle$.)

10.3.3 The tessellation of §9.4.5 revisited

Let $V = \mathbb{R}^n$ and σ be as in the previous subsection. For every subset $I \subset [0,1,\dots,n]$ of length $1 \leqslant |I| \leqslant n$ and any set of positive scalars $\lambda_i^{(I)} > 0$ $(i \in I)$ we shall denote, as in §9.4.5,

$$e_I = \sum_{i\in I} \lambda_i^{(I)} v_i \in V; \tag{10.46}$$

no confusion should arise between these e_I and the basis vectors of the previous subsection. Similarly to (9.81), for any increasing chain of subsets

$$\mathscr{J} : J_1 \subset \cdots \subset J_n = [0,1,\dots,n] \tag{10.47}$$

with $|J_i| = i$, for $1 \leqslant i \leqslant n$, we can consider the conical domain

$$C_{\mathscr{J}} = \mathrm{CC}(e_{J_1},\dots,e_{J_n}) \subset V.$$

These give the required tessellation of V in §9.4.5.

If we consider now $\sigma' = [v'_0,\dots,v'_n]$ and other similar simplexes and vectors

$$e'_I = \sum_{i\in I} \mu_i^{(I)} v'_i; \quad \mu_i^{(I)} > 0,\ i \in I, \tag{10.48}$$

for the same chain $\mathscr{J}$ we can define new conical domains $C'_{\mathscr{J}} \subset V$ and a new tessellation of V. These two tessellations in §9.4.5 were linked together by a continuous, piecewise linear mapping:

$$\Psi = \Psi_{e_I,e'_I} : V \longrightarrow V; \tag{10.49}$$

it will be more convenient to write $\Psi = \Psi[e_I, e'_I]$. This mapping is characterised by the fact that $\Psi(e_I) = e'_I$ and that its restriction on each $C_{\mathscr{J}}$ of the first tessellation is a linear mapping of $C_{\mathscr{J}}$ onto $C'_{\mathscr{J}}$ of the second tessellation.

Example 10.11 When $V = \mathbb{R}$ is one-dimensional then Ψ becomes a dilation of $\mathbb{R}$ with possibly different dilating constants for the positive and negative half axes $V_\pm = [x \in \mathbb{R};\ \pm x > 0]$.

Key Example 10.12 Here we go back to the notation of §10.1.1 and Example 10.1, with $\square^r \subset \mathbb{R}^r = N$, $V = \mathbb{R}^{r-1}$. Note that $\dim V = n = r-1$ here. This example is related to the first basic construction of §§9.3–9.4.

We shall identify V with a hyperplane $V \subset N$ such that $V \cap \square^r = \{0\}$, that is, this hyperplane intersects the cube $\square^r$ only at $0 \in \partial_0 \square^r$. We shall denote by $\kappa\colon N \to V$ the affine mapping that is defined by the radial projection from the centre of the cube $(= (1/2,\dots,1/2))$ onto V. When ξ_F is the middle point of the face $F \subset \square^r$ of the cube that corresponds to the index $I \subset [1,\dots,r]$ and contains $0 \in F$ (here $|I| \leqslant r-1$ and we are using the notation of §9.4.1), we shall write $e_I = \kappa(\xi_I) \in V$. This notation is compatible with that in (10.46) when $v_i = \kappa(e_i)$ where now $e_1, \ldots, e_r$ are the basis vectors of $\mathbb{R}^r = N$. These vectors e_i can be identified with vectors along the edges of $\square^r$ (i.e. in $\partial_1 \square^r$) that contain 0.

With these e_I, in our definition of Ψ in (10.49) we shall take $e'_I = \zeta_I \in V$ as in (9.87). The mapping f_V defined in §§9.4.3–9.4.6 coincides then near 0 ($\in \partial_0 \square^r$) with

$$f_V = \Psi \circ \kappa; \quad \Psi = \Psi\left[e_I, e'_I\right]; \tag{10.50}$$

a similar identification can be made for all the other vertices in $\partial_0 \square^r$.

10.3.4 The linearization 'lemma' of Ψ near the origin

Here we shall present a technical lemma that in the next subsection will be used in the smoothing of the function f_V of §§9.4.3–9.4.6 and (10.50). The e_I, e'_I and $\Psi = \Psi[e_I, e'_I]$ are as in (10.49).

We shall modify the mapping $\Psi(v)$ so that it does not change when $|v|$ is large but it becomes a non-singular linear transformation $v \to Av$, with $A \in \mathrm{GL}(V)$, in some small neighbourhood of 0. In view of the transversality condition (see §9.3.3) we shall require from this modification to maintain the 'bijectivity near 0'.

More precisely, we shall provisionally denote by $\widetilde{\Psi}\colon V \to V$ this modification and impose the following conditions:

(i) A constant $c_1 > 0$ will be chosen and will stay fixed: say $c_1 = 10^{-100}$.
(ii) The modification $\widetilde{\Psi}$ to be constructed will satisfy $\widetilde{\Psi}(v) = \Psi(v)$ for $|v| > c_1$.
(iii) We shall show that it is possible to find some $c_3 < c_1$ and $A \in \mathrm{GL}(V)$, some non-singular linear transformation, such that the following hold:
  (iv) $\widetilde{\Psi}(v) = Av$ for $v \in B_{c_3} = [|v| \leqslant c_3]$;
  (v) (here is the crux of the construction:) for every $v \in B_{c_3}$ there exists one and only one $\omega \in V$ such that $\widetilde{\Psi}(\omega) = v$ and ω is given by $\omega = A^{-1}v$.

It goes without saying that $\widetilde{\Psi}$ is a Lipschitz mapping. A more accurate notation for this modification would have been $\widetilde{\Psi} = \Psi_A$ or even Ψ_{A,c_1,c_3}.

In the above definition of $\widetilde{\Psi}$ we have transgressed standard mathematical usage. We should have said the following: given $c_1 > 0$ there exists c_3, A and Ψ_A such that (iv) and (v) hold. We feel, however, that the way we have described the set-up is more transparent.

To see that the above construction is possible we start first from the special case when both the two simplexes σ, σ' of §10.3.3 are in canonical position (see §10.3.2). In that case we choose some $0 < c_2 < c_1$ and $\chi(v)$, with $v \in V$, some cut-off function that satisfies $\chi(v) = 1$ for $|v| < c_2$, $\chi(v) = 0$ for $|v| > c_1$ and $0 < \chi(v) < 1$ for $c_2 < |v| < c_1$. We shall then set

$$\widetilde{\Psi}(v) = \chi(v)v + (1 - \chi(v))\Psi(v); \quad v \in V \tag{10.51}$$

and show that the function will satisfy the required conditions with $A = \mathrm{I}$, the identity, and $c_3 > 0$ sufficiently small, that is, $\Psi_{\mathrm{I}} = \widetilde{\Psi}$.

To see this we consider the annulus $R = [c_2 < |v| < c_1]$. It is then clear that condition (iv) holds. Furthermore, if in condition (v) we impose the additional restriction that $\omega \notin R$ then (v) certainly holds. Now if I is some multi-index with $|I| = n$ (as in §10.3.3) and if $v \in R \cap \mathrm{CC}(v_i;\ i \in I)$ the intersection of the annulus with the convex conical domain generated by the v_i (see §9.4.5), then $\Psi(v) \in R' \cap \mathrm{CC}(v_i';\ i \in I)$ for some other annulus $R' = [c_2' < |v| < c_1']$ with $c_2' > 0$. The fact that σ, σ' are in canonical position guarantees from this that no convex combination of v and $\Psi(v)$ is close to 0 (draw a picture to see this). And this, in view of the definition (10.51) completes the proof of property (v) in this special case.

Now let $B, B' \in \mathrm{GL}(V)$ be two non-singular linear transformations. By our definition (see §10.3.3) we then have

$$\Psi[e_I, e_I'] = (B')^{-1} \circ \Psi[Be_I, B'e_I'] \circ B. \tag{10.52}$$

These two linear transformations can be chosen so that $B\sigma, B'\sigma'$ are in canonical position, where now we have started from two arbitrary simplexes σ, σ' as in §10.3.3. Formula (10.52) allows us therefore to reduce the general case to the special case that we have just treated. This completes the construction in general because we can then set $A = (B')^{-1}B$ in (iii).

10.3.5 The linearization of the Key Example 10.12

We shall go back to the mapping of §9.4.3,

$$f = (f_N, f_V) : \partial\,\square^r \longrightarrow N \leftthreetimes V = G_r, \tag{10.53}$$

of the first basic construction.

As in (10.4) the first step is to transform this into a mapping

$$f \colon S^{r-1} \longrightarrow G_r \cong \mathbb{R}^{2r-1}. \tag{10.54}$$

This is done by composing f in (10.1) with the radial projection from the centre of $\square^r$ onto the sphere S^{r-1} with the same centre as $\square^r$ that goes through $\partial_0\,\square^r$.

For x_0 a vertex on $\partial_0 \square^r \subset S^{r-1}$ and $c > 0$ small enough we write

$$B_c = B_c(x_0) = \left[x \in S^{r-1};\ |x - x_0| < c\right] \subset S^{r-1}, \tag{10.55}$$

which is a small neighbourhood of x_0, and consider the restriction of the mapping f_V on B_c,

$$f_V : B_c \longrightarrow V. \tag{10.56}$$

The same radial projection can again be used so that in (10.55), for c small enough, f can be identified with

$$f_V : T \longrightarrow V, \tag{10.57}$$

where T is a neighbourhood of size c of 0 in the tangent space $T_{x_0}(S^{r-1}) = V_{x_0}$ of S^{r-1} at x_0. As explained in Key Example 10.12, this mapping (10.57) can, in a small neighbourhood of 0, be identified with the mapping (10.49), (10.50),

$$\Psi\left[e_I, e_I'\right] : V_{x_0} \longrightarrow V. \tag{10.58}$$

This mapping can thus be modified and can be linearised as in §10.3.4. The choice of the constants $0 < c_3 < c_1$ in the 'linearization lemma' will be made in what follows and from that we can make the following construction.

There exists $c_3 < c_1$ that satisfies the conditions of §10.3.4. Furthermore, these constants can be chosen to be much smaller than the c of (10.55), (10.56). The local modification of the lemma will then be performed on f_V on $B_c(x_0)$ and this will be done for every $x_0 \in \partial_0 \square^r$.

If c is small enough, these modifications for the various vertices of $\partial_0 \square^r$ do not interfere with each other and they can be put together because the value $f_V(x)$ of the original definition does not change if x is not close to $\partial_0 \square^r$. More explicitly, we see that there exist small constants $0 < c' \ll c'' \ll 1$ such that if we denote by $\tilde{f}_V$ the modified function

$$\tilde{f}_V : S^{r-1} \longrightarrow V, \tag{10.59}$$

then $\tilde{f}_V$ is Lipschitz and satisfies

(i) $\tilde{f}_V(x) = f_V(x)$ if $d(x, \partial_0 \square) > c''$;
(ii) for each $x_0 \in \partial_0 \square^r$, $\tilde{f}_V\big|_{B_{c'}(x_0)}$ is the composition of the radial projection κ with a non-singular linear function from $T_{x_0}S^{r-1}$ to V (abusing notation we use the same notation for this radial projection as in (10.50)).

Since in (10.53) the function f_N is constant near the vertices $\partial_0 \square^r$ we can choose c'' small enough and define a modification of the global function (10.53) by setting

$$\tilde{f} = \left(f_N, \tilde{f}_V\right) : S^{r-1} \longrightarrow G_r \tag{10.60}$$

and this can be done so that $\tilde{f}$ is Lipschitz and satisfies the transversality properties of §9.3.3.

Recapitulation of the essential features

There exist $c, c^* > 0$ such that the small neighbourhoods of the vertices $\partial_0\,\square^r$ correspond in a one-to-one fashion to a small neighbourhood of 0 in V via the projection of G_r on V. Explicitly we have

$$\tilde{f}^{-1}[nv;\ n \in N,\ v \in V,\ |v| < c^*] \subset [\bigcup B_c(x_0);\ x_0 \in \partial_0\,\square^r], \tag{10.61}$$

$$\tilde{f}(B_c(x_0)) \subset \{x_0\} \times V \subset G_r; \quad x_0 \in \partial_0\,\square^r, \tag{10.62}$$

and furthermore, for each $x_0 \in \partial_0\,\square^r$, the mapping

$$\tilde{f}\colon B_c(x_0) \longrightarrow \{x_0\} \times V \tag{10.63}$$

is, modulo the radial projection κ, a non-singular affine mapping.

The final point is that this modification has not changed the Lipschitz properties of f and

$$\tilde{f}\left(S^{r-1}\right) = \widetilde{S} \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset G_r. \tag{10.64}$$

To see this recall that f_N has not been modified and therefore (10.64) amounts to the fact that there exist constants such that

$$\tilde{f}_V \in \mathrm{Lip}\left(C(\log R)^c\right); \quad R \geqslant C. \tag{10.65}$$

This is obvious because by construction the Lipschitz norm of $\tilde{f}_V$ depends only on the Lipschitz norm of f_V and c_1, c_2, A and χ in §10.3.4.

In this recapitulation it is worth observing that all this elementary but lengthy work was done for one purpose and one purpose only: namely to linearise and achieve the additional linearity property of (10.63). Having achieved this linearization we shall show in the next subsection that the function $\tilde{f}$ that has been constructed and satisfies (10.61), (10.62), (10.63) and (10.65) can be smoothed and can in addition be assumed $\tilde{f} \in C^\infty$. The proof below is ad hoc but, quite generally, any Lipschitz mapping $f\colon S^d \to G$ into some Lie group (and other more general Riemannian manifolds) can be smoothed out while keeping the Lipschitz constant under control. Exponential coordinates can be used to identify G locally with $\mathbb{R}^n$. Then a simplicial decomposition of S^d can be used to localise. The idea is simple but the details are involved because among other things the decomposition of S^d has to depend on $\|f\|_{\mathrm{Lip}}$.

10.3.6 Smoothing by convolution

We shall show here how the smoothness properly asserted at the end of the previous subsection is achieved. And to make the presentation as clear as possible we shall start from the case $d = 2$ so that then S^{d-1} can be identified with the one-dimensional torus $\mathbb{T}$ which is a group. We can therefore use ordinary convolution to regularise and approximate uniformly any vector-valued Lipschitz function $f\colon \mathbb{T} \to E$. This is done by a sequence

$$C^\infty \ni f_n = f * \theta_n \longrightarrow f; \quad \overline{\lim} \|f_n\|_{\text{Lip}} \leqslant C\|f\|_{\text{Lip}}, \tag{10.66}$$

where $\theta_n \in C^\infty(\mathbb{T})$, $\operatorname{supp}\theta_n \subset [-1/2, 1/2]$ and $\int_{\mathbb{T}} \theta_n\, dt = 1$ (see Katznelson, 1968, Chapters 1 and 2).

This convolution applied to the function $\tilde{f}$ of the previous section shows that the property $\tilde{f} \in \text{Lip}(C(\log R)^c)$ of (10.65) is maintained after regularisation. The vector space E is now $G_r = N \leftthreetimes V$ with its usual (see §9.2) vector space structure $\mathbb{R}^{2r-1}$.

The additional transversality and linearity conditions (10.61), (10.62) and (10.63) that $\tilde{f}$ satisfies show that, in addition, in this case $f_n \to \tilde{f}$ converges in the C^∞ topology (see Katznelson, 1968, Chapter 1) near the vertices of $\partial_0 \square$. In addition, near these vertices for n large enough, f_n are one-to-one and they induce embeddings of the four neighbourhoods $B_c(x_0), x_0 \in \partial_0 \square$ of (10.55) into V for c small enough. Properties (i) and (iii) of §10.1.8 are therefore satisfied by f_n if n is large enough.

Exercise Verify the above. The reason for the $\overline{\lim}$ in (10.66) is that the Euclidean structure on E and the Riemannian structure on G_r are not identical. To deal with this difficulty we use §8.3.5, which gives explicitly the Riemannian structure on G_r in terms of the Euclidean coordinates of E and this allows us to dispose with the localisation (see the end of §10.3.5) and it simplifies matters. The fact that f_n remains one-to-one near the vertices comes from (10.63) and the affine nature of the function is used.

This completes the proof for $d = 2$.

In the general case $d \geqslant 2$, S^{d-1} is not a group. But the convolution with smooth measures $\mu = \varphi\, ds$ on the orthogonal group SO_d can be defined by the action $x \to sx$ on S^{d-1}, for $x \in S^{d-1}$, $s \in \text{SO}_d$. We set

$$f * \mu(x) = \int_{\text{SO}_d} f(sx)\, d\mu(s) \tag{10.67}$$

for any reasonable function f on S^{d-1}. When $\varphi \in C^\infty$, the above convolution gives a smooth function and when $\int d\mu = 1$ and support $\mu \to$ (the identity of SO_d) we obtain the same approximating sequence as in (10.66), which has exactly the same properties.

This convolution can then be applied to the function $\tilde{f}$ of (10.59) in the general case and finishes the proof.

Exercise Fill in the details in the above argument.

Recapitulation The function $\tilde{f}$ that is finally obtained, together with the properties in the recapitulation of §10.3.5, is in addition C^∞ and satisfies the transversality property (iii) in §10.1.8.

10.3.7 Smoothing the second basic construction of §9.5

In this subsection we shall consider the Abelian case of §9.5. We recall some of the notation used there where the simply connected soluble C-group Q of §10.3.1 was instead denoted by G and

$$\begin{gathered} G = N \leftthreetimes \left(A' \oplus A\right) = N \leftthreetimes V, \quad G' = N' \leftthreetimes V \subset G, \\ A = \mathbb{R}^s, \quad A' = \mathbb{R}^{r-1}; \quad s \geqslant 0,\ r \geqslant 2. \end{gathered} \tag{10.68}$$

Here, as before, N is a simply connected nilpotent group and $\mathbb{R}^r \cong N' \subset N$ and, with the notation of §§9.2.1, 9.5.1, $G' \cong G_r \times A$. Here and throughout this section we use the notation of §9.5.

In the construction in §9.5 we consider a unit cube $\square^r \subset N' = \mathbb{R}^r$ and this was used to make the first basic construction of $S \subset G_r$. Then a new cube was considered,

$$\begin{aligned} \square^{r+s} &= \square^r \times [-1,1]^s \subset \mathbb{R}^{r+s} = N' \oplus A, \\ \square^s_* &= [-1,1]^s \subset A, \end{aligned} \tag{10.69}$$

and the original cube $\square^r$ was identified with

$$P = \square^r \times \{0\} \subset \square^{r+s}; \quad 0 = \text{ centre of } \square^s_*. \tag{10.70}$$

The set of vertices $\partial_0\,\square^r$ is then identified with the subset $P_0 \subset P$. In this construction $\square^r$ and $\square^s_*$ play different roles. We constructed in §9.5 a mapping

$$f = (f_N, f_{A'}, f_A) : \square^{r+s} \longrightarrow G \tag{10.71}$$

with the three components in N, A', A respectively. The two key conditions on f are

$$f\left(\partial\,\square^{r+s}\right) = \Omega \text{ is an } \mathrm{LL}(R) - \partial\,\square^{r+s} \subset G \tag{10.72}$$

and the transversality conditions (9.147) are satisfied in small neighbourhoods of the points of $f(P_0)$. Observe also that here P_0 is not the set of vertices of the cube (10.69); it is the set of vertices of the slice P of that cube in (10.70).

As before we shall first show how we start by smoothing out $\partial\,\square^{r+s}$ so that it becomes the unit sphere S^{r+s-1} and then we smooth out f so that it preserves the key properties and becomes a smooth function $f\colon S^{r+s-1} \to G$. The first step is to write

$$\partial\,\square^{r+s} = (\partial\,\square^{r} \times \square^{s}_{*}) \cup (\square^{r} \times \partial\,\square^{s}_{*}) = L \cup M. \tag{10.73}$$

As in §10.3.5 we shall use the radial projection to identify $\partial\,\square^{r}$ with S^{r-1} and then L becomes a cylinder with base S^{r-1} and edge $\square^{s}_{*}$, that is, $L = S^{r-1} \times \square^{s}_{*}$. We then use (10.73) to make the identification of $\partial\,\square^{r+s}$ with S^{r+s-1} in such a way that L is identified smoothly with the belt

$$S^{r-1} \times \square^{s}_{*} \subset S^{r+s-1}, \tag{10.74}$$

around the subsphere $S^{r-1} \subset S^{r+s-1}$.

This situation becomes simple to describe when $s = 1$ because what we have, in effect, is the construction of S^{r} as the suspension of the lower dimension sphere: $S^{r} = SS^{r-1}$ (see Hilton, 1953). In that suspension, S^{r-1} becomes the equator and the two components of M in (10.73) collapse to the north and south poles respectively.

In the identification of $\partial\,\square^{r+s}$ with S^{r+s-1}, P_0 in (10.70) is identified with a set

$$P_0^{S} \subset S^{r-1} \times \{0\} \subset S^{r-1} \times \square^{s}_{*} \subset S^{r+s-1} \tag{10.75}$$

where (10.74) is used. We then proceed exactly as in (10.1), (10.2) and we identify the function f of (10.71) with a function

$$f\colon S^{r+s-1} \longrightarrow G. \tag{10.76}$$

This new function is again in $\mathrm{Lip}(C(\log R)^{c})$ for appropriate constants and its image $\Omega = f(S^{r+s-1}) \subset G$ satisfies the transversality conditions (9.147) as explained in §9.5.5.

We shall use the notation of §§10.1.1–10.1.4 where $d = r+s$, $V = A' \oplus A$, $\dim V = d-1$, and $\pi\colon N \leftthreetimes V \to V$ is the canonical projection. Then the above function in (10.71), (10.76) can be rewritten

$$f\colon \partial\,\square^{d} \longrightarrow N \leftthreetimes V, \qquad f\colon S^{d-1} \longrightarrow N \leftthreetimes V \tag{10.77}$$

and the transversality condition can be formulated as in §10.1.4.

Our aim is to proceed as in §§10.3.5–10.3.6 and smooth out the function f so as to satisfy the additional conditions (i)–(iii) of §10.1.8. Given the work that has already been done, this smoothing of f is now quite easy to do.

Let us go back to the construction of §9.5 for the function f in (10.71) and recall that we first used the first basic construction to construct

$$f' = (f_{N'}, f_{A'}); \quad \partial \square^r \longrightarrow N' \leftthreetimes A' = G_r \subset G. \tag{10.78}$$

Then we obtain f by identifying $\partial \square^r$ with

$$\partial \square^r \times \{0\} \subset \partial \square^r \times \square^s_* = L \subset \partial \square^{r+s} \tag{10.79}$$

and by extending the definition of f' from $\partial \square^r$ to the whole of $\partial \square^{r+s}$. In terms of coordinates we have

$$\begin{gathered} f(\alpha,\beta) \in G; \quad \alpha \in \partial \square^r,\ \beta \in \square^s_*,\ (\alpha,\beta) \in L, \\ f(\alpha,0) = f'(\alpha) \in A' \subset V. \end{gathered} \tag{10.80}$$

Let us also recall that in the construction of §9.5, a dilation was used (see end of §9.5.3 and (9.147) versus (9.148)) to identify $\square^s_*$ with the cube

$$[-N,N]^s \subset A \subset V \text{ with } N \sim (\log R)^c. \tag{10.81}$$

Going through the construction in §9.5 it is clear that in the extension from f' to f we can take (in fact this is the natural thing to do)

$$f(\alpha,\beta) = f'(\alpha,0) + N\beta \in G' \oplus A \subset G \tag{10.82}$$

as long as $|\beta| < 10^{-10}$. Here the dilation from $\square^s_* (\subset \mathbb{R}^s) \to A = \mathbb{R}^s$ given by $\beta \mapsto N\beta$, with $\beta \in \mathbb{R}^s$, is used. This is clear enough when $s = 1$ in the modification in the construction in §9.5.3. For $s \geqslant 1$ one has to follow the way the constructions of §§9.5.4–9.5.5 were made.

The conclusion is that once we have modified and linearised f' near $\partial_0 \square^r$ in the sense of §10.3.4, formula (10.82) automatically gives a linearization of f and this mapping is seen to be affine in some small neighbourhood of the vertices $P_0^S \subset S^{r+s-1}$ of (10.75). Here we understand the term 'affine function on S^{r+s-1}' in the sense of (10.63), that is, after composition with, say, a radial projection from S^{r+s-1} onto the tangent space.

Once we have achieved this linearization we can use convolution exactly as in §10.3.6 and guarantee that f can also be taken to be C^∞ and satisfies conditions (i) and (iii) of §10.1.8. The bijectivity near P_0^S of the function after convolution is seen as in the first exercise of §10.3.6. We shall omit the details.

10.3.8 The smoothing for the Heisenberg alternative

The Heisenberg versions of the two basic constructions were carried out in §§9.2.5 and 9.5.6. In each of these two cases, the function f that we need to smooth out is defined on the circle S^1. As a result, this linearization is as in Example 10.11 and is much easier,

not to say entirely trivial. One formal difference from the Abelian case is that here we linearise on the six vertices of §9.2.5 (see also §9.5.7) and this contrasts with the four vertices of $\partial_0 \square^2$ for the Abelian case of the same dimension.

The other changes in §10.3.6 that have to be made for the smoothing are entirely formal and involve only a change in the notation. The details will be left as an easy exercise for the reader. Notice, finally, that in Figure 9.4 we tried to draw this embedded circle in $\mathbb{R}^4$.

10.3.9 Smoothing of the extension mapping F of §10.1.5

Here we shall go back to the mapping $F\colon B^d \to Q$ that extends $f\colon \partial B^d \to Q$ to the whole of the unit ball B^d as in §10.1.5. We shall assume that we have already performed the smoothing of f as explained and that $f \in C^\infty$. A priori, F is not smooth but there are several easy ways of modifying F to guarantee that $F \in C^\infty$ and that conditions (i) and (iii) of §10.1.8 are satisfied. The procedure that we shall describe in the next few lines is as good as any. Let us define

$$\begin{aligned}\widetilde{F}(x) &= F(2x); \quad x \in \mathbb{R}^d,\ |x| \leqslant \tfrac{1}{2},\\ \widetilde{F}(x) &= f(\sigma) = F(\sigma) \text{ for } x = r\sigma \in \mathbb{R}^d,\ \tfrac{1}{2} < r < 2,\ \sigma \in S^{d-1},\end{aligned} \tag{10.83}$$

that is, we use polar coordinates in $1/2 < |x| < 2$. The mapping $\widetilde{F}$ is Lipschitz in $|x| < 2$ and there exists C such that

$$\|\widetilde{F}\|_{\mathrm{Lip}} \leqslant C\|F\|_{\mathrm{Lip}}; \quad \widetilde{F}\Big|_{S^{d-1}} = f. \tag{10.84}$$

We can then replace $\widetilde{F}$ by $F_\varepsilon = \widetilde{F} * \varphi_\varepsilon$ for some mollifier $\varphi \in C_0^\infty$ with $\int \varphi_\varepsilon = 1$, and $\operatorname{supp}\varphi_\varepsilon \subset [|x| < \varepsilon]$, for $\varepsilon > 0$, and we have

$$\overline{\lim}\|F_\varepsilon\|_{\mathrm{Lip}} \leqslant C\|F\|_{\mathrm{Lip}}; \quad F_\varepsilon|_{S^{d-1}} = f_\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} f, \tag{10.85}$$

for the C^∞ topology (cf. the first exercise in §10.3.6). We can then replace F, f by $F_\varepsilon, f_\varepsilon$ for ε sufficiently small and all is well in §10.1.8 (i), (iii).

A final remark on the proof of Proposition 10.5 In the next section, this proposition and (10.14) will be proved under the assumption that F is smooth. Explicitly, it will be seen that for $\varepsilon > 0$ small enough the function F_ε in (10.85) will satisfy

$$\|F_\varepsilon\|_{\mathrm{Lip}} \geqslant CR^c; \quad R \geqslant C \tag{10.86}$$

with constants that are independent of $\varepsilon > 0$. If we combine this with (10.85) we have a proof for the original function F.

10.4 The Second Proof of Proposition 10.5

10.4.1 The mapping F in §10.1.5 can be made to be an embedding

In the previous section we went through the lengthy but elementary process that shows that in §10.1.5 and in the proposition the mappings f, F can be assumed to be C^∞ and also satisfy the properties of §§10.1.1–10.1.5 and of (i), (iii) of §10.1.8.

It remains to be seen how we can also assume without loss of generality that $F\colon B^d \to Q$ is in addition an *embedding* of a manifold with boundary as asserted in §10.1.8(ii). This will use standard but non-trivial facts from differential topology and can only be achieved directly if the condition

$$p = \dim Q \geqslant 2d + 1 \tag{10.87}$$

is satisfied. We shall make that assumption in §10.4.2 and complete the proof of the proposition.

In §10.4.3 we shall use a simple device that 'jacks up' the dimension of Q in Proposition 10.5 while keeping d fixed. This will allow us to make the assumption in the proof that (10.87) is satisfied and this completes the proof.

10.4.2 Use of facts from differential topology

Here the assumption (10.87) that $p \geqslant 2d + 1$ will be made and all the notation introduced in this paragraph will be maintained. The function $F\colon B^d \to Q$ in Proposition 10.5 will here be assumed to be C^∞ as explained in §10.4.1 and we shall apply the Whitney approximation theorem to this mapping (see Hirsch, 1976, Chapter 2; the fact that B^d has a boundary makes no difference because among other things, we can extend the definition of F to some neighbourhood of B^d). This theorem says that we can find new functions

$$\widetilde{F}\colon B^d \longrightarrow Q, \qquad \tilde{f} = \widetilde{F}\Big|_{S^{d-1}} : S^{d-1} \longrightarrow Q \tag{10.88}$$

that approximate F, f as close as we like in the C^∞ topology and for which the mappings are embeddings of manifolds (with boundary); that is, one-to-one and immersions (i.e. the differentials of the mappings are non-singular). These new mappings therefore satisfy all the properties of §§10.1.1–10.1.4 and also properties (i), (ii) and (iii) of §10.1.8 provided that the approximation in (10.88) is close enough.

We shall complete the proof for these two new functions $\widetilde{F}, \tilde{f}$ and to simplify the notation we shall drop the tildes '~' and denote these functions by F, f

instead. As in (10.12), we shall write

$$B = F(B^d) \subset Q, \quad \partial B = S = f(S^{d-1}), \\ V_a = [v \in V;\ |v| < a], \quad \pi\colon Q \longrightarrow V, \tag{10.89}$$

where a will be chosen appropriately small. We can then assert that $T \subset V_a$, the set of *regular values* of $\pi \circ F$ and $\pi \circ f$ is of full Lebesgue measure in V_a. (This is the content of Sard's theorem; see de Rham, 1960; Hirsch, 1976. To elaborate, $x \in B$ (resp. $x \in \partial B$) is called a *singular point* of $\pi \circ F$ (resp. $\pi \circ f$) if at that point the rank of $d(\pi \circ F)_x$ (resp. $d(\pi \circ f)_x$) is $<$ the dimension of V. Then the Lebesgue measure of the image of singular points by $\pi \circ F$ (resp. $\pi \circ f$) is zero. It is the complement of that image that we call the set of *regular* values.)

On the other hand, by (10.10) there exist constants such that

$$d(s,s') \geqslant cR^c; \quad R > C,\ x \in V_a,\ s,s' \in S_x = \pi^{-1}(x) \cap S,\ s \neq s', \tag{10.90}$$

and recall that S_x consists of m points, an even number ($m = 2^r$ in the Abelian case or 6 in the Heisenberg case). The distance in (10.90) is measured for either of the two Riemannian metrics $|\ |_2$ or $|\ |_1$ of §10.1.2. Therefore, by elementary differential topology, for all $x \in T$ we can write

$$B_x = \pi^{-1}(x) \cap B = M_x^{(1)} \cup \cdots \cup M_x^{(q)}; \quad q = \frac{m}{2}, \tag{10.91}$$

where each $M_x^{(j)}$ is a 'neat submanifold' of B (see Hirsch, 1976, §1.4), that is, it is a one-dimensional embedded submanifold of B for which $\partial M_x^{(j)}$ consists of two distinct points on S_x.

If we combine (10.90), (10.91) with the equivalence of Riemannian metrics explained in §10.1.2, it follows that there exist constants such that for the Riemannian metric $|\ |_1$ their lengths satisfy

$$l_x^{(i)} = \text{Length}\left(M_x^{(i)}\right) \geqslant cR^c; \quad R \geqslant C,\ i = 1, \ldots, q. \tag{10.92}$$

We shall now consider the Riemannian d-dimensional volume of B for the metric $|\ |_1$ in §10.1.2. By elementary differential geometry (this is sometimes called the coarea formula) we have

$$\text{Vol}_d B \geqslant \text{Vol}_d\left(\pi^{-1}(T) \cap B\right) \geqslant \sum_{i=1}^{q} \int_T l_x^{(i)}\, dx \geqslant cR^c; \quad R \geqslant C, \tag{10.93}$$

for appropriate constants. From this, exactly as in §10.2, we deduce that

$$\text{Lip}\,\widetilde{F} \geqslant cR^c; \quad R \geqslant C, \tag{10.94}$$

where the constants depend only on the constants of (10.90) and where we go back to the notation with tilde for the approximating function $\widetilde{F}$ in (10.88). If that approximation is close enough in the C^∞ topology it follows that (10.94) also holds for the original function F and this is the required estimate (10.14) for the proof of Proposition 10.5. We have thus settled the special case $p \geqslant 2d+1$.

10.4.3 Getting round the constraint (10.87) on the dimensions

It is easy to see that without condition (10.87) the Whitney approximation theorem may fail. The standard counterexample is provided by a figure 8 immersed in the plane.

Observe that in the examples of §10.1.1 it is only Examples 10.3 and 10.4 for $\dim N$ large enough that satisfy condition (10.87). This gives us a clue as to how to get round this difficulty.

Let us start from an arbitrary $Q = N \leftthreetimes V$ as in §10.1.1 and let us choose $A \geqslant 1$ sufficiently large so that for Q^* we have

$$Q^* = N^* \leftthreetimes V = \mathbb{R}^A \times Q, \quad N^* = \mathbb{R}^A \times N; \quad \dim Q^* \geqslant 2d+1, \tag{10.95}$$

with d as in (10.87). We shall denote here $\pi^* : Q^* \to V$, the canonical projection, and since $Q \subset Q^*$ canonically we can identify the functions F, f of the proposition that take their values in Q, with functions

$$F : B^d \longrightarrow Q^*, \quad f : S^{d-1} \longrightarrow Q^*. \tag{10.96}$$

On the mappings (10.96) we can therefore apply the treatment of §10.4.2 and approximate these mappings by new mappings

$$F^* : B^d \longrightarrow Q^*, \quad f^* = F^*|_{S^{d-1}} : S^{d-1} \longrightarrow Q^*, \tag{10.97}$$

where these mappings are embeddings and for which all the conditions of Proposition 10.5 and (i), (ii) and (iii) in §10.1.8 are satisfied. The idea here is that to perform the approximation of the functions of the proposition we had to spill out of Q in Q^*.

Now we can apply the result for the special case of §10.4.2 to Q^*, F^*, f^* and π^*. We deduce that

$$\operatorname{Lip} F^* \geqslant cR^c; \quad R \geqslant C, \tag{10.98}$$

where the constants in (10.98) are independent of the particular approximation (10.97) that we used and depend only on F, f.

By making this approximation sufficiently close we deduce that the same thing holds for F and

$$\operatorname{Lip} F \geqslant cR^c; \quad R \geqslant C.$$

This completes the proof of the proposition in full generality.

11

The Metric Classification

Overview of Chapter 11 (and a Preview of Part III)

In this chapter we shall give the geometric classification theorems B and NB for general groups. We shall see in particular that for this classification the special groups that are soluble and simply connected and which have preoccupied us in the first four chapters of the theory play the role of building blocks. We recall (see §7.3) that these groups are diffeomorphic to some Euclidean space (and vice versa!; see §11.2.1). The tool that allows us to use these as a building block in the theory is the notion of quasi-isometries (see §7.1; i.e. diffeomorphisms $M_1 \xrightarrow{\varphi} M_2$ with $|d\varphi|, |d\varphi^{-1}|$ bounded).

Theorem 11.14 already gives the distinct flavour of what is achieved. This theorem says that as long as the group G is *simply connected* then G is quasi-isometric with $U \times K$, where K is some compact group and U is a simply connected soluble group. Furthermore, U is a C-group (resp. NC-group) if G is a B-group (resp. NB-group). By examples one easily sees that the simple-connectedness is essential for such a quasi-isometric classification to hold.

Despite this, by passing to the simply connected cover of any connected group G, we can claim that with the above and with Chapters 7–10 we have what we wanted: namely a geometric B–NB classification of G.

Whether this classification should be considered satisfactory is a matter of opinion. One thing that is certain is that to go further we need new ideas.

One such idea is the use of the *coarse* quasi-isometries of §11.1.1 below. Without rewriting the definition here, we can say that this new metric notion captures the global large distances 'outlook' of some metric space while ignoring the local differences that could be very drastic. For instance, with this definition, $\mathbb{R}$ and $\mathbb{Z}$, with their natural distances, are coarse quasi-isometric. Locally of course, these two spaces could hardly look more different.

One way to 'popularise' the notion is to place an observer very far out: for such an observer the distinctions between consecutive integers in $\mathbb{Z}$ blur out and on the whole $\mathbb{Z}$ and $\mathbb{R}$ look the same.

Once a precise definition is given, we shall prove in §11.2 that if U_1, U_2 are two simply connected soluble groups and if, with their natural distances, they are coarse quasi-isometric then if one is a C-group (resp. NC-group) then so is the other. We shall also show (see §11.1.5) that if G is some connected B-group (resp. NB-group) then there exists U, some simply connected soluble C-group (resp. NC-group), such that $G \simeq U$ (coarse quasi-isometric).

Theorem 11.16 is in fact more precise and, in view of preparing the ground for Part III of the book, we shall say a few more things on this. In that theorem we start from some connected Lie group G and then construct a new Lie group G_1, and what the theorem says can be highlighted schematically as follows:

$$G \underset{1}{\sim} G_1 \underset{2}{\sim} Q \times K; \quad Q \underset{3}{\sim} U.$$

Here Q is some connected soluble group, U is some simply connected soluble group and K is some compact group.

Here $\underset{3}{\sim}$ is some coarse quasi-isometry which will be constructed explicitly in §11.3.3 and is quite simple. The definitions of $\underset{1}{\sim}$ and $\underset{2}{\sim}$ are even simpler: $\underset{1}{\sim}$ is given by $G_1 \to G_1/F = G$, the quotient mapping by some finite central subgroup F. As for $\underset{2}{\sim}$, it is a genuine quasi-isometry (and not a coarse one). Finally, and this by now goes without saying, all the groups G, G_1, Q and U are simultaneously B or NB.

It is this sequence of equivalences $\sim$ (: 1,2 and 3) that prepares the ground both for Appendix F and Part III of the book. Let us allow ourselves to end up by giving a preview of what happens there.

The fundamental topological invariants of homotopy and homology theory will come into play. More explicitly, in Part III we shall develop an adapted homology theory that takes into account the metric character of the space (here this space is the group G) and which is invariant by the polynomial equivalence of §7.1. Then the above equivalences $\sim$ (: 1,2 and 3) will allow us to start from any group and get back to our building blocks, that is, the simply connected soluble groups.

On these building blocks the homology will be ‘computed’ and on the basis of this we shall capture once more the B–NB classification of the group.

In the final remark of this chapter one will also find a direct and easier construction of the coarse quasi-isometry $G \simeq U$ that always exists for any given connected Lie group G and some simply connected soluble group U.

11.1 Definitions and Statement of the Metric Theorems

11.1.1 Definitions of quasi-isometries

Depending on the use one wants to make, different definitions of quasi-contractions and quasi-isometries between two metric spaces (M_1, d_1) and (M_2, d_2) can be used. Here M_i denotes the space and $d_i(\cdot,\cdot)$ denotes the distance in (M, d). We shall explain the notions involved in Definitions 11.1, 11.3, 11.4, 11.5 below.

Definition 11.1 Let $f\colon (M_1, d_1) \to (M_2, d_2)$ be a mapping between two metric spaces. We say that f is Lipschitz or a quasi-contraction if there exists some constant $A > 0$ such that

$$d_2(f(x), f(y)) \leqslant A d_1(x, y); \quad x, y \in M_1. \tag{11.1}$$

Remark 11.2 If M_1, M_2 as above are Riemannian manifolds and if f satisfies (11.1) then the differential df exists almost everywhere and

$$df\colon TM_1 \longrightarrow TM_2; \quad |df| \leqslant A \text{ a.e.} \tag{11.2}$$

(see §7.1.1). The converse is not quite true and (11.2) does not in general imply (11.1). The classical counterexample is the singular increasing function from [0,1] to [0,1] with zero derivative almost everywhere. If we make, however, the additional hypothesis that $f \in C^1$ or even locally Lipschitz then (11.2) does imply (11.1). Locally Lipschitz means that for every $m \in M_1$ there exists $m \in \Omega \subset M_1$ some neighbourhood of m on which f is Lipschitz with a constant A_Ω in (11.1) that may depend on Ω.

Together with the notion of a quasi-contraction comes the notion of a quasi-isometry.

Definition 11.3 Let (M_1, d_1), (M_2, d_2) be two metric spaces as above. We then say that M_1, M_2 are quasi-isometric if we can find

$$M_1 \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} M_2, \tag{11.3}$$

where φ, φ^{-1} are two bijective mapping inverses of each other such that φ, φ^{-1} are both quasi-contractions.

We shall now generalise the above notions so that only large distances are taken into account.

Definition 11.4 Let $f\colon (M_1, d_1) \to (M_2, d_2)$ be as in Definition 11.1. We then say that f is a coarse quasi-contraction if there exist constants $C > 0$ such that

$$d_2(f(x), f(y)) \leqslant C d_1(x, y) + C; \quad x, y \in M_1. \tag{11.4}$$

Notice that in the above definition f does not have to be continuous.

Definition 11.5 Let M_1, M_2 be two metric spaces as above. We then say that they are coarse quasi-isometric if there exists a constant C and coarse quasi-contractions α, β that satisfy

$$M_1 \underset{\beta}{\overset{\alpha}{\rightleftarrows}} M_2;$$

$$d_1\,(\beta\circ\alpha(x),x)\leqslant C,\quad d_2(\alpha\circ\beta(y),y)\leqslant C;\quad x\in M_1,\ y\in M_2. \tag{11.5}$$

Two important examples follow.

Example 11.6 Let us go back to our original definition of a compactly generated locally compact group G. Then, as we saw in §1.1, every compact generating neighbourhood of the identity $e\in\Omega\subset G$ induces a natural left-invariant distance d_Ω. We have previously pointed out (see §1.1) that the identity mapping on G is a coarse quasi-isometry for the two distances that correspond to two different such neighbourhoods Ω_1 and Ω_2.

Example 11.7 The above example can be elaborated as follows. Let G be some compactly generated locally compact group and let $d=d_\Omega$ be the above distance that corresponds to some compact neighbourhood of the identity as above. Let $K\subset G$ be some compact subgroup and let

$$\frac{G}{K}=[\dot g=gK;\ g\in G];\quad \pi\colon G\longrightarrow\frac{G}{K} \tag{11.6}$$

be the right homogeneous space and π the natural projection. When K is normal, G/K is the quotient group. Even when K is not normal the distance d on G induces on G/K the distance

$$\dot d(\dot x,\dot y)=\inf[d(x,y);\ x\in\dot x,\ y\in\dot y], \tag{11.7}$$

provided that the distance d is K-right invariant, that is, $d(xk,yk)=d(x,y)$, for $x,y\in G$, $k\in K$. (Notice that, by replacing Ω if necessary with $[\bigcup k^{-1}\Omega k;\ k\in K]$, we can always achieve this on d_Ω up to coarse quasi-isometry. Analogously, a left-invariant Riemannian structure, as in §1.4.1, can be assumed to be K-right invariant up to quasi-isometry: see §14.2.4.1 below.)

When K is normal this is again the left-invariant distance on the group G/K. For these distances π is a quasi-contraction. It is always possible to construct $s\colon G/K\to G$, a Borel mapping that satisfies $\pi\circ s(\dot g)=\dot g$ for all $\dot g\in G/K$. Such a mapping is usually referred to as a Borel section. The two mappings $G \underset{s}{\overset{\pi}{\rightleftarrows}} G/K$ induce a coarse quasi-isometry. In many important examples in this chapter the section s will be a C^∞ embedding.

From §6.1 and the above it follows, in particular, that every connected locally compact group (: LCG) as above is coarse quasi-isometric to some connected Lie group. As a consequence, several of the results of this chapter extend to this larger class of groups.

Example 11.8 Another natural example of coarse quasi-isometry occurs in the left-invariant group distances and the natural injection $i\colon R \to G$ where G is some LCG that is compactly generated and R is a closed cocompact subgroup. This means that there exists $K \subset G$ some compact subset such that $G = R \cdot K$. This example was examined in §2.12.2 where, in effect, we showed that R and G are coarse quasi-isometric. From this example we see in particular that coarse quasi-isometric spaces can in fact look very different, for example $\mathbb{Z} \subset \mathbb{R}$, the integers as a subset of the reals gives a coarse quasi-isometry. On the other hand, $\mathbb{Z}^r$ and $\mathbb{Z}^s$ are coarse quasi-isometric only if $r = s$ (prove that).

The next example that we shall consider is important and we shall use it explicitly later on.

Example 11.9 Let M be some C^∞ manifold and let R be some abstract group that acts cocompactly on the left and gives a group of diffeomorphisms on M. We denote that action by

$$m \longmapsto r \cdot m \in M; \quad m \in M,\ r \in R. \tag{11.8}$$

We shall assume that there exists $K \subset M$ some compact subset such that

$$R \cdot K = [r \cdot k;\ r \in R,\ k \in K] = M. \tag{11.9}$$

We shall assign M with two Riemannian structures, that is, two scalar products $\langle \cdot,\cdot \rangle_1$, $\langle \cdot,\cdot \rangle_2$ on the tangent space TM that are both invariant by the group action. This means that

$$\langle drX, drY \rangle_i = \langle X,Y \rangle_i; \quad r \in R,\ \ X,Y \in TM,\ i = 1,2, \tag{11.10}$$

where r is identified with the diffeomorphism (11.8). The conclusion is that these two Riemannian structures are quasi-isometric.

Exercise 11.10 Verify this. The proof is immediate since we have $\langle X,X \rangle_1 \sim \langle X,X \rangle_2$ for $X \in T_m M$, $m \in K$ by the compactness of K. We then apply (11.10).

Example 11.11 A further example of a coarse quasi-isometry that will be used systematically is given by $M \simeq M \times K$, where both M and K are metric spaces and K is a metric space of bounded diameter, for example some compact space and the direct sum distance is assigned on the product. In that case the two mappings α, β of (11.5) are supplied by the canonical injection and projection $M \to M \times K \to M$ that are induced by the Cartesian product.

Finally, it is clear that both the quasi-isometries and the coarse quasi-isometries are equivalence relations between metric spaces. Verification of this is immediate and of course without this property the notions would have been of very little use.

11.1.2 The building blocks and the C–NC classification theorem

The basic building blocks for the geometric characterisation of the classification of the C–NC conditions that we shall give are the *soluble simply connected* Lie groups that we examined in detail in Chapters 7–10.

The following is an essentially elementary result but it needs proving. We shall give the proof in §11.2 below.

Theorem 11.12 *Let U_1, U_2 be two simply connected soluble Lie groups assigned with their left-invariant Riemannian structure and let us suppose that they are coarse quasi-isometric* (: $U_1 \simeq U_2$ *as in Definition 11.5*)*. Then U_1 is a* C- (*resp.* NC-) *group if U_2 is a* C- (*resp.* NC-) *group.*

This result will be combined with the observation in Example 11.11 and we can deduce the following result.

Corollary 11.13 *Let U_1, U_2 be two simply connected soluble Lie groups assigned with their left-invariant Riemannian structures. Further, let K_1, K_2 be two bounded metric spaces* (*e.g. both compact*) *such that*

$$U_1 \times K_1 \simeq U_2 \times K_2 \quad \text{coarse quasi-isometric} \tag{11.11}$$

as in Definition 11.5. Then U_1 is a C- (*resp. NC-*) *group if and only if U_2 is.*

11.1.3 The classification theorem for simply connected Lie groups

Theorem 11.14 *Let G be some simply connected Lie group. Then there exists U some simply connected soluble Lie group and K some compact Lie group and*

$$G \underset{\varphi}{\longrightarrow} U \times K \underset{\varphi^{-1}}{\longrightarrow} G, \tag{11.12}$$

two diffeomorphisms that are bijective and inverses of each other and that give a quasi-isometry $G \simeq U \times K$ as in Definition 11.1. Furthermore, U is a C- (*resp.* NC-) *group if G is a* B- (*resp.* NB-) *group.*

This theorem should be combined with Corollary 11.13.

It is easy to show by examples (see §11.3.4 below) that the simple-connectedness is essential for (11.12) to hold.

For any connected Lie group we can consider the simply connected cover $\pi: \overline{G} \to G$ that is B (resp. NB) if G is. One could therefore argue that the above theorem, together with Theorems 7.10 and 7.11, gives the required geometric B–NB classification for general connected Lie groups.

11.1.4 Soluble non-simply-connected groups

It is here that the notion of coarse quasi-isometry takes its full significance. We recall also that a Lie group T is called a torus or a toroidal group if $T \simeq \mathbb{T}^k$ with $\mathbb{T} = \mathbb{R}(\mathrm{mod}\,1)$.

Theorem 11.15 *Let Q be some connected soluble Lie group. Then there exists U, some soluble simply connected Lie group, and T some torus and*

$$Q \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} U \times T, \tag{11.13}$$

a bijective diffeomorphism that induces a coarse quasi-isometry. Furthermore, T is isomorphic with a maximal compact subgroup of Q (see Hochschild, 1965, §XV.3.1) and U is a C- (*resp.* NC-) *group if G is.*

By concrete examples we shall see that it is impossible in general to improve (11.13) and have genuine quasi-isometries.

The complete proof of this theorem will only be given in Appendix F. The main application of this theorem is that the simply connected soluble group U is coarse quasi-isometric with Q and this is quite easy to see directly (see §11.3 below).

11.1.5 One more geometric classification of connected Lie groups

Here we shall consider a general connected Lie group G.

Theorem 11.16 *Let G be a connected Lie group. Then there exist Q, some connected soluble Lie group, and K, some connected compact Lie group. There exist also G_1, some other connected Lie group, and $F \subset G_1$, some finite central group such that G is homomorphic with G_1/F and G_1 is smoothly quasi-isometric with $Q \times K$. Furthermore, the soluble Lie group Q is a* C- (*resp.* NC-) *group if G is a* B- (*resp.* NB-) *group.*

By smoothly we mean of course that there exist diffeomorphisms

$$G_1 \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} Q \times K \tag{11.14}$$

that give a quasi-isometry.

Now if we invoke Examples 11.11 and 11.7 and Theorem 11.15, we finally see that a coarse quasi-isometry exists as in Definition 11.5,

$$G \underset{\beta}{\overset{\alpha}{\rightleftarrows}} U, \tag{11.15}$$

where U is simply connected soluble. But now a priori the α, β are not even continuous, let alone smooth. The U in (11.15) is C (resp. NC) if G is B (resp. NB). This follows from Theorems 11.15 and 11.16 and in view of Example 11.7, (11.15) also holds for general connected LCGs. A direct proof of this important consequence of Theorem 11.16, that does not use the theorem, will be given at the end of the chapter.

Remark (Not used in our theory) The amenability property of §3.1 is not preserved by quasi-isometries (see (11.42)). It is, however, preserved if we restrict ourselves to unimodular groups. One can use the HLS estimates of Varopoulos et al. (1992) (see §1.3.2) to see this. Alternatively, one can use the Følner condition (cf. Paterson, 1988, §§4.6(iii), 4.13) directly. The details are easy but will not be given in this chapter. This, on the other hand, together with §§1.2, 2.5.4, highlights the role that unimodularity plays in the theory.

11.2 Proof of Theorem 11.12

11.2.1 A reformulation of Theorem 11.12

We recall first Definition 7.12, the polynomial filling property (PFP).

Let M be some Riemannian manifold. We say that M admits the PFP if for all $n \geqslant 1$ there exists $C = C_n$ such that for all $R \geqslant 100$ and all

$$f\colon \partial\square^n \longrightarrow M, \quad f \in \operatorname{Lip} R, \tag{11.16}$$

there exists an extension

$$F\colon \square^n \longrightarrow M; \quad F|_{\partial\square^n} = f,\ F \in \operatorname{Lip}(R^C). \tag{11.17}$$

As in §7.5.1, $\square$ and $\partial\square$ denote the unit cube and its boundary in $\mathbb{R}^n$.

In view of the theorems of §§7.4–7.5, our Theorem 11.12 admits the following equivalent formulation.

Theorem 11.12′ *Let U_1, U_2 be two coarse quasi-isometric soluble simply connected groups. Then U_1 admits the PFP if and only if U_2 does.*

The PFP is in general not invariant under coarse quasi-isometries and the special nature of U_1, U_2 is needed. The proof of this theorem will be done in the next subsection. This proof, although not deep or difficult, is submerged in a 'torrent' of notation. The reader could, or indeed should, skip the details in a first reading.

It is worth noting, however, that this theorem has very little to do with group theory and is a consequence of the following result.

Lemma 11.17 *Let U_1, U_2 be two homogeneous smooth Riemannian manifolds that satisfy the following conditions*:

(a) *If $x, y \in U_i$, with $i = 1, 2$, are two points, then there exists T, a Riemannian isometry on U_i, such that $Tx = y$. We shall further assume that*

  (b) *both U_1, U_2 are diffeomorphic with a Euclidean space and*

  (c) *U_1, U_2 are coarse quasi-isometric by the mappings $U_1 \underset{\beta}{\overset{\alpha}{\rightleftarrows}} U_2$ as in Definition 11.5.*

Then if U_1 admits the PFP, so does U_2 also.

It is clear that the lemma implies Theorem 11.12′ because of the fundamental fact (see Varadarajan, 1974 and §7.3) that the simply connected soluble group satisfies the conditions of the lemma. But what is also true is that as soon as U is a Lie group that satisfies condition (b) of the lemma, then U has to be a soluble simply connected group (see the exercise below). As a result, for all practical purposes, the formulation of the lemma is essentially an equivalent, but perhaps more transparent, reformulation of Theorem 11.12′.

Exercise 11.18 Verify the above. When the Lie group U is as in (b) of the lemma, then since U is simply connected we use the Levi decomposition and the (generalised – see §4.6) Iwasawa decomposition $U = Q \leftthreetimes NAK$, $R = Q \leftthreetimes NA$, with K compact and simply connected. Therefore R is soluble and simply connected and K has to reduce to the identity. (There are all sorts of ways of seeing this last point! For example, the homology $H_n(K) \neq 0$ when $n = \dim K$.)

A digression: general locally compact groups revisited (see Chapter 6) Let G be some *connected locally compact group* such as those we considered in §6.1. Then by the general theory on these groups (see Montgomery and Zippin, 1955) there exists $K \subset G$, some compact normal subgroup, such that G/K is a Lie group. Using Example 11.7 and (11.15) we deduce therefore that G is coarse quasi-isometric with some soluble simply connected Lie group. Furthermore, if we use the previous lemma we can give a geometric (metric) B–NB classification for these connected locally compact groups. Moreover, this classification is consistent with the classification that is suggested in §6.1 and which is based on the analytic theory of Part I of the book.

This matter will, however, not be pursued any further.

11.2.2 Proof of Lemma 11.17

11.2.2.1 Notation We first recall notation from §§7.5, 9.2 and denote by $\square$ the unit cube, and by $\square_R$ the cube of size $R \gg 10^{10}$ in $\mathbb{R}^n$. For simplicity in the notation we shall suppress the dimension n in $\square = \square^n$. These are assigned with their Euclidean distance and all the Lipschitz conditions below refer to that distance. All the positive constants $c_1, c_2, \ldots$ that we shall introduce in the proof depend only on U_1, U_2 in the dimension n and the two mappings α, β, but they are independent of R. These constants are introduced successively and c_j will also depend on the previous ones, $c_{j-1}, \ldots, c_1$.

In what follows we shall assume that R, the size of the cube is large and for a larger integer $m \geqslant R$ we shall denote by $(\square_R)_m$ the grid points of the cube $[0, R/m, 2R/m, \ldots, R]^n \subset \square_R^n$ at a distance R/m apart that subdivide the cube into m^n small subcubes. The notation $(\partial\,\square)_m = (\square_R)_m \cap \partial\,\square_R$ will also be used. Two points on this grid will be called adjacent if they are vertices of the same small subcube as above. These grid points will be denoted by z_i^m and the ones that lie in the boundary will be denoted by $\dot{z}_i^m$ (see Figure 11.1).

11.2.2.2 The picture in two dimensions, $n = 2$. Cube of size R We start by subdividing $\square_R$ into subcubes of size $\lesssim 1$ as explained.

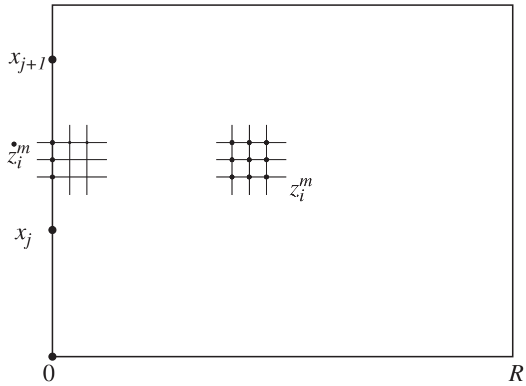


Figure 11.1 The points x_j, x_{j+1} are adjacent, i.e. a distance ≈ 1 apart on the boundary. For large m, the grid points of $(\square_R)_m$ are z_i^m, where $i \in \mathbb{Z}^2$ is a double index. The notation $\dot{z}_i^m$ indicates grid points that lie on the boundary.

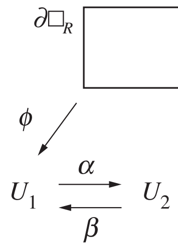


Figure 11.2 The mappings $\phi\colon \partial\square_R \to U_1$, and α, β are given.

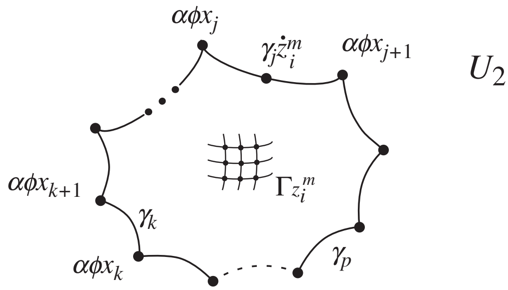


Figure 11.3 We write $\gamma = \gamma_1 \cup \gamma_2 \cup \cdots$. The term Γ denotes the extension of γ to $\square_R$, and defines the new points Γz_i^m. The following sets of points are close: $\Gamma\dot{z}_i^m = \gamma_j \dot{z}_i^m \sim \alpha\phi x_j$ and then $\beta\Gamma\dot{z}_i^m \sim \beta\alpha\phi x_j \sim \phi x_j \sim \phi\dot{z}_i^m$.

What makes the dimension $n = 2$ special is that the adjacent points x_j, x_{j+1} (see Figure 11.1) can be denoted as successive points as we enumerate them round the boundary.

Here $\phi\colon \partial\square_R \to U_1$, with $\phi \in \mathrm{Lip}(1)$, is given. (This, modulo scaling, is the same as some $\mathrm{Lip}(R)$ mapping $\square \to U_1$.) We shall assume that U_2 has the PFP and proceed with a construction that will show that U_1 also has the PFP.

11.2.2.3 The first extension (See Figures 11.1, 11.2 and 11.3.)We construct paths γ_j that join $\alpha\phi x_j$ with $\alpha\phi x_{j+1}$ and parametrise them to be $\gamma_j(t)$, with $t \in [j, j+1]$, and $\gamma_j \in \mathrm{Lip}(c_1)$. By following $\gamma_1, \gamma_2, \ldots$, one after the other, we obtain $\gamma \in \mathrm{Lip}(c_1)$, $\gamma\colon \partial\square_R \to U_2$. For this construction only properties (b) and (c) of the lemma have been used and also, at least the way it looks, the fact that the dimension $n = 2$. We shall see, however, below that this can be generalised to higher dimensions. But before this we shall finish the proofs in this special case.

We use the PFP on U_2 and extend γ to $\Gamma\colon \square_R \to U_2$ such that $\Gamma|_{\partial\square_R} = \gamma$ and $\Gamma \in \mathrm{Lip}(R^{c_2})$, with $R \geqslant 10^{10}$ where, to apply the PFP, we scale first to bring

$\square_R$ to the unit cube $\square$. From this it follows that we can find $R \leqslant m \leqslant R^{c_3}$ such that for the two adjacent points in the grid $(\square_R)_m$ we have

$$d\big(\Gamma(z_i^m), \Gamma(z_j^m)\big) \leqslant 1; \quad z_i^m,\ z_j^m \in (\square_R)_m \text{ adjacent}. \tag{11.18}$$

Let us fix some $j = 1, 2, \ldots$ and consider a point $\dot{z}_i^m$ between x_j and x_{j+1}. Then $\Gamma(\dot{z}_i^m) = \gamma_j(\dot{z}_i^m)$ is close to $\alpha\phi x_j$ and therefore $\beta\Gamma(\dot{z}_i^m)$ is close to $\beta\alpha\phi x_j$ which in turn is close to ϕx_j and $\phi\dot{z}_i^m$ (see Figure 11.3). Explicitly,

$$d\big(\beta\Gamma(\dot{z}_i^m), \phi\dot{z}_i^m\big) \leqslant c_4. \tag{11.19}$$

The bottom line is this: from $\square_R$ into U_1 we have defined two mappings:

(i) The first is $\phi\colon \partial\square_R \to U_1$ defined only on the boundary.
(ii) The second is $\beta\Gamma(z_i^m)$ and is defined only on $(\square_R)_m$.

We shall now glue together these two mappings and define

$$Y = \partial\square_R \cup (\square_R)_m \longrightarrow U_1 \tag{11.20}$$

and to be able to do this we shall first need to modify the value of the second mapping of (ii) as follows: on $\dot{z}_i^m \in \partial\square_R$ we shall not give the value $\beta\Gamma(\dot{z}_i^m)$ but the value $\phi\dot{z}_i^m$ which is close, by (11.19). The function so obtained is clearly a Lipschitz function from Y with the induced Euclidean distance to the space U_1 with the Riemannian distance. For this use (11.18), (11.19).

To render the following (and final) construction more general and more canonical we shall scale in $\mathbb{R}^n$ (here $n = 2$) so that $\square_R$ becomes $\square_{R_1}$ with $R_1 \simeq m$ and $(\square_R)_m$ becomes the 'unit' grid, that is, the lattice points $X_0 = \square_{R_1} \cap \mathbb{Z}^n$. Our set Y and our mapping in (11.20) then becomes

$$\begin{gathered} f_0\colon Y_0 = \partial\square_{R_1} \cup X_0 \longrightarrow U_1, \\ f_0 \in \operatorname{Lip}(c_5). \end{gathered} \tag{11.21}$$

The proof of Lemma 11.17 will now be a consequence of the extension properties to be described below (see Bott and Tu, 1982, p. 147).

11.2.2.4 The second extension Here there is no real advantage in assuming that $n = 2$. This extension property says that we can extend f_0 to a Lipschitz mapping on the whole cube:

$$f\colon \square_{R_1} \longrightarrow U_1, \quad f \in \operatorname{Lip}(c_6). \tag{11.22}$$

With this, if we scale back, we obtain a proof of our lemma.

11.2.2.5 The first extension revisited The first extension property is a simpler version of the mapping §11.2.2.4. The original function

$$f\colon X_0 \longrightarrow U_1, \quad f_0 \in \mathrm{Lip}(c) \tag{11.23}$$

is defined on $X_0 = \square_R \cap \mathbb{Z}^n$ for some $c > 0$ and $R \geqslant 10^{10}$. Then there exists C depending on c but independent of R such that we can extend f_0 to a function

$$f\colon \square_R \longrightarrow U_1, \quad f \in \mathrm{Lip}(C), \quad f|_{X_0} = f_0. \tag{11.24}$$

Note that here the Lipschitz property (11.23) for the function f_0 can be reformulated by saying that

$$d\big(f_0(x_1), f_0(x_2)\big) \leqslant c \quad \text{for adjacent points } x_1, x_2 \in X_0. \tag{11.25}$$

Notice that the argument of §11.2.2.3 started with the construction of the paths γ_j and these are the special case for $n = 1$ of this extension which we applied separately on the four sides of the square $\square_R^2$. Using the first extension property on each of the 2^n faces of $\square_R$ again, we can therefore repeat what was done in §§11.2.2.2–11.2.2.3 for any dimensions $n \geqslant 2$. The proof of the lemma as we finished it in §11.2.2.4 works therefore for all dimensions.

11.2.2.6 Proof of the extension properties Together with $X_0 = \square_R \cap \mathbb{Z}^n$ (here R is a large number) we shall denote by X_j the union of the j-dimensional boundaries ∂_j (see §9.4) of all the subcubes of size 1 with lattices on X_0. We have $X_0 \subset X_1 \subset \cdots \subset X_{n-1} \subset X_n = \square_R$. To prove the first extension property of (11.24), (11.25) we shall construct inductively $f_j\colon X_j \to U_1$, with $f_j \in \mathrm{Lip}(c_j)$, that give successive extensions, that is, $f_{j+1}\big|_{X_j} = f_j$.

This is immediate by the triviality of the $\pi_j(U_1) = 0$ homotopy group. The homogeneity property (a) in the lemma allows us to make these extensions with Lipschitz constants that are independent of the position of the particular face $F_{j+1} \in \partial_{j+1}$ whose boundary $\partial F_{j+1} \subset \partial_j$. More explicitly, f_j is already defined on ∂F_{j+1} and we extend it on F_{j+1}.

This is exactly what is also done for the second extension property where the inductive construction of the functions $f_j\colon \partial\square_R \cup X_j = Y_j \to U_1$ is done as before. With the same notation if $F_{j+1} \subset \partial\square_R$ then $F_{j+1} \subset Y_{j+1}$ and we do not need to do anything. If not, we have again $\partial F_{j+1} \subset Y_j$ and of course ∂F_{j+1} is bi-Lipschitz homeomorphic to $\partial\square^{j+1}$. As before, the triviality of the homotopy group $\pi_j(U_1)$ is used for the extension and just the definition of the homotopy groups is used (see Hilton, 1953).

11.3 Soluble Connected Groups

The groups that will be considered in this section will be soluble and connected, but not necessarily simply connected. Further, and more technical, results on these groups will be given in Appendix F. Here we shall give a direct proof of the main corollary of §11.1.4 which says that *every soluble connected group is coarse quasi-isometric with some soluble simply connected group*. Also, in a number of examples we shall see how the notions on metric spaces that were introduced in §11.1 apply on soluble Lie groups.

11.3.1 The maximal central torus

Let G be some connected Lie group. We shall consider toroidal subgroups, or a torus, $T \subset G$; that is, T is closed and $\cong \mathbb{T}^k$, where $\mathbb{T} = \mathbb{R} \pmod 1$. If the group $T \lhd G$ is a normal torus, it is necessarily central because the inner automorphism induced by G on T has to be a discrete group since the automorphism group of T is discrete. It follows that the inner action of G on T has to be trivial. Note also that if $T_1, T_2 \subset G$ are two such normal tori then so is T_1T_2 because that group is compact and soluble (see Hochschild, 1965 §XIII.1.3). Consequently, there exists one such torus that is maximal.

We shall denote by T_G this maximal compact central subgroup of G. This subgroup T_G is fully invariant, that is, it is invariant by all automorphisms $\alpha \in \mathrm{Aut}(G)$, $\alpha(T_G) = T_G$. As a consequence, if $H \subset G$ is some connected closed normal subgroup we have $T_H \subset H \cap T_G$ because T_H is normal in G. Easy examples show (see below) that we do not in general have $T_H = H \cap T_G$. This, however, is the case when H is assumed nilpotent (see next subsection). Observe finally that the quotient group G/T_G contains no non-trivial central tori, for otherwise its inverse image in G would be soluble and compact and strictly larger than T_G. One other incidental fact that we shall not actually use is that since T_G is central it lies in the nilradical $N \lhd G$ and therefore $T_G = T_N$.

Example Let T be a one-dimensional torus and $F \subset T$ a finite subgroup. Let S be some compact semisimple group and $F_1 \subset S$ some finite central subgroup, $F \simeq F_1$ (see Helgason, 1978, §§II.6.9 and VII.6). Let G be the group that we obtain from $T \times S$ after we identify F with F_1, that is, quotient by $(x, -x)$, with $x \in F$. Then T can be identified with some central subgroup of G and $T \cap S_1 = F_1$ (with the obvious identification of S with a subgroup $S_1 \subset G$). On the other hand, $T_{S_1} = \{0\}$.

11.3.2 Nilpotent groups

The central tori in a nilpotent group N are easy to describe. The reason is that if $\pi : \widetilde{N} \to N$ is the universal covering group then $D = \ker \pi \subset \widetilde{N}$ is a central discrete subgroup $\cong \mathbb{Z}^k$ and therefore by the nilpotency it follows that $D \subset Z(\widetilde{N})$ the analytic centre of $\widetilde{N}$ (see Hochschild, 1965, §XVI.1.1; Varadarajan, 1974, §3.6.4).

We have then the following natural identifications:

$$N = \frac{\widetilde{N}}{D}, \quad \frac{Z(\widetilde{N})}{D} = V \times T;$$
$$V \cong \mathbb{R}^p, \quad T = \mathbb{T}^q, \quad T_N = T \subset N; \tag{11.26}$$
$$N_1 = \frac{N}{T} = \text{ simply connected.}$$

Here we use the fact that every connected Abelian Lie group is of the form $V \times T$. These can be summarised by the exact sequence

$$0 \longrightarrow T_N \longrightarrow N \longrightarrow N_1 \longrightarrow 0. \tag{11.27}$$

Exercise 11.19 With the help of Varadarajan (1974, §3.6), verify the above facts.

Now if $K \subset N$ is some compact subgroup of N then $T_N K$ is also compact, and by (11.27) the image of that group in N_1 is also compact, and therefore by the simple-connectedness of N_1 this image is $\{e\}$. As a consequence $K \subset T_N$.

Example 11.20 The Heisenberg algebra $\mathfrak{h} = (\zeta, \mu, \nu)$ is generated by the three vectors such that $\zeta = [\mu, \nu]$ lies in the centre of $\mathfrak{h}$ (Definition 9.5). Using the exponential coordinates of §7.3.1 we see that the corresponding simply connected group $\mathbb{H}$ is generated by the one-parameter subgroups

$$\begin{aligned} &\mathbb{H} = ZXY; \quad X = \left[X(t_1) = e^{t_1 \nu}\right], \\ &Y = \left[Y(t_2) = e^{t_2 \mu}\right], \ Z = \left[Z(\tau) = e^{\tau \zeta}\right]; \quad t_1, t_2, \tau \in \mathbb{R}. \end{aligned} \tag{11.28}$$

Let

$$D = [Z(n);\ n \in \mathbb{Z}] \subset Z \subset \mathbb{H}. \tag{11.29}$$

Then we shall consider the identifications

$$\mathbb{H}_T = \frac{\mathbb{H}}{D} = TXY; \quad T = \frac{Z}{D} = T_{\mathbb{H}_T}, \tag{11.30}$$

where (11.30) denotes the product of the one-parameter subgroups

$$\mathbb{H}_T = [Z(\tau)X(t_1)Y(t_2);\ t_1, t_2 \in \mathbb{R},\ \tau \in \mathbb{R}(\bmod 1) = \mathbb{T}]. \tag{11.31}$$

Here we use the same notation as (11.28) and identify the subgroups with the images of the subgroups (11.28) by the mapping $\mathbb{H} \to \mathbb{H}_T$. Of course $Z(\tau)$ is the subgroup $Z/D = T_{\mathbb{H}_T}$, hence the periodic values of τ in (11.31).

This is therefore a version of 'periodic' exponential coordinates similar to §7.3.1.

These exponential coordinates give a C^∞ identification

$$\mathbb{H}_T \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} \mathbb{T} \times \mathbb{R}^2 = \mathbb{H}_0. \tag{11.32}$$

Here on $\mathbb{H}_0$ we shall give the direct product group structure and by the discussion in Example 11.7, φ is a coarse quasi-isometry for the left-invariant Riemannian structures that are involved.

These types of coarse quasi-isometries are fundamental in the theory. They illustrate the notions of §11.1.1 very well because in particular $\mathbb{H}_T$ and $\mathbb{H}_0$ are not quasi-isometric. To see that the mappings of (11.32) are not quasi-isometries we consider the corresponding mappings on the simply connected covering groups: these cannot be quasi-isometries since then the two sides of would have different volume growths at infinity (see Varopoulos et al., 1992). The volume growth, on the other hand, is clearly a quasi-isometric invariant. The fact that no quasi-isometry exists between the two groups of (11.32) will not be used but nevertheless for the proof we can again appeal to volume growth, together with the "lifting" argument that we use in the more involved, but similar, situation developed in §11.3.4 below.

11.3.3 Tori in connected soluble groups

The type of argument that we used in §11.3.2 applies to connected soluble groups Q as follows. Let $N \lhd Q$ be the nilradical (a closed subgroup; see Varadarajan, 1974, §3.18.13). Then Q/N is Abelian and of the form $V \times T$ as in (11.26). Let $Q_1 = \pi^{-1}(V)$ for the canonical projection $\pi\colon Q \to Q/N$. Then, by Example 11.8, Q_1 is coarse quasi-isometric with Q. On the other hand, if T_N is the maximal central torus of N then the group Q_1 is, by Example 11.7, coarse quasi-isometric with $Q_1/T_N = U$ and this group is simply connected soluble and, furthermore, U is a C- (resp. NC-) group if Q is a C- (resp. NC-) group.

Exercise Use §2.2 to verify this last point.

With the above we have kept the promise that we made at the beginning of this section that Q is coarse quasi-isometric with U.

Using this coarse quasi-isometry for soluble connected groups we shall see in the next section that this can in fact be achieved for all the connected Lie groups. Nonetheless, it is worth noting that the proof that we gave also works for all amenable groups (see §3.1).

Exercise Refer to §F.3, (F.9) and verify this.

11.3.3.1 Exponential coordinates revisited We shall consider a special case of §11.3.3 and assume that $Q/N = V$ is a Euclidean space. In other words, with our previous notation, we shall assume that $T = 0$ but that the central torus T_Q is not necessarily trivial. We then have

$$\pi\colon Q \longrightarrow \frac{Q}{T_Q} = Q^* = \text{ simply connected.} \tag{11.33}$$

We shall express Q^* in terms of exponential coordinates of the second kind

$$Q^* = \{e_1^*(r_1)\cdots e_n^*(r_n);\ r_j \in \mathbb{R},\ 1 \leqslant j \leqslant n\}, \tag{11.34}$$

where we use the notation $e_j^*(r_j) = e^{r_j \xi_j^*}$ for the one-parameter subgroups as in §7.3 for appropriate vectors $\xi_j^* \in \mathfrak{q}^*$ the Lie algebra of Q^*. We then lift the ξ^* by $d\pi(\xi_j) = \xi_j^*$ in an arbitrary manner to vectors $\xi_j \in \mathfrak{q}$ in the Lie algebra of Q and write $e_j(r_j) = e^{r_j \xi_j} \in Q$ for the corresponding one-parameter subgroups; these can be used to define the products and the bijective mapping

$$\begin{gathered} \sigma\colon e_1^*(r_1)\cdots e_n^*(r_n) \longrightarrow e_1(r_1)\cdots e_n(r_n), \\ \Sigma = [e_1(r_1)\cdots e_n(r_n) \in Q;\ r_j \in \mathbb{R}] \end{gathered} \tag{11.35}$$

and this, by (11.34), can be identified to a section (see §8.4)

$$Q \underset{\pi}{\longrightarrow} Q^* \underset{\sigma}{\longrightarrow} \Sigma, \quad \pi \circ \sigma = \text{Identity}. \tag{11.36}$$

We have thus obtained a bijective diffeomorphism

$$\begin{gathered} Q \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} \frac{Q}{T} \times T, \quad T = T_Q, \\ \frac{Q}{T} \rightleftarrows \Sigma, \end{gathered} \tag{11.37}$$

and as we saw in Example 11.7, φ, φ^{-1} give a coarse quasi-isometry. From this we also obtain exponential coordinates of the second kind on Q itself, exactly as in (11.31). This can be done because T_Q is central (fill in the details).

11.3.4 The role of the fundamental group

We shall present here an example that illustrates a number of important points.

Example 11.21 The notation is as in Example 11.20. Once again $\mathfrak{h} = (\mu, \nu, \zeta)$ is the Heisenberg algebra and $\mathfrak{a} = \xi\mathbb{R}$ is the one-dimensional Abelian algebra. As in Definition 9.5 we define the C-algebra of Heisenberg type $\mathfrak{q} = \mathfrak{h} \leftthreetimes \mathfrak{a}$

given by the action $[\xi,\mu]=L\mu$, $[\xi,\nu]=-L\nu$, $[\xi,\zeta]=0$ for some $L\neq 0$. The simply connected group that corresponds to $\mathfrak{q}$ is $\tilde{Q}=\mathbb{H}\leftthreetimes\mathbb{R}$ where $\mathbb{H}$ is the Heisenberg group as in §11.3.2. With the notation of §11.3.3,

$$D=[Z(n);\ n\in\mathbb{Z}]\subset Z\subset\mathbb{H}\subset\tilde{Q} \tag{11.38}$$

is again a discrete central subgroup in $\tilde{Q}$ and in (11.33) we can take $\tilde{Q}/D=Q_T=\mathbb{H}_T\leftthreetimes\mathbb{R}$ with the natural identifications and the natural action, and the notation of (11.30). Here the central torus $T_{Q_T}\subset Q_T$ is $T=Z/D$ and the group Q_T/T is $\mathbb{R}^2\leftthreetimes\mathbb{R}$ with the action $[\xi,\mu']=L\mu'$, $[\xi,\nu']=-L\nu'$ for the basis (μ',ν') of $\mathbb{R}^2$ and ξ the basis vector in $\mathbb{R}$. This is the simplest C-group of Abelian type of §§9.1.7 and 9.2.2. Because of (11.32), topologically we do have a diffeomorphism

$$Q_T=\mathbb{H}_T\leftthreetimes\mathbb{R}\ \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}}\ T\times\left(\mathbb{R}^2\leftthreetimes\mathbb{R}\right)=Q_0. \tag{11.39}$$

What we shall show is that not only can we not choose such a diffeomorphism φ to be a quasi-isometry but that φ cannot even be chosen to be a polynomial map as in §7.1. In showing this we shall illustrate in addition the role that is played by the fundamental group in these types of considerations.

To see this let $\widetilde{Q}_0=\mathbb{R}\times(\mathbb{R}^2\leftthreetimes\mathbb{R})$ denote the simply connected covering group of Q_0 and let $\tilde{\varphi}\colon\widetilde{Q}\to\widetilde{Q}_0$, the lifting of the mapping φ where we shall assume, as we may, that $\varphi(e)=e$, $\tilde{\varphi}(e)=e$ for the corresponding neutral elements. Our assertion that φ cannot be polynomial is a consequence of the following observations.

(i) We have $\tilde{\varphi}(Z(n))=n\subset\mathbb{Z}\subset\mathbb{R}=$ the first factor in the product that defines $\widetilde{Q}_0$ (we prove this below).

(ii) The distance in $\widetilde{Q}_0$ between e and $\tilde{\varphi}(Z(n))$ is $\sim n$. This follows from (i) and the product structure of $\widetilde{Q}_0$.

(iii) The distance between $Z(n)$ and e in $\tilde{Q}$ is $\lesssim\log n$. The proof of this is a variant of the argument of §9.2.5 as indicated in Figure 9.4; the reader is invited to prove this for themselves. Here we have in fact a typical example of what in Varopoulos (2000a) – see also §8.2.2 – is called a subgroup of 'strict exponential distortion' and if you get stuck with the proof and absolutely wish to see it you can always look it up there.

Proof of (i) This is a consequence of the winding number that we obtain from some path $\gamma\subset\tilde{Q}$ (i.e. $\gamma(t)\in\tilde{Q}$, $0\leqslant t\leqslant 1$) and its image $\tilde{\varphi}(\gamma)\subset\widetilde{Q}_0$ once we project these paths on the two groups Q_T and Q_0. Indeed, as long as these projected paths are closed they have the same winding number on the cylinder $T\times\mathbb{R}^3$ which is topologically diffeomorphic to Q_T and Q_0. □

This example shows that the group Q_T cannot be polynomially equivalent to any group of the form $K \times U$ with K compact and U simply connected soluble. Had that been the case we would have $K \simeq \mathbb{T}$ and U some three-dimensional C-group. This forces U to be the group in §9.2.2!

Exercise Verify the above statements

This example should be compared with results in Appendix F.

11.4 General (Not Necessarily Soluble) Groups and Theorems 11.14 and 11.16

In the previous two sections, we have exclusively considered soluble groups. There the main tool from the point of view of global structure theory was the use of the exponential coordinates of the second kind of §7.3. It is a surprising feature of our geometric theory that these soluble groups are pivotal and that they in fact capture the general classification of all the groups. We shall first deal with the case where these groups are simply connected.

11.4.1 Notation and structure theorems

Let G be some simply connected Lie group. We shall make essential use of the Levi decomposition (see Varadarajan, 1974, §3.18)

$$G = Q \leftthreetimes S; \quad Q \text{ radical of } G, \tag{11.40}$$

and S some semisimple closed Levi subgroup. Several choices of (11.40) exist and we shall fix one. The subgroup S may or may not be compact and as we saw in §3.1 this is what makes the distinction between amenable and non-amenable groups. We can write in both cases (somewhat abusively) the Iwasawa decomposition

$$\begin{gathered} S = NAK; \quad (n,a,k) \to nak, \\ n \in N,\ a \in A,\ k \in K \text{ is a diffeomorphism.} \end{gathered} \tag{11.41}$$

For simply connected closed subgroups N, A, K, we have N is nilpotent, A is Abelian. When S is of non-compact type this is the classical Iwasawa decomposition that we explained in Appendix A. In the general case we shall use the terminology of §4.6 and write a simply connected $S = S_n \oplus S_c$ for its compact and non-compact components and $S_n = NAK_n$ the classical Iwasawa decomposition. Then in (11.41) we set $K = K_n \oplus S_c$. The centre $Z(S)$ is a closed discrete subgroup and $Z(S) \subset K$ with a compact quotient $K/Z(S)$.

Most people that have had some exposure to Lie group theory are probably already familiar with the above facts and notation. A classical reference is Helgason (1978). At any rate we have already made use of (11.40), (11.41) in our definition of the Iwasawa radical $R = QNA$. This is a soluble closed simply connected subgroup and R is a C- (resp. NC-) group if and only if G is a B- (resp. NB-) group. For the above the reader can also refer to Chapter 4 and Appendix A. This in fact gives the definition of the B-condition for G. In Chapter 4, in the case when K is compact, the product space $X = R \times K$ has already played a crucial role in the analysis of the convolution operators on G.

Whether K is compact or not we shall push the product structure a step further and together with the original group G we shall consider a new group G_0,

$$G = RK, \qquad G_0 = R \times K, \tag{11.42}$$

for the direct product group structure on G_0. An alternative way of viewing (11.42) is to say that we have given on the same space G two *different* Lie group structures.

To handle the two groups in (11.42) we shall need to recall an additional feature of the centre $Z(S) \subset S$. We saw in §4.6.1 (in the proof that we gave there S was of non-compact type, but the proof works in general) that there exists $Z \subset Z(S)$ a subgroup of finite index (i.e. $Z(S)/Z$ is finite; recall also that $Z(S)$ is finitely generated and $Z(S) \subset K$) such that $Z \subset Z(G)$, that is, $zg = gz$, $z \in Z, g \in G$.

Proof of Theorem 11.14 The first step for the proof of this theorem is to prove that the two groups G and G_0 assigned with their left-invariant structures are quasi-isometric.

This is a direct consequence of Example 11.9. To see this with Z as defined in the previous subsection we define two subgroups

$$H = RZ \subset G, \quad H_0 = R \times Z \subset G_0. \tag{11.43}$$

(These subgroups are closed but this is not relevant in what follows.) The fact that Z is central has to be used to see that H is a subgroup. What counts is that in both cases H, H_0 act cocompactly on the two manifolds G and G_0 by left group action. Indeed, the homogeneous spaces in both cases can be identified with the compact group K/Z (see Helgason, 1978, Chapter 6).

A digression To finish the proof of our theorem we shall need to use the following additional information. *The group K is a direct product*:

$$K = E \times K_0; \quad E = \mathbb{R}^k \text{ is a Euclidean space}, K_0 \text{ is compact}. \tag{11.44}$$

The proof of this is difficult and furthermore it is not really part of general knowledge on Lie groups. It is, however, a direct consequence of the fact that the mapping

$$K \longrightarrow \frac{K}{Z(S)} = \text{compact} \tag{11.45}$$

is a locally injective map. Proofs can be found in Hochschild (1965, §XIII.2.1) even in the context of general connected locally compact groups.

Return to the Proof of Theorem 11.14 Combining the quasi-isometry so obtained in (11.42) and (11.44) we get a quasi-isometry

$$G = REK_0 \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} G_0 = R \times E \times K_0 = R_E \times K_0; \quad R_E = R \times E \tag{11.46}$$

where $\times$ indicates direct product group structures. Here R_E is a C-group if and only if R is (see §2.2), that is, if and only if G is a B-group. This completes the proof of Theorem 11.14. □

Remark It is much easier to prove that there exists $E \simeq \mathbb{R}^k$ that is a closed subgroup of K with $Z \subset E$ (see Hochschild, 1965, Chapter XVI) than the full thrust of (11.44). With this and Example 11.8 we already have (11.15) for simply connected groups.

Remark 11.22 (The double coset decomposition) In (11.46) we wrote $G = RK_0E$ and this a double coset decomposition with respect to the two subgroups R, E. Furthermore, it is easy to verify from well-known facts on the construction of the Iwasawa decomposition that $K_0 \subset S$ is a maximal compact subgroup of S.

Exercise 11.23 To see this let $P \subset S$ be some compact subgroups of the simply connected semisimple group. Then $P \subset K$ for some Iwasawa decomposition $S = NAK$. This holds because the image of P in $S/Z(S) = S_1 = NAK_1$ lies in K_1 (see Helgason, 1978, §6.1.1). As a consequence $P \subset \mathbb{R}^d \times K_0$ and therefore $P \subset K_0$. We shall not pursue the matter because it is not of great significance in our theory. From this it follows that K_0 is a maximal compact subgroup of G also, because for any compact subgroup M, $M \cap Q = \{e\}$ (see Hochschild, 1965, §XII.2.3) since Q is simply connected and soluble.

At any rate it is worth recording the above as follows:

The quasi-isometry of Theorem 11.14 is associated with the double coset decomposition with respect to the Iwasawa radical R and some Euclidean subgroup E and $R\backslash G/E = K_0$ can be identified with a maximal compact subgroup.

Towards the proof of Theorem 11.16 Before we give the proof we need to use the following additional information on the decomposition $K = E \times K_0$ of (11.44). For this we need to observe that K_0 is simply connected and compact. As a result, by the basic structure theorem on these groups (see Hochschild, 1965, §XIII.1.3), K_0 is a semisimple group and $Z(K_0)$, the centre, is finite. Using this fact we see that in our choice of $Z \subset Z(S)$ in (11.43) the projection of $Z \subset K$ by $K = E \times K_0 \to K_0$ lies in $Z(K_0)$. Therefore by taking a subgroup of Z of finite index we may suppose that

$$Z \subset E \cap Z(G). \tag{11.47}$$

This remark will be used in the proof of Theorem 11.16 that we shall give below.

11.4.2 The simply connected covering group

Let G be some connected Lie group and let $\pi\colon \overline{G} \to G$ be the covering by the simply connected group $\overline{G}$. For the group $\overline{G}$ we have the decompositions (11.40), (11.41) and with our previous notation we can summarise the information we have as follows:

$$\begin{aligned} &p\colon \overline{G} = Q \leftthreetimes S \longrightarrow S = NAK;\\ &K = E \times K_0,\ Z \subset E \cap Z(\overline{G}),\ [Z(S)\colon Z] < +\infty. \end{aligned} \tag{11.48}$$

Now let $D \subset \overline{G}$ be some discrete central subgroup $D \subset Z(\overline{G})$. Since $p(D) \subset Z(S)$ the subgroup

$$\Gamma = D \cap p^{-1}(Z); \quad [D\colon \Gamma] < +\infty \tag{11.49}$$

is a subgroup of finite index in D. We shall examine Γ in more detail and use the same notation, $g = q \leftthreetimes s \in \overline{G}$, for the product in $\overline{G}$ with $q \in Q$, $s \in S$ in (11.40) as we did in (8.54). For $\gamma \in \Gamma$ we can then write

$$\gamma = \gamma_Q \leftthreetimes \gamma_S \in \Gamma; \quad \gamma_Q \in Q,\ \gamma_S \in Z \subset Z(\overline{G}) \tag{11.50}$$

and therefore also

$$\begin{aligned} &\gamma_Q \in Z(\overline{G}), \quad \gamma \longrightarrow \gamma_Q, \quad \gamma \longrightarrow \gamma_S \text{ are group homomorphisms,}\\ &(q \leftthreetimes s)\gamma = q\gamma_Q \leftthreetimes s\gamma_S; \quad \gamma \in \Gamma,\ q \in Q,\ s \in S. \end{aligned} \tag{11.51}$$

Furthermore, if $\overline{G} = REK_0$ is the decomposition of (11.42), (11.44) we shall express the elements of $\overline{G}$ as products in that decomposition and we have

$$\begin{aligned} &g = rxk; \quad r \in R,\ x \in E,\ k \in K_0,\ \gamma = \gamma_Q \leftthreetimes \gamma_S \in \Gamma;\\ &g\gamma = (r\gamma_Q)(\gamma_S x)k; \quad r\gamma_Q \in R,\ \gamma_S x \in E. \end{aligned} \tag{11.52}$$

This holds because γ_Q, γ_S are central elements in $\overline{G}$.

We shall consider now the direct product group in (11.46),

$$\overline{G}_0 = R \times E \times K_0 = R_E \times K_0,$$

and $g_0 = r \times x \times k$ (to indicate the product in $\overline{G}_0$) and we shall define the discrete central subgroup

$$\Gamma_0 = [\gamma_Q \times \gamma_S \times e;\ \gamma = \gamma_Q \curlywedge \gamma_S \in \Gamma] \subset \overline{G}_0,$$

where $e \in K_0$ is the identity and we recall that $\gamma_S \in E$ by our choice of Z in (11.47) and of Γ in (11.49). The left (and right) action of Γ_0 on $\overline{G}_0$ is then given as follows. For

$$g_0 = r \times x \times k = r_E \times k \in \overline{G}_0 = R_E \times K_0; \quad R_E = R \times E, \tag{11.53}$$

$$\gamma_0 = \gamma_Q \times \gamma_S \times e \in \Gamma_0; \quad \gamma = \gamma_Q \curlywedge \gamma_S \in \Gamma, \tag{11.54}$$

we have

$$g_0\gamma_0 = r\gamma_Q \times x\gamma_S \times k = r_E\gamma_0 \times k. \tag{11.55}$$

Finally, if we put together (11.52), (11.55) we see that the actions of the groups Γ and Γ_0 intertwine the identification (11.46), that is,

$$\varphi(g\gamma) = \varphi(g)\gamma_0 \tag{11.56}$$

with γ, γ_0 as in (11.54).

Since the action of Γ (resp. Γ_0) is discrete we see that the quasi-isometric diffeomorphism φ in (11.46) induces a quasi-isometric diffeomorphism on the corresponding two quotient groups:

$$\overline{G} \longrightarrow \frac{\overline{G}}{\Gamma}; \quad \overline{G}_0 \longrightarrow \frac{\overline{G}_0}{\Gamma_0}. \tag{11.57}$$

We denote this as

$$\frac{\overline{G}}{\Gamma} \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} \frac{\overline{G}_0}{\Gamma_0} = \left(\frac{R_E}{\Gamma_0}\right) \times K_0 = P \times K, \tag{11.58}$$

where P is some connected soluble group, not in general simply connected, that is C- (resp. NC-) if G is B- (resp. NB-).

11.4.2.1 A recapitulation *We start from some central discrete subgroup $D \subset \overline{G}$. We can then find $\Gamma \subset D$, a subgroup of finite index, such that we have a smooth quasi-isometry as in* (11.58).

Proof of Theorem 11.16 We go back to the original group G of the theorem and to the covering map $\pi\colon \overline{G} \to G$ and apply the previous considerations to $D = \ker \pi$. We then have canonical mappings

$$\theta\colon \frac{\overline{G}}{\Gamma} \longrightarrow \frac{\overline{G}}{D} = G; \quad \ker\theta = \frac{D}{\Gamma} = \text{ finite}. \tag{11.59}$$

If we combine (11.58), (11.59) we finally see that we have a proof of Theorem 11.16. □

Remark The most important application of Theorem 11.16, at least for the time being, is to prove (11.15), which says that the connected Lie group G is coarse quasi-isometric with some connected soluble group that is C (resp. NC) if G is B (resp. NB). For this, much easier proofs can be given but they depend on proving first the easy fact that $R \subset G$, the analytic subgroup that corresponds to the Iwasawa radical $\mathfrak{r} = \mathfrak{q} + \mathfrak{n} + \mathfrak{a}$ as in §4.1, is a closed subgroup. The proof of this is left as an exercise (it is spelled out in Varopoulos, 1996b, §4.8).

With this we consider the connected semisimple group $S = G/Q$ where Q is the radical. If $Z(S)$ the centre of S is finite, the group R is cocompact and $G \simeq R$ are coarse quasi-isometric by Example 11.8, and we are done.

In the general case we proceed as in §4.6.3 and select $\bar{z}_1, \dots, \bar{z}_p \in Z(G)$ to be in the centre of G and such that their canonical images $z_1, \dots, z_p \in S$ are the free generators of a subgroup of finite index in $Z(S)$.

Write $\mathbb{Z}^p \simeq \Gamma = Gp(\bar{z}_1, \dots, \bar{z}_p) \subset G$. Then $\overline{R} = \Gamma \oplus R \subset G$ is a closed cocompact subgroup (exercise: verify this). Therefore $G \simeq \overline{R}$ as before. On the other hand, again by Example 11.8, $\overline{R}$ is coarse quasi-isometric with the connected soluble group $\mathbb{R}^p \oplus R$ which is a C-group if and only if R is (see §2.2). And we are done again.

Appendix F

Retracts on General NB-Groups (Not Necessarily Simply Connected)

F.1 Introduction

This appendix could have been presented as a separate chapter because in §F.4–F.5 it contains the proof of one of the main theorems of the subject. More precisely, we shall prove in this appendix the following result.

Theorem (NB–Pol) *Let G be some connected* NB *Lie group; then G is polynomially homotopic* (*see* §*7.4*) *to some compact manifold G_0.*

One of the main reasons that the homology theory of Part III of the book was developed is that it allows us to prove the above theorem the other way round, and show that when G is polynomially homotopic to a compact manifold, then G is an NB-group. In other words, we have a B–NB classification in terms of polynomial homotopy.

Maximal compact subgroups In the above theorem, G_0 can be chosen to be some maximal compact subgroup $G_0 \subset G$.

We encountered maximal compact subgroups of Lie groups in §11.4.1 and in Appendix C. Let us recall here some important features of these subgroups (see Hochschild, 1965, §XV.3.1; Helgason, 1978 §VI.2). These facts are deep and difficult to prove.

These subgroups are what their name says, that is, they are compact subgroups and are contained in no other compact subgroups. Furthermore, if $L \subset G$ is some compact subgroup and G_0 is such a maximal subgroup of G, there exists $g \in G$ such that $gLg^{-1} \subset G_0$. Maximal compact subgroups exist and they are connected when G is. Better still, if G is connected, when $G_0 \subset G$ is a maximal compact subgroup and if $d = \dim G - \dim G_0$ there exists a diffeomorphism

$$G \simeq G_0 \times \mathbb{R}^d. \tag{F.1}$$

Incidentally, it is this diffeomorphism that 'triggers' the third part of the book: see §12.1.3. Note, however, that apart from this motivation, (F.1) will not be used in any essential way

R-groups The other thing that will be proved in this appendix is the following result.

Proposition *When G is a connected R-group* (*see* §*2.2.2*) *then the diffeomorphism* (F.1) *can be chosen to be a polynomial equivalence* (*see* §*7.4*). *The converse also holds*.

The proposition is one of the ingredients for the proof of the above (NB–Pol) theorem. It will be proved in §§F.2–F.3 below, together with a number of other geometric properties of R-groups. We have already encountered most of these geometric properties before (in Chapters 7 and 8) and their proof is long overdue. These properties are, on the other hand, well known and their connections to the subject were studied long before the B–NB classification of this book (see Guivarc'h, 1973; Jenkins, 1973).

There are two reasons that justify the presentation of all this as an appendix. One is that although it comes up with one of the main results of the geometric theory, the proofs are 'odds and ends' of things that we have done elsewhere and the real advantage of putting this material in an appendix is that by doing so, we felt free to make the exposition less formal, lighter and more condensed.

The other reason is that while the proof of the main theorem in this appendix is not very hard, it does rely on the ideas that were presented in Appendix A, Appendix B and Appendix C. Note, on the other hand, that the simply connected case of the theorem follows directly from §11.1.3, and also that these ideas, which are non-trivial, are not essential if we are prepared to settle with something slightly less (see §12.6.2).

F.2 R-Groups

Throughout this section we shall consider connected R-groups (see §2.2) and we shall elaborate on the following easy proposition (see (8.9)).

Proposition *Let G be some connected R-group. Then the norm of the* Ad-*action of G on the Lie algebra $\mathfrak{g}$ grows polynomially, that is, there exist constants such that*

$$\|\mathrm{Ad}\, g\| \leqslant C(1+|g|)^C; \quad g \in G. \tag{F.2}$$

Here, as usual, $|g|$ denotes the distance from the identity e.

Proof By considering the simply connected cover $\pi\colon \tilde{G} \to G$ and by lifting any $g \in G$ to some $\tilde{g} \in \tilde{G}$ such that $\pi(\tilde{g}) = g$, $|\tilde{g}| \leqslant 2|g|$ (see Exercise 2.6), we may assume that G is simply connected. But then by the Levi decomposition (see §2.2.2 and Varadarajan, 1974, §3.18) we have $G = Q \leftthreetimes K$ where Q is soluble and K is compact. Then g in (F.2) can be written as $g = qk$, with $q \in Q$, $k \in K$, and since, by compactness, $\|\mathrm{Ad}\, k\| \leqslant C$ stays bounded and additionally $|q| \lesssim |g|$ (see §§1.1, 2.14), we may assume in (F.2) that $g \in Q$. To finish the proof, we look at the Ad-action of Q on $\mathfrak{g}$. The roots of that action, $\omega_1, \dots,$ are the exponentials of the roots of the ad-action of $\mathfrak{q}$, the Lie algebra of Q, on $\mathfrak{g}$. Now, $[\mathfrak{q}, \mathfrak{g}] \subset \mathfrak{q}$ and the roots of $\mathrm{ad}\,\mathfrak{q}$ on $\mathfrak{q}$ are pure imaginary by definition. It follows that $|\omega_1| = |\omega_2| = \cdots = 1$. For this, we then invoke Lie's theorem (see Varadarajan, 1974, §3.7.3, and also §2.3.3 above) and the discussion that follows (3.59) is used to cope with the exponential mapping on soluble groups (cf. §3.8.4). We then use the lemma proved in §3.9.1, and estimate (F.2) follows. □

The converse assertion, (F.2) $\Longrightarrow$ (G is an R-group), is also essentially contained in the above proof.

Exercise F1 Prove this assertion. The same lemma from §3.9.1 implies, when (F.2) holds, that the radical $Q \triangleleft G$ is an R-group (use $\mathrm{Ad}\, e^{\xi} = e^{\mathrm{ad}\,\xi}$). Now (F.2) will also hold on the semisimple group G/Q. Using the structure theorems on these groups, deduce that G/Q is compact. For more details on all that see Guivarc'h (1973), Jenkins (1973) and part (i) of the following exercise.

Exercise F2 Deduce from the proposition and Exercise F1 the following well-known and easy facts:

(i) The algebra $\mathfrak{g}$ is an R-algebra if and only if the characteristic roots of the $\mathrm{ad}\,\xi$-action on $\mathfrak{g}$, with $\xi \in \mathfrak{g}$, are pure imaginary (use the lemma in §3.4.4 or in §3.9.1 on the one-dimensional algebra $\{\mathrm{ad}\,\xi\}$ acting on the space $\mathfrak{g}$).

(ii) Use (i) and the action induced by ad to prove that subalgebras and quotients of the R-algebras are R.

(iii) Use the proposition and §8.1 to prove that in an R-group the following mappings are polynomial:

  (a) $G \to G$ given by $g \mapsto g^{-1}$;
  (b) $G \times \cdots \times G \to G$ given by $(g_1, \dots, g_n) \to g_1 \cdots g_n$ (group product);
  (c) $\mathfrak{g} \to G$ given by the exponential mapping $\xi \to \exp \xi$.

  We have already encountered (c) in Example 7.5 in the case of the nilpotent groups. (For the proof see Varadarajan, 1974, §3.6. For nilpotent groups, the Baker–Campbell–Hausdorff formula could be used.) For general R-groups, this property (c) will not be needed but it can be proved by the explicit formula for the differential of the exponential mapping (see Varadarajan, 1974, §2.14).

(iv) The correspondence that we obtain in (7.9) by the use of exponential coordinates of the second kind of a simply connected soluble R-group Q gives a polynomial equivalence between Q and R^d. Let us recall that two Riemannian manifolds are *polynomially equivalent* if

$$M_1 \underset{\phi^{-1}}{\overset{\phi}{\rightleftarrows}} M_2,$$

where both φ and φ^{-1} are polynomial diffeomorphisms (see §7.1).

To see this, with the notation of (7.9), we observe that the mapping $Q \ni g \mapsto t_d$, the last coordinate, is polynomial since it can be identified with $Q \mapsto Q/Q_1$, where Q_1 is the subgroup generated by $e_1, \ldots, e_{d-1}$. Now $Q = Q_1 \leftthreetimes \{e_d\}$ and therefore $q \to q_1$, the Q_1-coordinate of q, is also polynomial (by (F.2) and §8.1.2). Then use induction on the dimension for Q_1.

We have more generally the following result.

Proposition *Let G be some connected R-group. Then G is polynomially equivalent to $\mathbb{R}^d \times G_0$, where G_0 is some maximal compact subgroup of G.*

The obvious converse that any group that is polynomially equivalent to $\mathbb{R}^d \times K$, for some compact manifold K, is an R-group clearly holds. To see this, we can for instance use $\gamma(r)$, the volume growth of G that has to be polynomial (see Guivarc'h, 1973; Jenkins, 1973).

The proposition is important because it is one of the ingredients in the proof of the (NB–Pol) theorem in §F.1. In particular, this proposition, combined with §C.3, gives a proof of the (NB–Pol) theorem for unimodular groups. We shall give the proof of the proposition in the next section.

We finish this section with an exercise that puts in perspective the exponential coordinates of the second kind of §7.3.

Exercise Prove in full generality (i.e. G is simply connected and solvable, but is not necessarily an R-group) that the correspondence obtained between G and $t = (t_1, \ldots, t_d)$ cannot in general be made polynomial (as above). However, with a choice of an appropriate basis $\xi_1, \ldots, \xi_d$ of the Lie algebra, as in §7.3, we can have exponential bounds; that is, we have $|g| \leqslant \exp(C|t| + C)$ and the converse.

The proof of this uses the ideas of Chapter 8 and is not trivial. Since we shall make no use of this fact we shall simply indicate the choice of the ordered basis $\xi_1,\dots,\xi_d$ of $\mathfrak{g}$ that is used and the interested reader can then work out the proof if they so wish.

As in §8.4 we let $\mathfrak{n}\triangleleft\mathfrak{g}$ be the nilradical of the Lie algebra of G and let $\mathfrak{h}$ be some Cartan subgroup. The ordered basis of $\mathfrak{g}$ that we need to use for the exponential coordinates will be $v_1,\dots,v_n,\eta_{n+1},\dots,\eta_d$ where $v_1,\dots,v_n\in\mathfrak{n}$ is a basis of $\mathfrak{n}$ and $\eta_{n+1},\dots,\eta_d\in\mathfrak{h}$. The first thing that needs proving, exactly as in §8.4.3, is that such a choice is possible. Once this is done, one sees that the coordinates $t_{n+1},\dots,t_d$, which correspond to the vectors $\eta_{n+1},\dots,$ depend polynomially on g. The last step is similar to what we did in §8.1.4. More explicitly, when $g=e(vt)e(\eta\tau)$ (notation as in (8.55)) then $e(vt)$ depends exponentially on g. To finish up, we use the fact that $\mathfrak{n}$ is nilpotent.

F.3 Amenable Groups

Let $\mathfrak{g}$ be the Lie algebra of an amenable connected Lie group G. Let $\mathfrak{n}\triangleleft\mathfrak{q}\triangleleft\mathfrak{g}$ be the nilradical and the radical of $\mathfrak{g}$. We then have the Levi decomposition $\mathfrak{g}=\mathfrak{q}\leftthreetimes\mathfrak{s}$, where $\mathfrak{s}$ is a semisimple algebra of compact type (see §3.1). It is important to recall here that this fact in the Levi decomposition can be used to define the amenability on the group; that is, G is amenable if and only if S, some – and therefore all – Lie group that corresponds to $\mathfrak{s}$, is compact (see Helgason, 1978, §II.6.9 and Varadarajan, 1974, §4.11.6; see also Appendix A). We have already pointed out that we can identify $\mathfrak{g}/\mathfrak{n}\simeq\mathfrak{a}\times\mathfrak{s}$ where $\mathfrak{a}\simeq\mathfrak{q}/\mathfrak{n}$ is Abelian. The product is direct rather than semidirect because $[\mathfrak{g},\mathfrak{q}]\subset\mathfrak{n}$ (see Varadarajan, 1974, §3.8.3). If we already know that the group G is simply connected, this fact about the Lie algebras implies that G/N is of the form $A\times S$, where here N is the nilradical of G and where A is a Euclidean space and S is some compact semisimple group. But quite generally, without the condition of simple-connectedness, we do have

$$G/N\simeq V\times K, \tag{F.3}$$

where $V\simeq\mathbb{R}^d$ is a vector space and K is again compact (but not necessarily semisimple).

Steps in the proof of (F.3) To see this let $(G/N)^{\sim}$ be the simple connected cover of G/N. Then $(G/N)^{\sim}\simeq A\times S$, where A is a Euclidean space and S is compact and semisimple. As a consequence, $G/N\simeq(A\times S)/\Gamma$ for some discrete central subgroup Γ. But since the projection of Γ on the factor S lies in the finite centre of S, the index $[\Gamma\colon A\cap\Gamma] < +\infty$ is finite. It follows that to obtain G/N we first factor $(G/N)^{\sim}/(A\cap\Gamma) = A_1 \times S$ where $A_1 \simeq \mathbb{R}^a \times \mathbb{T}^b$ is Abelian and then factor $(A_1 \times S)/F$ by some finite group $\simeq \Gamma/A\cap\Gamma$. But then $F \subset \mathbb{T}^b \times S$ and with $K \simeq (\mathbb{T}^b \times S)/F$ we have the required isomorphism (F.3).

Short exact sequences We shall preserve all our earlier notation, in particular that of (F.3), and denote by T the maximal central torus of G. Since T is central it is a central subgroup of N and it coincides with T_N, the maximal central torus of N (see §11.3.2). More explicitly, T_N is invariant by all automorphisms of N, and therefore T_N is a normal subgroup G and as a consequence $T_N = T$. On the other hand, the automorphism group of a torus is discrete and this implies that T_N is central in G, as asserted. We then have two exact sequences

$$0 \to N/T \to G/T \underset{\pi}{\to} G/N = V \times K \to 0, \tag{F.4}$$

$$0 \to Q^* = \pi^{-1}(V) \to G/T \underset{\alpha}{\to} K \to 0. \tag{F.5}$$

The group N/T is simply connected (see Varadarajan, 1974, §3.6.4) and so therefore is Q^*, which is clearly also soluble: we have $0 \to N/T \to Q^* \to V \to 0$. These properties of Q^* and the fact that K is compact imply that the exact sequence (F.5) splits and

$$G/T = Q^* \leftthreetimes K^* \tag{F.6}$$

for some compact subgroup $K^* \subset G/T$ that is mapped isomorphically on K by α (see Hochschild, 1965, §XII.3.2; Varadarajan, 1974, Exercise 3.37).

We shall denote by $\theta\colon G \to G/T$ the canonical projection and write $Q_1 = \theta^{-1}(Q^*)$, $K_1 = \theta^{-1}(K^*)$. We have

$$G = Q_1 \cdot K_1, \quad Q_1 \cap K_1 = T, \tag{F.7}$$

where T is now a central torus on Q_1 (in fact the maximal central torus in that group – see the 'proof' below), and where $Q_1/T = Q^*$ is simply connected. We can then construct the smooth coarse quasi-isometry of (11.37):

$$Q_1 \underset{\phi^{-1}}{\overset{\phi}{\rightleftarrows}} Q^* \times T. \tag{F.8}$$

It is also easy to see that K_1 is a maximal compact subgroup of G. This point is not very important, so we shall be brief.

A quick 'proof' Let $P \subset G$ be such a maximal compact subgroup. Since T is a central torus in G we must have $T \subset P$; but then $\theta(P) \subset Q^* \leftthreetimes K^*$ is a compact subgroup. Therefore $Q^* \cap \theta(P) = \{\mathrm{e}\}$ because Q^* is simply connected and soluble (see Varadarajan, 1974, Exercise 3.36). This means that $\dim\theta(P) \leqslant \dim K^*$ and thus $\dim P \leqslant \dim K_1$. The fact that the maximal compact subgroups are all conjugate finishes the proof.

What is important instead is the following 'twist' that we can make in the above construction. By (F.7), every element $g \in G$ can be represented as a product $g = q_1 k_1$, with $q_1 \in Q_1$, $k_1 \in K_1$. Modulo T, this representation is unique, that is, all other representations are $g = (q_1\tau)(k_1\tau^{-1})$ for some τ in the central subgroup T. We can then specify one particular such representation as follows. Denoting with a dot $\dot{g} = \theta(g) \in G/T$, we first consider $G \to G/T$ and decompose uniquely $\dot{g} = \dot{q}\dot{k}$, with $\dot{q} \in Q^*$, $\dot{k} \in K^*$, using the semidirect product (F.6). We now use the mapping $\sigma\colon Q^* \to \Sigma \subset Q_1$ for the 'section' that was defined in (11.36), and write $g = \sigma(\dot{q})[(\sigma(\dot{q}))^{-1}g]$ where clearly the cofactor $(\sigma(\dot{q}))^{-1}g \in K_1$. This therefore gives a decomposition of $g = \sigma k_1$ with $\sigma \in \Sigma$, $k_1 \in K_1$ which is clearly unique because if $\sigma k_1 = \sigma' k_1'$ then $\sigma, \sigma' \in Q_1$ have the same image in $Q_1 \to Q_1/T = Q^*$ by (F.6). We obtain in this way a smooth identification ψ:

$$G \underset{\psi^{-1}}{\overset{\psi}{\rightleftarrows}} Q^* \times K_1, \tag{F.9}$$

which clearly is a coarse quasi-isometry because we always have

$$|\mathrm{d}_G(\sigma(\dot{q}_1)k_1, \sigma(\dot{q}_2)k_2) - \mathrm{d}_{Q^*}(\dot{q}_1, \dot{q}_2)| \leqslant C,$$

when $\dot{q}_1, \dot{q}_2 \in Q^*$, $k_1, k_2 \in K_1$. This is just a consequence of the definition of the section Σ and the compactness of K_1; see §11.3.3 and Example 11.7.

Finally, let us specialise G to be an R-group, which, we recall, are amenable. In that case, by the semidirect product (F.6) and §F.2 (see the exercise), the mapping $G/T \ni \dot{g} \to \dot{q} \in Q^*$ is polynomial in the sense of §7.2. But again, by §F.2 and (11.36), the mapping σ is polynomial. Therefore in the identification (F.9) both ψ and ψ^{-1} are polynomial mappings. We have therefore obtained a proof of the proposition in §F.2.

F.4 Homotopy Retracts for Groups That Are Not Simply Connected

To make the exposition as light as possible we shall give the definitions and notation below in a slightly informal manner.

Let M be some Riemannian manifold with a base point O as in §7.1 and let $Y \subset M$ be some compact subset. We shall consider smooth, or at least locally

Lipschitz, homotopies as in §7.4 that have the following properties:

$$H(x,1)=x, \quad H(y,t)=y, \quad H(x,0)\in Y; \quad x\in M,\ y\in Y,\ 0\leqslant t\leqslant 1, \tag{F.10}$$

and as in §7.4 we shall impose the additional condition that

$$|\mathrm{d}H(x,t)|\leqslant C(1+|x|_M)^C; \quad x\in M,\ 0\leqslant t\leqslant 1, \tag{F.11}$$

for some $C>0$ with $|\ |_M$ denoting the Riemannian distance from the base point. If such a homotopy exists, for brevity we shall denote that homotopy by $H_{M,Y}(x,t)$ and simply say that $H_{M,Y}$ exists. If M is such that for all $a\geqslant 0$ we can find some compact subset $Y\subset M$ that contains the ball of radius a, (i.e. $[m\in M;\ |m|\leqslant a]\subset Y$) and such that $H_{M,Y}$ exists, we shall say that M has property-$\mathscr{H}$. In general, when $\mathscr{H}$ holds, several such Y exist and they are all of the same homotopy type. Informally, this says that the homotopy retracts M can be chosen so that it also leaves unchanged the points of an arbitrary compact subset.

Let us now specialise and assume that $M=G$ is some connected Lie group with its intrinsic Riemannian structure (see Chapters 1 and 7). Then from (F.1) it follows that if we assume that G has property-$\mathscr{H}$, then H_{G,G_0} exists for any maximal compact subgroup $G_0\subset G$. This holds because, if we use (F.1), we can follow $H_{G,Y}$ with some 'local homotopy' that shrinks Y into G_0 (see also the end of this appendix). The converse is also easy to see and we actually have $\mathscr{H} \iff (H_{G,G_0}\text{exists})$. This converse will not be needed and we shall not elaborate on this (see, however, the next exercise).

From this it follows in particular that if the group G has property-$\mathscr{H}$ then it is polynomially homotopically equivalent to the compact manifold G_0 and satisfies the condition of the theorem in §F.1.

Examples The following are clear:

(i) The direct product of two groups that have property-$\mathscr{H}$ also has property-$\mathscr{H}$.
(ii) Simply connected soluble NC-groups have property-$\mathscr{H}$. In that case, $G_0=\{e\}$. This follows from Theorem 7.10 and the special case of our previous assertion when $G_0=\{e\}$.
(iii) R-groups have property-$\mathscr{H}$. This follows from the proposition in §F.2.
(iv) Simply connected NB-groups have property-$\mathscr{H}$. This follows from Theorem 11.14 and (i), (ii).

Exercise Verify (ii) above. Let $B_1,B_2\subset G$ be the balls of radius 1 and 2 respectively. Furthermore, let $H_0(t,s)\in G$, for $0\leqslant t\leqslant 1$, $s\in G$, be given by the homology of Theorem 7.10 and be such that $H_0(0,s)=e$ and $H_0(1,s)=s$.

Then construct $H(t,x) \in G$, for $0 \leqslant t \leqslant 1$, $x \in G$, such that $H(t,s) = H_0(t,s)$, $H(t,b) = b$, $H(1,x) = x$, for $0 \leqslant t \leqslant 1$, $b \in B_1$, $x \in G$, $s \notin B_2$. For this we use the fact that G is diffeomorphic with a Euclidean space, and Federer (1969, §2.10.43). Then we follow H with some local homotopy that leaves the points of B_1 unchanged and shrinks $H(0,G)$ into B_1. Notice also that by Examples 7.5 and 7.6 the problem is much easier when G is simply connected and nilpotent because it reduces to the same problem on Euclidean spaces. This special nilpotent case suffices for the applications at the end of this subsection.

The aim of the rest of this appendix will be to give the proof of the following result.

Proposition *Every connected NB-group admits property-$\mathscr{H}$.*

This, as promised, will imply the NB–Pol theorem of §F.1.

As we pointed out, property-$\mathscr{H}$ behaves well under direct products. It turns out that it also behaves well under semidirect products provided that an additional condition is imposed, which we now explain.

Let $G = N \leftthreetimes M$ be the semidirect product of two connected Lie groups that each possess property-$\mathscr{H}$. Let $\mathfrak{g} = \mathfrak{n} \leftthreetimes \mathfrak{m}$ be the corresponding semidirect product of the Lie algebras. We shall make the following hypothesis which we shall call the 'glueing condition' for the semidirect product.

Glueing condition *There exists $\xi \in \mathfrak{m}$ that has the following two properties*:

GC1: *The characteristic roots $\lambda_1, \ldots, \lambda_n$ of the action* $\operatorname{ad}\xi$ *on* $\mathfrak{n}$, *with* $n = \dim \mathfrak{n}$, *satisfy* $\operatorname{Re}\lambda_j > 0$ *for* $1 \leqslant j \leqslant n$.

GC2: *The characteristic roots $\mu_1, \ldots, \mu_m$ of the* $\operatorname{ad}\xi$ *action on* $\mathfrak{m}$, *with* $m = \dim \mathfrak{m}$, *satisfy* $\operatorname{Re}\mu_j \geqslant 0$ *for* $1 \leqslant j \leqslant m$.

In §F.5 below, we shall prove the following result.

The glueing lemma *Let $G = N \leftthreetimes M$ be the semidirect product of two connected Lie groups. Assume that both N and M satisfy the $\mathscr{H}$-property and that their Lie algebras satisfy the glueing condition. Then G satisfies the $\mathscr{H}$-property.*

Before we give the proof let us go back to Appendix B and Appendix C. For any NB-group G we saw in (C.2) that $G = N \leftthreetimes M$, where both N and M satisfy the $\mathscr{H}$-condition. To see this for N we use part (ii) of the above example. For M we use (C.5) and parts (i) and (iii) of the example. For the second factor NA in the right-hand side of (C.5) we can use part (ii) or we can again use the glueing lemma because now we are in a situation where the $\mathscr{H}$-property is clearly satisfied for the factors A and N (see the previous exercise).

Back to our proposition. We shall assume that the NB group of the proposition is not unimodular and we shall use §B.6. In the unimodular case, as we have already pointed out, the proposition in §F.2 directly implies the (NB–Pol) theorem.

Now the Lie algebra $\mathfrak{m}$ of M is $\mathfrak{g}_R \oplus \mathfrak{s}_n$ (see Appendix B). Since Hypothesis (H) of §B.7 is satisfied, the vector $\xi \in \mathfrak{q}_R$ (which is the radical of $\mathfrak{g}_R$) in Hypothesis (H) satisfies GC1. Condition GC2 is automatically satisfied since $\mathfrak{g}_R$ is an R-algebra and as a consequence $\operatorname{Re}\mu_j = 0$ (see Exercise F.2(i)). The final conclusion is that the above glueing lemma applies and completes the proof of our proposition.

F.5 Proof of the Glueing Lemma

F.5.1 The exponential retract

We shall start from a more general situation where G is some connected Lie group that we can retract to some compact subset $P \subset G$, that is, $H(g,1) = g$, $H(p,t) = p$, $H(g,0) \in P$, where $g \in G$, $p \in P$, $0 \leqslant t \leqslant 1$, as in (F.10). But now instead of (F.11) we consider the following weaker exponential bound:

$$|\mathrm{d}H(g,t)| \leqslant C\exp\left(C|g|^C\right); \quad g \in G, \tag{F.12}$$

for some $C > 0$. It is 'probably provable' (with the methods of Chapters 8 and 11) that such a retract always exists, even with an exponent $C = 1$ on $|g|$ (see the final 'comment' at the end of this appendix). This general fact (if correct), however, will not be needed here because in what follows we shall stick to groups for which this exponential retract comes for free.

The perturbation This perturbation of the homotopy H will depend on two things. First, on some $\xi \in \mathfrak{g}$ the Lie algebra of G that has the property that $\rho_1, \dots$, the characteristic roots of $\operatorname{ad}\xi$ on $\mathfrak{g}$, have non-negative real parts: $\operatorname{Re}\rho_j \geqslant 0$, for $j = 1, \dots$. For the same reasons as in Exercise F.2(i), this is equivalent to the fact that the norm of $\operatorname{Ad}\exp = e^{\operatorname{ad}}$ satisfies

$$\|\operatorname{Ad} e^{-t\xi}\| \leqslant C(1+t)^C; \quad t > 0 \tag{F.13}$$

for appropriate constants.

The other ingredients on which our perturbation of H will depend are two C^∞ functions, $0 \leqslant \alpha(t), \beta(t) \leqslant 1$ for $0 \leqslant t \leqslant 1$ with the following properties:

$$\left.\begin{array}{l}\beta(0) = \beta(1) = 0; \quad \beta(t) \equiv 1; \quad t \in \left[\frac{1}{10}, \frac{9}{10}\right],\\ \alpha(t) = 0 \text{ for } t \in \left[0, \frac{4}{10}\right]; \quad \alpha(t) = 1 \text{ for } t \in \left[\frac{6}{10}, 1\right].\end{array}\right\} \tag{F.14}$$

This means that β is constant when it reaches its 'top' value in a large central interval, and α 'picks up' all its increment in a much smaller central interval (see §8.2.3 where a very similar construction was carried out).

With a large value of $C_0 > 0$, to be chosen later, we shall then define the required perturbation by

$$H_0(g,t) = H(g,\alpha(t))\theta(g,t), \quad \theta(g,t) = \exp\left(C_0\beta(t)|g|^{C_0}\xi\right); \quad g \in G,\ 0 \leqslant t \leqslant 1. \tag{F.15}$$

With this notation H_0 is the product in G of H and θ. We can then use formula (8.6) for the differentials to obtain

$$\mathrm{d}H_0(g,t) = (\mathrm{Ad}\,\theta(g,t))^{-1}\mathrm{d}H + \mathrm{d}\theta(g,t). \tag{F.16}$$

Here, the bounds

$$\|\,\mathrm{Ad}(\theta(g,t))^{-1}\| \leqslant C(1+|g|)^C; \quad |\mathrm{d}\theta| \leqslant |g|^C + C \tag{F.17}$$

hold for appropriate constants, by (F.13) and the fact that $\big|\mathrm{d}|g|\big| \leqslant C$. In computing $\mathrm{d}\theta$ observe that θ is a time-changed one-parameter subgroup.

Straightaway therefore we can deduce from (F.16) and (F.17) that the following polynomial bound holds:

$$|\mathrm{d}H_0(g,t)| \leqslant C(1+|g|)^C; \quad g \in G,\ t \in \left[0,\tfrac{4}{10}\right] \cup \left[\tfrac{6}{10},1\right], \tag{F.18}$$

for some $C > 0$.

Indeed, in the interval $\left[0,\frac{4}{10}\right]$, the first factor in (F.15) reduces to e. In the interval $\left[\frac{6}{10},0\right]$ this first factor is identical to g and the differential $\mathrm{d}H$ stays bounded. We can therefore use (F.16). This is parallel with what we did in §8.2.3: the choice of ξ, together with §3.9.1, is used.

The issue is therefore to prove (F.18) in the central interval $\left[\frac{4}{10},\frac{6}{10}\right]$ and for this, additional conditions in G clearly have to be imposed.

F.5.2 A special exponential retract

The additional conditions that we must impose on G are that it satisfies the conditions of the glueing lemma of §F.4. Here, $G = N \leftthreetimes M$ and the notation of the previous subsection is retained.

We shall also denote by H_N, H_M the two polynomial retracts on the groups N and M that satisfy (F.10) and (F.11), where P in §F.5.1 is taken to be some large compact subset in N or M, as the case may be. Then using the same notation (which is not standard) for the product in G of two elements (one from N and the other from M) as in (8.54) we shall define the following homotopy on G:

$$H(g,t) = H_N(n,t) \leftthreetimes H_M(m,t); \quad n \in N,\ m \in M,\ g = n \leftthreetimes m,\ 0 \leqslant t \leqslant 1. \tag{F.19}$$

By the definition of H_N, H_M, this does retract G to some compact subset as needed and we shall prove in the next few lines that it satisfies the properties of §F.5.1. For this we apply as before the formula from §8.1.3 to obtain

$$\mathrm{d}H(g,t) = \mathrm{Ad}\, H_M^{-1}\big(\mathrm{d}H_N(n,t)\big) + \mathrm{d}H_M(m,t). \tag{F.20}$$

From this we can easily verify (F.12). The only point that has to be treated with care in this verification is that the differentials in (F.20) are taken with respect to the variables g and t. Therefore, on the right-hand side of (F.20) we are considering the compositions of $\mathrm{d}_{n,t}H_N \circ \mathrm{d}_g n$ and $\mathrm{d}_{m,t}H_M \circ \mathrm{d}_g m$. (To be formally accurate, if slightly pedantic, we should write $\mathrm{d}_{g,t}(n,t) = \mathrm{d}_g n \oplus \text{Identity}$ here, rather than $\mathrm{d}_g n$, and the same thing for $\mathrm{d}_g m$)

The mapping $g \mapsto m$ given by the splitting $g = n \leftthreetimes m$ is a homomorphism, so the second composition of differentials that we need gives no problem because of the polynomial properties of $\mathrm{d}_{m,t}H_M$. For the first composition of differentials we have

$$|\mathrm{d}_{n,t}H_N(n,t)| \leqslant C(1+|n|)^C; \quad n \in N, \tag{F.21}$$

$$\left.\begin{aligned} |\mathrm{d}_g n(g)| &\leqslant C\exp(C|g|), \\ |n(g)| &\leqslant C\exp(C|g|); \quad g \in G. \end{aligned}\right\} \tag{F.22}$$

Inequality (F.21) follows by the hypotheses on H_N, and for (F.22) we use $n = gm^{-1}$ and §§8.1.2, 8.1.3, 8.1.4. If we use (F.21), (F.22) and (F.20), estimate (F.12) follows at once. For this we must observe in addition that $|H_M(g,t)| \leqslant |g|^C + C$ because H_M is polynomial, and then use the general fact $\|\mathrm{Ad}\, h\| \leqslant c\exp(c|h|)$, for $h \in G$, where c depends only on the group: see (8.8). This explains the $|g|^C$ in the exponential in (F.12) and, to avoid this, something special clearly has to be done. It was avoided in Example 8.1 only by imposing additional conditions; see, however, the final comments of this appendix.

F.5.3 The perturbations and the proof of the glueing lemma

We shall preserve the notation of the previous subsection and construct the perturbation (F.15) with a $\xi \in \mathfrak{m}$ that satisfies conditions GC1 and GC2 of the glueing lemma and where $H(g,t)$ is as constructed in (F.19). Formulas (F.16) and (F.20) will be combined, but now we shall restrict ourselves to the central range $t \in \left[\frac{4}{10}, \frac{6}{10}\right]$ that was left undone in §F.5.1. In that range $\theta(g,t) = \exp(C_0|g|^{C_0}\xi)$ and $\mathrm{d}H_0(g,t)$ is therefore the sum of three terms that will be examined separately as follows. For constants C that depend on C_0 we have

(i) $|\mathrm{d}\theta| \leqslant C|g|^C + C$ as before;

(ii) $|(\mathrm{Ad}\,\theta^{-1})\mathrm{d}H_M| \leqslant C(1+|g|)^C$: this follows from the polynomial property (F.11) that holds for M, combined with the facts that $g \mapsto m$ in the decomposition $g = n \leftthreetimes m$ is a homomorphism, and that from GC2, $\|\,\mathrm{Ad}\,\theta^{-1}\big|_{\mathfrak{m}}\| \leqslant C(1+|g|)^C$.

(iii) $\mathrm{Ad}\,\theta^{-1}\,\mathrm{Ad}H_M^{-1}\mathrm{d}H_N$ remains, and this can be estimated by the product of the three norms

$$\|\,\mathrm{Ad}\,\theta^{-1}\big|_{\mathfrak{n}}\|\,\|\,\mathrm{Ad}H_M^{-1}\|\,|\mathrm{d}H_N|; \tag{F.23}$$

we saw in §F.5.2 that the product of the last two factors in (F.23) can be estimated by $C_1\exp(|g|^{C_1})$, with $g \in G$, for some $C_1 > 0$ that does not depend on C_0 of (F.15).

On the other hand, by GC1 and, of course, the use of the lemma in §3.9.1, we deduce that there exists $c_1 > 0$ such that the first factor in (F.23) can be estimated by

$$C(1+|g|)^C\exp(-c_1C_0|g|^{C_0}),$$

where here C depends on C_0 but $c_1 > 0$ does not. If we combine these estimates and make the choice of C_0 sufficiently large, we finally conclude that the norm of the term (iii) is bounded in the range $t \in \left[\frac{4}{10}, \frac{6}{10}\right]$.

The bottom line is that (F.11), as well as the first and third conditions in (F.10), holds for H_0. The second condition in (F.10), however, breaks down. To handle the problem and guarantee that $H(y,t) = y$, with $y \in Y$, $0 \leqslant t \leqslant 1$, we must modify the perturbing factor θ and make it $\equiv e$ when $|g|$ is small. To fix ideas we could for instance set $\theta(g,t) = \exp(C_0\beta(t)|g|_P^{C_0}\xi)$ where $|g|_P =$ (distance of g from P) and $P \subset G$ some large compact subset. The verification that all the above estimates for this new perturbation still hold will be left to the reader.

With this we have completed the proof of the glueing lemma and have shown that an NB-group admits the $\mathscr{H}$-property.

An open problem Let us restrict ourselves to soluble groups: we saw in §11.3.4 that, for C-groups at least, it is not always possible to make them polynomially equivalent to a group of the form $U \times T$, where U is simply connected and T compact. We have, on the other hand, seen that this is possible for R-groups. Whether it is possible for NC-groups is unclear. We can go a step further and ask the same thing for general NB-groups (not necessarily soluble). This general problem may be difficult to tackle but, on the other hand, there may be interesting classes of groups for which there is a positive answer. The techniques that we have developed for our specific purposes may lead to results of that type but we shall leave it at that.

A comment on the construction If we use further structure theorems for a simply connected soluble R-group G we can construct a polynomial retract H on G that, in addition, satisfies $|H_M(g,t)| \leqslant C|g| + C$ (see Example 8.1). This certainly presents some independent interest and for us here it allows us to dispose with the artificial exponent C in $|g|^C$ on the right-hand side of (F.12). Since, however, the structure theorems needed for this are non-trivial, the use of $|g|^{C_0}$ in the perturbing factor θ was, we felt, preferable (although, admittedly, it is ugly to look at). No doubt this special type of polynomial retract can be constructed for any NB-group and this completes the picture nicely. Some indication of how to go about this is given in Varopoulos (2000b, §§5.3–5.4), if the group is simply connected; the general problem is, however, 'esoteric' and technical so we shall say no more.

Another point worth noting is that the use we made at the beginning of §F.4 of the general fact (F.1) can be avoided. One way to see this is to build the required condition explicitly in the construction of the two homotopies H_N, H_M of §F.5.2. The details are left to the reader.

PART III

HOMOLOGY THEORY

12

The Homotopy and Homology Classification of Connected Lie Groups

12.1 An Informal Overview of the Chapter and of Part III

12.1.1 A review of what has already been achieved in the geometric classification

The task we set off to accomplish was to give geometric conditions on a general connected Lie group G that characterise the B and NB conditions. This, one could argue, has been achieved already. Indeed, we showed in Chapter 11 that G is coarse quasi-isometric to some soluble simply connected Lie group Q, and Q is B or NB at the same time as G. These soluble simply connected Lie groups should be considered as the basic building blocks of the theory. For these the C–NC conditions have been characterised geometrically in the course of Chapters 7–10 and this was done by homotopy considerations.

This way of going about things is, however, unsatisfactory. Quite apart from having to go about it in two stages, we are mixing here the coarse quasi-isometry which is a rough (almost discrete) equivalence relation $G \simeq Q$ with the smooth homotopy properties of Q.

The aim of the third part of the book is to overcome this objection and obtain the geometric classification in one go on G by homotopic or homological considerations.

12.1.2 The use of the Poincaré equation

For a soluble simply connected group Q we have indicated already (without proofs) in Chapter 7 how the Poincaré equation $d\theta = \omega$ for closed smooth forms ω (closed means that $\mathrm{d}\omega = 0$) can be used to characterise the C–NC condition.

Let us recap. Let $\omega = \sum a_i \mathrm{d}x_i + \sum a_{ij} \mathrm{d}x_i \wedge \mathrm{d}x_j + \cdots$ be, in local coordinates, a closed smooth differential form with vanishing constant component. Global coordinates could also have been used here because $Q \cong \mathbb{R}^n$ (diffeomorphism). From this it also follows that ω can be written $\mathrm{d}\theta = \omega$ for some other differential form θ. This is standard for $\mathbb{R}^n$ (see Warner, 1971 or de Rham, 1960, which is the reference that we shall follow very closely in Part III of the book).

The issue now is this. Assume that ω grows polynomially. This notion was used in Chapter 7 and it will be developed in detail in this chapter. Is it then possible to solve the equation $\mathrm{d}\theta = \omega$ by a form θ that also grows polynomially? It turns out that this is possible if and only if Q is an NC-group.

This is one of the main results in the geometric theory and it will be the topic in this and the next chapter.

12.1.3 Homology and the Poincaré equation for general connected Lie groups

This characterisation of the NC-condition cannot possibly hold as stated in the previous subsection for general groups. The obvious example is the torus $\mathbb{T} = [\varphi \in \mathbb{R};\ \mathrm{mod}\, 1]$ where the form $\omega = \mathrm{d}\varphi$ cannot be represented by $\omega = \mathrm{d}f$ for some smooth function f on $\mathbb{T}$. This example already shows that the problem lies with compact groups.

To explain further we first invoke a deep fundamental theorem (see Hochschild, 1965, §XV.3.1).

Theorem 12.1 *Let G be some connected Lie group. Then there exist some compact Lie group K, some Euclidean space $E = \mathbb{R}^n$ and φ some diffeomorphism such that*

$$E \times K \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} G. \tag{12.1}$$

12.1.3.1 Maximal compact subgroups It is possible to make the construction of (12.1) in such a way that $\varphi(0 \times K) = K_0 \subset G$ is a compact subgroup that is maximal. This is equivalent, here, to saying that if $L \subset G$ is some compact subgroup then there exists $a \in G$ such that $aLa^{-1} \subset K_0$. In relation to this last point, observe that from (12.1) it follows that K_0 is connected.

No direct use will be made of this difficult result but it will be an important guiding line for the theory that will be developed in Part III.

One important consequence of (12.1) is that we can use the radial retract of the Euclidean space E to 0 and this induces a retract of G to the maximal compact subgroup K_0. This means that there exists some smooth function $F(g,t)$ of the two variables $g \in G$ and $0 \leqslant t \leqslant 1$ such that $F(g,1) = g$, $F(k_0,t) = k_0$ $(k_0 \in K_0)$ and $F(g,0) \in K_0$.

The question arises of how fast we have to move to achieve this, that is, how big the gradient ∇F of F has to be. This gradient is taken in the left-invariant Riemannian structure of G. When ∇F can be controlled by a polynomial in terms of $|g|$, the distance from the identity, we shall say that G admits a polynomial retract to a compact. This notion generalises the notion of the polynomially retractable simply connected soluble Lie groups Q where then K_0 reduces to the identity $\{e\}$. The main result in Chapters 8–10 was to characterise the NC-condition on Q by the existence of a polynomial retract to a point. This fact generalises to a general group G and it will be examined in detail in §12.2 below.

What is even more important from our point of view is that (12.1) implies that $H(G,\mathbb{R}) \cong H(K,\mathbb{R})$. Explicitly, this says that the homology of G is identical to the homology of K and is therefore a finite-dimensional vector space. This can among other things be seen by the retract F, and can be reformulated in terms of the Poincaré equation as follows.

There exist finitely many smooth closed differential forms $\Omega_1,\dots,\Omega_m$ on G such that for every smooth closed differential form ω on G there exist some smooth differential form θ and scalars $\lambda_1,\dots,\lambda_m \in \mathbb{R}$ such that

$$\omega = \mathrm{d}\theta + \lambda_1\Omega_1 + \cdots + \lambda_m\Omega_m. \tag{12.2}$$

Let us now restrict ourselves to the subspace of closed smooth differential forms that grow polynomially. The question then arises whether the Ω_j can be chosen once and for all as above so that they have polynomial growth and such that the θ that solves (12.2) can also be chosen to be of polynomial growth when ω is of polynomial growth.

The answer is striking: this can be done if and only if G is an NB-group.

The proof of this fact when G is simply connected will be done with the methods that we shall develop in this chapter and the next. To get round the simple-connectedness, as is often the case with questions involving homology, is a difficult problem and this will be the subject matter of Chapter 14.

12.1.4 The homology on manifolds. The use of currents

The homology $H(G;\mathbb{R})$ of the previous subsection that will be used will be the homology developed by de Rham (1960) in his pioneering work. This is based on the systematic use of smooth differential forms on a manifold M and their dual space which is the space of currents on M. The point of view that is taken in de Rham (1960) is that currents are generalised forms. This is like the distributions on $\mathbb{R}^n$ that, as well as being elements of a dual space (see Schwartz, 1957) could and should be considered as generalised functions.

Ideally, before embarking on the chapter the reader should spend some time looking at the first three chapters of de Rham (1960), which are very readable. One way or another, familiarity with currents will be essential in this chapter. The reader who tries to tackle the second half of this chapter and Chapter 13 below and get away with just the overview that we gave in §10.2 may find several points hard going. To help the reader, a further overview on currents and how to read de Rham (1960) will be given in this chapter.

12.1.5 Content of Chapter 12

In this subsection we shall give a more detailed description of Chapter 12 and of how it stands with respect to Part III. This should help the reader cope with a chapter that is unusual in the sense that what it mostly does is collect together and try to synthesise background material from several branches of algebra, analysis and geometry.

The sections in this chapter fall into three groups:

first group: §§12.2–12.6;
second group: §§12.7–12.13;
third group: §§12.14–12.17.

First group This is easy to read and with it one can find what Part III is all about. The first thing we do is formalise some notions that we have encountered before from the theory of currents, de Rham cohomology and homotopy theory. In particular, the pivotal definition of polynomial growth in de Rham cohomology is hammered out in its final form.

With this background material we are able to state precisely the theorems that will be proved in Part III, which are among the most important results in the subject. In fact, the results on polynomial homotopy come straightaway in §12.2.

The proofs of these theorems will be given in Chapter 13 for the special case of soluble simply connected groups (these are the 'models' of §1.5) and in the general case in Chapter 14. For the latter, some sophisticated ideas from algebraic topology have to be recalled and we shall say much more about this when we come to it.

Second group This is considerably more technical and gives background material from the different subjects:

(i) further aspects from the theory of currents – homotopies and regularisation of currents;

(ii) some basic definitions from homological algebra and algebraic topology;

(iii) some notions from the theory of topological vector spaces that go beyond the normed and Banach spaces that everyone knows.

In the proofs of Chapter 13, these subjects and ideas are blended. Because of this, in an effort to save space, we have tended to mix these topics in a manner that is somewhat artificial. A more compartmentalised exposition would have been more satisfactory, but this would have been longer and, given the size of the book already, we felt we could not afford to do that.

Third group Here we process the notions from the second group so that they will be ready to be used in Chapter 13.

From the above, it should be clear that only those readers that intend to study carefully the proofs of Chapters 13–14 need to go any further than the first group in this chapter.

12.2 Definitions and the Main Theorem Related to Homotopy

We shall first collect, in a systematic and unified manner, some of the definitions already given in Chapter 7.

12.2.1 Homotopies. Homotopic equivalence

Let $f, g\colon X \to Y$ be two continuous mappings between two topological spaces. We say that $f \simeq g$ are homotopic if there exists a continuous mapping F,

$$F\colon X \times [0,1] \longrightarrow Y; \quad F(x,1) = f(x),\ F(x,0) = g(x),\ x \in X. \tag{12.3}$$

We say that two topological spaces are *homotopically equivalent*, written $X \simeq Y$, if there exist continuous mappings α, β,

$$X \xrightarrow{\alpha} Y \xrightarrow{\beta} X, \tag{12.4}$$

such that $\beta \circ \alpha \simeq$ the identity mapping on X and $\alpha \circ \beta \simeq$ the identity mapping on Y. Both the above $\simeq$ are equivalence relations. Hilton (1953) is a good reference for these notions.

12.2.2 Homotopy retracts

Let $X \supset A$ be some connected topological space and some subspace. We say that X retracts (homotopically) to A if there exists some homotopy $F(x,t)$ as in (12.3) such that

$$F(x,1) = x, \quad F(x,0) \in A; \quad x \in X. \tag{12.5}$$

This is a very weak notion and it does not even imply that $X \simeq A$, but we shall make systematic use in the following two cases:

(1) $A = \{O\}$ for some point $O \in X$. We then say that X retracts to a point.
(2) $A \subset X$, some compact subset. In this case we say that X is retractable to a compact set.

Exercise Try to convince yourself, using ideas from the exercise in §14.1.2, that only a very restrictive class of manifolds are retractible to a compact set in the above sense. This is not easy to see but has nothing to do with the subject of the book. On the other hand, by Theorem 12.1, connected Lie groups *do* admit this property.

12.2.3 Smooth manifolds

The above definitions will be applied to C^∞ manifolds which, more often than not, will in fact be Lie groups. In that case all mappings considered will be smooth or at least locally Lipschitz (see §7.1). This assumption will be made tacitly throughout without further mention.

A number of technical issues concerning the above definitions will also be tacitly ignored. Here is an example. Assume that X, Y are C^∞ manifolds and that the two mappings f, g in (12.3) are C^∞. Assume also that the homotopy F in (12.3) is locally Lipschitz. We can then regularise and obtain some smooth homotopy $\widetilde{F}$ between f and g that still satisfies (12.3). This fact is necessary if we wish to link up two smooth homotopies $f \simeq g$ and $g \simeq h$ and prove that a smooth homotopy is an equivalence relation. To write down a proof of this in full generality might take some doing.

In our case, however, where the homotopies considered will be constructed explicitly, the smoothing whenever necessary will be seen to be automatic.

This and similar issues will therefore be bypassed without mention.

12.2.4 Riemannian manifolds and polynomial mappings

Let us first go back to a definition that we introduced in §7.1. Let M be some C^∞ connected Riemannian manifold and let $\mathrm{O} \in M$ be some base point. Let M_1

be some other C^∞ Riemannian manifold and let f be some smooth, or at least locally Lipschitz mapping

$$f\colon (M,\mathrm{O}) \longrightarrow M_1. \tag{12.6}$$

We say that f is polynomial if

$$|df(m)| \leqslant C(1+|m|)^C; \quad m \in M, \tag{12.7}$$

where $|m|$ is the Riemannian distance from m to O, for constants C independent of m. Note that for a Lipschitz mapping, df is defined almost everywhere; see Federer (1969, §3.1.4). Note also that the 'a.e.' has been suppressed in (12.7). This will be standard practice in similar situations where the 'a.e.' is obvious. The norm $|\ |$ on the left of (12.7) is of course taken with respect to the Riemannian norm on the tangent spaces.

We saw in §7.1 that if we fix some base point $\mathrm{O}_1 \in M_1$ and write $|m_1|_1 = d_1(\mathrm{O}_1, m_1)$ for the distance in M_1 then there exist constants such that

$$|f(m)|_1 \leqslant C(1+|m|)^C; \quad m \in M. \tag{12.8}$$

This fact implies that the composition of two polynomial mappings,

$$(M,\mathrm{O}) \xrightarrow{f} (M_1, \mathrm{O}_1) \xrightarrow{g} M_2, \tag{12.9}$$

is polynomial.

Similarly, if $f, g\colon (M,\mathrm{O}) \to M_1$ are two polynomial mappings we say that they are polynomially homotopic if there exists F, a homotopy as in (12.3), that in addition satisfies

$$|dF(m,t)| \leqslant C(1+|m|)^C; \quad m \in M,\ 0 \leqslant t \leqslant 1, \tag{12.10}$$

for appropriate constants, where $|dF|$ is taken for the product Riemannian structure on $M \times [0,1]$ and where also F is assumed to be locally Lifschitz for the product distance.

Similar definitions are given for two polynomially homotopically equivalent manifolds $(M,\mathrm{O}) \simeq (M_1, \mathrm{O}_1)$ and a polynomial retract of (M,O) to $A \subset M$.

These notions are clearly invariant by Riemannian quasi-isometries (see §1.4.1) and by a change of the base point, although then the constants in say (12.8) or (12.10) depend on the base point.

Extensive use will be made of the above notions when the manifolds are connected Lie groups assigned with their left-invariant metrics (see §1.4). In that case, more often than not, we shall tacitly take as a base point the identity element $e \in G$. And, of course, the particular choice of the left-invariant Riemannian metric used makes no difference to the definitions.

12.2.5 Simply connected groups

With the previous definitions we can now express the essential content of Chapters 7–10 as follows.

Theorem 12.2 *Let G be some simply connected soluble Lie group. Then G is an* NC*-group if and only if it is polynomially retractable to $e \in G$, or in other words, if and only if it is polynomially homotopically equivalent to the one-point space $\{e\}$.*

To put this result in the correct perspective, we shall start from two elementary exercises on the Euclidean space $\mathbb{R}^d$ that will be left for the reader to do.

Exercise 12.3 Let $c < C$ be two constants; then there exists some smooth homotopy $F_1(x,t)$ $(x \in \mathbb{R}^d,\ 0 \leqslant t \leqslant 1)$ such that

$$\begin{aligned} &F_1(x,1) = x; \quad x \in \mathbb{R}^d, \\ &F_1(x,0) = \mathrm{O} = \text{ origin of } \mathbb{R}^d; \quad |x| \leqslant c, \\ &F_1(x,t) = x; \quad |x| \geqslant C,\ 0 \leqslant t \leqslant 1. \end{aligned} \tag{12.11}$$

This says that we can continuously shrink the c-ball to the origin O without moving the points that are far out at all.

Exercise 12.4 Let $\gamma(t) \geqslant 0$ $(0 \leqslant t \leqslant 1)$ be continuous with $\gamma(1) = 0$, and $C \geqslant \sup \gamma(t) + 1$. Then there exists a homotopy $F_2(x,t)$, for $x \in \mathbb{R}^d, 0 \leqslant t \leqslant 1$ such that

$$\begin{aligned} &F_2(x,1) = x; \quad x \in \mathbb{R}^d, \\ &F_2(x,t) = x; \quad |x| \geqslant C,\ 0 \leqslant t \leqslant 1, \\ &F_2(x,t) = \mathrm{O}; \quad |x| \leqslant \gamma(t),\ 0 \leqslant t \leqslant 1. \end{aligned} \tag{12.12}$$

Here we are as in Exercise 12.3, but the shrinking of the small balls is progressive with time and it is a matter of constructing a continuous function of x and t such that $F(x,t) = x$ for $|x| \geqslant C$, $F(x,1) = x$ for $|x| \leqslant C$ but $F(x,t) = \mathrm{O}$ for $|x| \leqslant \gamma(t)$.

We now recall that every soluble simply connected group Q is diffeomorphic with some Euclidean space; see Varadarajan (1974, §3.18.11). This fact, combined with Theorem 12.2 and Exercises 12.3 and 12.4, leads to the following two facts.

Theorem 12.5 *Let Q be some soluble simply connected group. Then Q is* NC *if and only if Q is polynomially retractable to a compact.*

It suffices for this to follow the original contracting homotopy to a compact set by the homotopy F_1 of Exercise 12.3. Furthermore, let $F(g,t) \in Q$ be some polynomial homotopy such that $F(\mathrm{O},1) = F(\mathrm{O},0) = \mathrm{O}$. Since Q is diffeomorphic to $\mathbb{R}^d$ we can use (12.12) and compose $F_2(F(g,t),t) = \widetilde{F}(g,t)$ and we have a new polynomial homotopy for which the base point stays put:

$$\widetilde{F}(\mathrm{O},t) = \mathrm{O}; \quad 0 \leqslant t \leqslant 1. \tag{12.13}$$

In this context, note that the original shrinking homotopy that we constructed for NC-groups in Appendix F and §8.5.2 already satisfies (12.13).

A consequence of Theorem 12.5 and the above observation is the following important theorem.

Theorem 12.6 *Let G be some simply connected Lie group. Then G is NB if and only if G is polynomially retractable to a compact.*

To prove this we use the previous theorem and in addition Theorem 11.14, which says that the group G of Theorem 12.6 is smoothly quasi-isometric to $Q \times K$ where Q is simply connected soluble and K is compact. And G is NB if and only if Q is. But also clearly G is polynomially contractible to a compact set if and only if Q is. This gives Theorem 12.6.

Exercise 12.7 For the last point, if we can retract $G = Q \times K$ to a compact $P \subset G$ by a polynomial homotopy $H(g,t)$, and if $Q \xrightarrow{i} G \xrightarrow{p} Q$ are the canonical injections and projections then $p[H(i(q),t)] = H_Q(q,t)$ is a homotopy that retracts Q to a compact set.

If we use Appendix F we obtain the following more general result.

12.2.6 Retract to a maximal compact subgroup

Theorem 12.8 *If G is a connected* NB*-group, then there exists $K_0 \subset G$ some compact subgroup and a polynomial retract $F(g,t)$ of G onto K_0 that has the additional property that*

$$F(k,t) = k; \quad k \in K_0,\ 0 \leqslant t \leqslant 1. \tag{12.14}$$

It follows in particular that G is a polynomially homotopically equivalent to K_0. Furthermore, the compact subgroup K_0 can be chosen to be a maximal compact subgroup of G. This last point is not of great significance; it was discussed in §F.4 and we shall not come back to it again.

12.2.7 General connected Lie groups

One of the main results of Part III is that we have a converse of Theorem 12.8 and thus a homotopy B–NB characterisation.

Theorem 12.9 *Let G be some connected Lie group that can be retracted polynomially to a compact. Then G has to be an* NB*-group.*

As a consequence, a connected Lie group G is NB *if and only if G is polynomially retractable to a compact set.*

The hypothesis of the theorem is as in §12.2.2 and it does not automatically imply the existence of a homotopy that satisfies (12.14). There is an interesting gap in the above 'if and only if'. This, among other things, will be examined in Chapter 14. For this, one uses (12.1), which is quite deep and is not proved in this book, combined with (14.5) and the discussion at the end of §14.1.2.

Exercise 12.10 Prove a reduction. For the proof of the first part of Theorem 12.9 we may assume that G is soluble. This reduction illustrates things that we have already done but it will not be used for the proof of the theorem. To prove that reduction we use §11.1.5 and find new groups $\widetilde{G}$, Q, K, with Q soluble and K compact, and a quasi-isometry $\simeq$ such that

$$\widetilde{G} \simeq Q \times K, \qquad \widetilde{G}/F = G, \tag{12.15}$$

where F is a central finite subgroup.

Now it is elementary to see, and quite in the spirit of fibre spaces (see Hilton, 1953, §V.1 or Steenrod, 1951, §11), that the polynomial homotopy that retracts G to a compact can be lifted to a similar polynomial homotopy on $\widetilde{G}$. If we intertwine this homotopy with i and p as in Exercise 12.7 we see that the soluble group Q is already polynomially contractible to a compact set. As a consequence at this point, if we can conclude from this that Q is NB, then the same thing will hold for the original group G because of Theorem 11.16. Note that a similar use of §11.1.5 will be made in Exercise 12.22 down the road but in the context of homology theory.

Note finally that the complete dénouement for Theorem 12.9 will only happen in Chapter 14.

12.3 A Review of the General Theory of Currents: A Reader's Guide to the Literature

12.3.1 Scope of §§12.3, 12.4

This section and the next will be a rapid review of the general theory of currents and how they are used to define a homology theory on C^∞ manifolds. The key reference in the subject is de Rham (1960). The style of the book is old-fashioned but we have found this to be a great advantage, especially in its first part, which has aged very well. We hope the reader will agree!

We have already used currents and given an informal introduction to the subject in §10.2. Here we shall need another aspect which is more formal and is related to their invariance properties as they are defined on a general manifold. For this the reader will have to fall back on de Rham (1960). We shall describe things and give precise references as we go along, and since there will be many of them will simply specify the section or page as 'de Rham (1960, §x)', say. But the task of seeing exactly how things are done by de Rham will have to be carried out by the reader. Note, however, that this is not the only section in which we shall give background material on the theory of currents. This task will be picked up again in §§12.8–12.10

12.3.2 Even and odd forms

Apart from the style of de Rham (1960) we must clarify another technical but important point. The manifolds M considered in that book are not assumed to be orientable. As a result de Rham (1960, §5) considers two kinds of (smooth) differential forms: the even ones and the odd ones, and similarly for even and odd currents (de Rham, 1960, §8). In those considerations below, all our manifolds will be orientable and we shall dispense with this distinction. All the forms that we shall consider are even forms in the sense of de Rham; similarly, all the currents are odd, being the elements of the dual of the space of even forms.

One standard way of saying that a manifold M is orientable is to demand that there exists some non-vanishing real smooth differential form of maximal degree. The above definition of orientability, in the terminology of de Rham (1960, §5), implies the existence of a C^∞ odd form ε of degree 0 which in terms of local coordinates is a scalar that is $\varepsilon = \pm 1$. De Rham calls this an orientation and it can be used to turn even forms into odd forms and vice versa by $\omega \rightleftarrows \varepsilon\omega$.

Exercise 12.11 Prove that Lie groups are orientable. Use the left-invariant (Haar measure) volume form.

De Rham's terminology for odd forms seems to have fallen out of use (see Bott and Tu, 1982, §7 for the modern terminology) but the great advantage built into the formalism is that the integral

$$\int_M \text{(odd form of maximal degree)} = \text{ can be defined (see de Rham, 1960, §5).} \tag{12.16}$$

For even forms ω this can only be done by using an orientation, if it exists, by passing to the corresponding odd form $\varepsilon\omega$ (see de Rham, 1960, §6). More explicitly, for an orientable manifold M we shall fix the orientation ε (there are two choices for this) and for an even form ω we shall write

$$\int \omega \quad \text{instead of} \quad \int \varepsilon\omega \quad \text{(terminology of de Rham, 1960, §6).} \tag{12.17}$$

There are therefore two possible choices for this integral on the left.

12.3.3 Further notation

Let M be some C^∞ orientable manifold. The space of all (even) smooth differential forms on M is denoted by $\mathscr{E} = \mathscr{E}(M)$ and the subspace of compactly supported forms is denoted by $\mathscr{D} = \mathscr{D}(M)$.

For their natural topologies (see de Rham, 1960, §9), $\mathscr{D}' = \mathscr{D}'(M)$ is the dual of $\mathscr{D}$ and is the space of currents (see de Rham, 1960, §8) on M, and $\mathscr{E}' = \mathscr{E}'(M)$ is the dual of $\mathscr{E}$ and is the space of the currents with compact support.

We shall use the notation (see de Rham, 1960, §8)

$$\begin{aligned} (\alpha, T) = (T, \alpha) = T[\alpha]; \qquad & \alpha \in \mathscr{D},\ T \in \mathscr{D}', \\ (\beta, S) = (S, \beta) = S[\beta]; \qquad & \beta \in \mathscr{E},\ S \in \mathscr{E}'. \end{aligned} \tag{12.18}$$

Here, again for the natural topologies, we are in a reflexive situation and the duals, written as $(\ )^*$ are (see de Rham, 1960, §10)

$$(\mathscr{D}')^* = \mathscr{D}, \qquad (\mathscr{E}')^* = \mathscr{E}. \tag{12.19}$$

However, in what follows, the dual of a topological vector space E will be denoted by E' or E^*, whichever of the two is more natural; and sometimes it will be more natural to use $\langle\cdot,\cdot\rangle$ rather than $(\cdot,\cdot)$.

12.3.4 An illustration: chains

Let M be some C^∞ manifold (orientable or not!) and let c be some chain in M. This we recall (de Rham, 1960, §6) is

$$\pi\colon \Pi \longrightarrow M; \quad \Pi \subset \mathbb{R}^p, \tag{12.20}$$

where Π is some non-singular simplex $\Pi \subset \mathbb{R}^p$ together with some orientation ε on $\mathbb{R}^p$ and π defined and C^∞ in some neighbourhood of Π in $\mathbb{R}^p$.

Now let $\alpha \in \mathscr{D}$ run through the space of (even) forms, the current associated to c; the integral of α on c is then defined in de Rham (1960, §6, §8, Example 1) as

$$(\alpha, c) = \int_c \alpha = \int \varepsilon f \pi^*(\alpha), \tag{12.21}$$

where f is the characteristic function of Π in $\mathbb{R}^p$ and π^* is the pullback mapping on the differential forms. Explicitly, to avoid de Rham's notation, we should consider only forms of degree p and write

$$(\alpha, c) = \int_\Pi \pi^*(\alpha) = \pm \int f(x) a(x)\, \mathrm{d}x_1 \cdots \mathrm{d}x_p; \tag{12.22}$$

$$\pi^*(\alpha) = a(x)\, \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_p; \quad (x_1, \ldots, x_p) \text{ Euclidean coordinates of } \mathbb{R}^p;$$

that is, $\pm$ the integral with respect to Lebesgue measure with the $\pm$ being decided by the orientation on $\mathbb{R}^p$. It is the orientation that by definition determines whether $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_p$ is $+$ or $-$ the Lebesgue measure. For forms α that are not of degree p, the scalar product in (12.22) is 0, by definition.

12.3.5 Currents as forms with distribution coefficients

This point is very important in the theory and is in complete analogy with distribution theory (Schwartz, 1957).

Distributions on $\mathbb{R}^n$ are defined as elements of the dual of $C_0^\infty(\mathbb{R}^n)$ (this space is denoted by $\mathscr{D}$ in Schwartz, 1957) but they can and they should also be considered as generalised functions and this is in conformity with the historical definition in the pioneering work of Sobolev et al. in partial differential equations.

This is done as follows. We start from some function f say in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and associate the linear functional

$$\varphi \longmapsto f[\varphi] = \int \varphi(x) f(x)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n; \quad \varphi \in \mathscr{D}(\mathbb{R}^n). \tag{12.23}$$

The same thing can be done on differential forms on an orientable manifold M (whose orientation has been fixed). To wit, let α be some (even) differential form on M with L^1_{loc} coefficients; we can then define (see de Rham, 1960, §8, Example 2)

$$\varphi \longmapsto \alpha[\varphi] = \int \alpha \wedge \varphi; \quad \varphi \in \mathscr{D}(M). \tag{12.24}$$

A special case of (12.24) is (12.23) where $M = \mathbb{R}^n$, $\alpha = f\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n$ and where the orientation of $\mathbb{R}^n$ is the one that makes $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n =$ Lebesgue measure.

This way of looking at currents goes a long way. In this spirit, many authors prefer to define currents as differential forms with distribution coefficients. The drawback with this approach is that we run into difficulties with the transformation rule for local coordinates. The correct way of presenting this interpretation of currents can be found in de Rham (1960, §8, p. 42).

Be that as it may, these identifications allow us to write

$$\mathscr{D} \subset \mathscr{E}', \qquad \mathscr{E} \subset \mathscr{D}', \tag{12.25}$$

and consider these as subspaces of the largest possible space $\mathscr{D}'$.

12.3.6 The differential and the boundary operators

In §12.3.4 and (12.24) we gave examples of currents associated to a chain c or a differential form α. These and Stokes' theorem motivate the following definitions that we have already used in §10.2 (cf. de Rham, 1960, §11).

For $T \in \mathscr{D}'(M)$ we define

$$\begin{aligned} (bT, \alpha) &= (T, \mathrm{d}\alpha); \quad \alpha \in \mathscr{D}(M), \\ dT &= wbT. \end{aligned} \tag{12.26}$$

Finally, a form ω (resp. a current T) is called closed if $\mathrm{d}\omega = 0$ (resp. $\mathrm{d}T = 0$).

Here w is the linear operator on $\mathscr{D}'$ (see de Rham, 1960, §11) which multiplies currents of degree p by $(-1)^p$. And we say that a current T is of *degree* p if $(T, \alpha) = 0$ for all $\alpha \in \mathscr{D}$ that are not of degree $n - p$, $n = \dim M$. We then also say that T is of *dimension* $n - p$ (see de Rham, 1960, §8).

The reader can easily verify that the above definitions are consistent with the degree of the differential forms in (12.24) and with the geometric dimension and boundary of c in (12.21). Seen like this, (12.26) is just a rewriting of Stokes' theorem (de Rham, 1960, §6).

12.3.7 The pullback of forms and the direct image of currents

Conforming with the notation of de Rham (1960, §§3, 11) let $\mu\colon W \to V$ be some C^∞ mapping between two orientable C^∞ manifolds. The pullback mapping $\mu^*\colon \mathscr{E}(V) \to \mathscr{E}(W)$ for differential forms is then defined. For example, if $\alpha \in \mathscr{E}(V)$ is of degree 0, that is, a function on W, then $\mu^*\alpha = \alpha \circ \mu$ is the function obtained by composition.

The direct image on currents is defined by duality (see de Rham, 1960, §p. 55) by $\mu T[\varphi] = T[\mu^*\varphi]$ in de Rham's notation ($T \in \mathscr{E}'$, $\varphi \in \mathscr{E}$). The direct image cannot in general be extended to $\mathscr{D}'$ unless μ is a proper mapping, that is, if μ^{-1} of a compact set is compact.

Note that when T is a Radon measure, that is, a current of dimension 0 (see de Rham, 1960, §8) then Bourbaki (1963) uses the notation $\check{\mu}$ for this direct image. For additional clarity we shall sometimes denote this mapping on currents by μ_* and call it the 'pushforward' (as opposed to the 'pullback' μ^* on forms).

12.4 Homology. Review of the Definitions of the General Theory

12.4.1 General definitions. Some classical examples

To define a homology theory like the one that we shall need to use we shall start from a sequence of real vector spaces (Λ_n; $n \in \mathbb{Z}$) and a sequence of linear mappings

$$\Lambda : \cdots \longrightarrow \Lambda_{n-1} \xrightarrow{d_{n-1}} \Lambda_n \xrightarrow{d_n} \Lambda_{n+1} \longrightarrow \cdots$$

that turn these into a complex in the sense that $d_{n+1} \circ d_n = 0$ (see Cartan and Eilenberg, 1956, §IV.3).

More often than not, natural locally convex linear topologies will be assigned to these spaces so that they become topological vector spaces (TVS).[1] In that case the d_n will be required to be continuous.

Example 12.12 All the spaces are finite-dimensional and up to equivalence only one separated topology can be assigned to these spaces (namely, the Euclidean topology). The standard notation and definitions are as follows:

$$H_n = H_n(\Lambda) = \frac{\operatorname{Ker} d_n}{\operatorname{Im} d_{n-1}} = \text{ the } n\text{th homology group} \tag{12.27}$$

(actually a vector space) and

$$\text{Ker} = \text{kernel of the mapping}, \qquad \text{Im} = \text{image of the mapping}. \tag{12.28}$$

[1] Throughout, such spaces will be assumed to be locally convex and their topologies will be separated (i.e. Hausdorff). For these spaces we shall use the references Bourbaki (1953), Grothendieck (1958) and Jarchow (1981).

Sometimes one also writes

$$\Lambda = \sum_{-\infty}^{+\infty} \Lambda_n; \quad d\colon \Lambda \longrightarrow \Lambda; \quad d|_{\Lambda_n} = d_n,\ d^2 = 0;$$
$$H = \sum_{-\infty}^{+\infty} H_n = \frac{\operatorname{Ker} d}{\operatorname{Im} d}, \tag{12.29}$$

and also, abusing notation[2] somewhat, we write instead $d\colon \Lambda \to \Lambda$.

One other item of notation that is often used and is useful is the image for the canonical projection in (12.29):

$$\operatorname{Ker} d \longrightarrow \frac{\operatorname{Ker} d}{\operatorname{Im} d} = H; \quad \xi \longrightarrow [\xi] \in H; \quad \xi \in \operatorname{Ker} d. \tag{12.30}$$

When $[\xi_1] = [\xi_2]$ we say that ξ_1 and ξ_2 are homologous. Also, when it is necessary to be more specific, the complex Λ is denoted (Λ, d).

One says that the complex Λ is geometric if $\Lambda_n = 0$ for $n < 0$ and for $n \geqslant n_0 \geqslant 0$. When all the homology spaces H_n are finite-dimensional we shall say that the complex Λ is *finite* or has *finite homology*.

Example 12.13 (i) The complex associated to a finite geometric simplicial complex is a geometric complex and all the spaces Λ_n are finite-dimensional.

(ii) Let us consider Λ the complex of singular simplexes in a nice topological space, for example, some compact manifold (see e.g. Dubrovin et al., 1990; Hilton and Wylie, 1960 and §12.4.3 below). Then Λ is finite but the spaces Λ_n are not finite-dimensional. Note that for the complexes that occur in topology we have that $\dim H_0$ equals the number of components of the space. In what follows it will also be convenient to adopt a terminology that is not standard. We shall say that a complex is *acyclic* if $H_n = 0$ for all $n \neq 0$.

(iii) Complexes of differential forms and currents on a manifold M will be considered in the next subsection. There the spaces Λ_n will be TVS in a natural way.

12.4.2 C^∞ manifolds and complexes of differential forms

Here and throughout this subsection M will denote some C^∞ differential manifold and the notation $\mathscr{D}$, $\mathscr{D}'$, etc. will be as in §12.3.3. We shall denote by $\mathscr{D}_n \subset \mathscr{D}$, $\mathscr{E}_n \subset \mathscr{E}$, $\mathscr{D}'_n \subset \mathscr{D}'$, $\mathscr{E}'_n \subset \mathscr{E}'$ the corresponding forms and currents of

[2] Despite its ambiguity, this is the notation used in the standard texts on algebraic topology. We shall be using the following references systematically: Godement (1958), Hilton and Wylie (1960), Dubrovin et al. (1990), Cartan (1948), Massey (1991).

degree n with $0 \leqslant n \leqslant \dim M$. Currents of dimension p will be denoted by $\mathscr{D}_{(p)}$, $\mathscr{E}_{(p)}$.

The exterior differential d and the boundary operator b in §12.3.3 turn $\mathscr{D}, \mathscr{E}$, $\mathscr{D}', \mathscr{E}'$ into complexes. For example, we have

$$\mathscr{D}\colon \cdots \longrightarrow \mathscr{D}_{n-1} \xrightarrow{\mathrm{d}} \mathscr{D}_n \xrightarrow{\mathrm{d}} \mathscr{D}_{n+1} \longrightarrow \cdots,$$
$$\mathscr{E}'\colon \cdots \longrightarrow \mathscr{E}'_{n-1} \xrightarrow{\mathrm{d}} \mathscr{E}'_n \xrightarrow{\mathrm{d}} \mathscr{E}'_{n+1} \longrightarrow \cdots.$$

The above complex $(\mathscr{E}', \mathrm{d})$ can be written alternatively and equivalently in terms of the boundary operator

$$\mathscr{E}'\colon \cdots \longleftarrow \mathscr{E}'_{(p-1)} \xleftarrow{b} \mathscr{E}'_{(p)} \xleftarrow{b} \mathscr{E}'_{(p+1)} \longleftarrow$$

where $\mathscr{E}'_{(m)} = \mathscr{E}'_{N-m} =$ compactly supported currents of dimension m with $N = \dim M$. Here of course the reversal of the arrow for the b operator in $\mathscr{E}'$ appears simply as a superficial change of notation. But this reversal of the arrow also reflects another point that has played an historically important role in the development of algebraic topology.

This is the distinction between cohomology theory as given by the complex $(\mathscr{E}, d)$ and homology theory as given by the complex $(\mathscr{E}', b)$. The difference between homology and cohomology is that when $X \xrightarrow{f} Y$ is some C^∞ mapping between two C^∞ manifolds then it induces

$$H\left(\mathscr{E}'(X)\right) \longrightarrow H\left(\mathscr{E}'(Y)\right) \tag{12.31}$$

by the direct mapping on currents (see de Rham, 1960, §11) but a mapping in the opposite direction,

$$H\left(\mathscr{E}(X)\right) \longleftarrow H\left(\mathscr{E}(Y)\right), \tag{12.32}$$

by the pullback mapping of differential forms (see de Rham, 1960, §§5, 18). In the terminology of algebraic topology this says that homology is a covariant functor and that cohomology is a contravariant functor.

This distinction between homology and its dual, cohomology, appears only in a marginal way in de Rham (1960, §21), where cochains are considered for the first and last time. The point of view adopted by de Rham, and this is also the point of view that is best suited for our purposes, is the following.

The homology is defined on the largest (universal in some sense) complex $(\mathscr{D}', d)$ and insofar that the other complexes $\mathscr{D}$, $\mathscr{E}$, $\mathscr{E}'$ can be considered as subcomplexes of $\mathscr{D}'$ the definition of their homologies can be 'subordinated' to this. This of course does not necessarily mean a priori that, say, $H(\mathscr{E}')$ can be identified to a subspace of $H(\mathscr{D}')$. Indeed, two closed currents $T_1, T_2 \in \mathscr{E}'$

could be homologous in $\mathscr{D}'$ without being homologous in $\mathscr{E}'$; see de Rham (1960, §18) for more on this important point.

The fact, on the other hand, that we can identify $H(\mathscr{E}) = H(\mathscr{D}')$ is an important consequence of the regularisation of de Rham (1960, §15) to which we shall return in §12.10.3 below.

From our point of view the relation between the homology of Λ in §12.4.1 and the homology of its dual Λ^* (see §12.4.4 below) is something that will play a pivotal role in this chapter and the next.

12.4.3 More on singular homology. Connections with algebraic topology

In de Rham (1960, §21) one finds a proof of the equivalence between the homologies of $\mathscr{D}$ and $\mathscr{E}$ and the simplicial homology of the manifold M. Even more direct is the connection with the singular homology $\mathscr{S}\colon \cdots \to \mathscr{S}_{n-1} \xrightarrow{\partial} \mathscr{S}_n \xrightarrow{\partial} \mathscr{S}_{n+1} \to \cdots$ that we obtain on M by the *Lipschitz singular simplexes*.

To be more precise, we consider $\mathscr{S}_n$ the space spanned by all Lipschitz mappings

$$f\colon \Pi \longrightarrow M \quad \text{where } \Pi = (0, e_1, \ldots, e_p) \subset \mathbb{R}^p \tag{12.33}$$

is the standard simplex and $e_j = (0, \ldots, 1, 0, \ldots, 0)$ are the unit coordinate vectors. An orientation is then induced by the natural order of these vectors and the chain c and the current $\int_c \alpha = \int_\Pi f^*\alpha$ can be defined as in §12.3.4. These singular simplexes can thus be identified to currents in $\mathscr{E}'(M)$ by the definitions of §12.3.4 and de Rham (1960, §8). This identification intertwines the boundary operators b of (12.26) and ∂ in here. We obtain thus a mapping $H(\mathscr{S}) \to H(\mathscr{E}')$. That this mapping induces an isomorphism of the two homologies is a long story. For instance, one can use de Rham (1960, §21) and also the not-so-obvious but nonetheless well-known fact that singular homology and simplicial homology coincide on M. This can be found in standard algebraic topology references such as Massey (1991) and Hilton and Wylie (1960). (For the above 'Lipschitz' variant one can use, e.g., a modification of Bott and Tu, 1982, §15.8. This proof is, however, rather sophisticated because it goes through Čech cohomology. Or see Massey (1991), Appendix A, for a direct proof.)

12.4.4 Intermediate spaces and complexes

As before, M will be some C^∞ manifold and $\mathscr{D} = \mathscr{D}(M) \subset \mathscr{D}' = \mathscr{D}'(M)$ are in some sense the largest and the smallest spaces of forms that make sense to consider. All other natural spaces of forms are intermediate and together with the definition of such a space Λ we usually assign a topology to Λ which, if the definition of the subspace is natural, should be built into the definition. A number of examples that we shall need later will be given in the next subsection, but of course, every aspect of modern analysis is full of such spaces; see, for example, Schwartz (1957).

The dual space Λ' of Λ can then be defined and since we shall also need this to be an intermediate space

$$\mathscr{D} \subset \Lambda \subset \mathscr{D}', \qquad \mathscr{D} \subset \Lambda' \subset \mathscr{D}', \tag{12.34}$$

we shall often demand in the original definition of the TVS Λ that $\mathscr{D}$ is dense in Λ because this guarantees the second inclusion in (12.34).

The problem here is that unless something very specific happens, with the exception of the obvious examples $\mathscr{D}, \mathscr{E}, \mathscr{D}', \mathscr{E}'$, there is no reason at all that such a space should give rise to a complex, that is, that $\mathrm{d}\Lambda \subset \Lambda$ for the differential of $(\mathscr{D}', \mathrm{d})$.

There are two natural ways to mend this. If we start from an arbitrary intermediate space Λ we can then define two complexes as follows:

$$\hat{\Lambda} = \Lambda \cap \mathrm{d}^{-1}\Lambda, \quad \check{\Lambda} = \Lambda + \mathrm{d}\Lambda. \tag{12.35}$$

These are subspaces of $\mathscr{D}'$ and that they are indeed complexes is obvious (for the first one uses $\mathrm{d}^2 = 0$ and $\xi \in \hat{\Lambda}$ certainly implies that $\zeta = \mathrm{d}\xi \in \mathrm{d}^{-1}\Lambda$ since $\mathrm{d}\zeta = 0$). The issue here is how we define the topologies in (12.35). This is done by standard ideas in TVS (see e.g. Bourbaki, 1953; Grothendieck, 1958).

We can identify $\hat{\Lambda}$ with a subspace of $\Lambda \times \Lambda$ by the inclusion map $\xi \to (\xi, \mathrm{d}\xi)$, for $\xi \in \hat{\Lambda}$, and the topology we give on $\hat{\Lambda}$ is the topology as a subspace of the Cartesian product. (This is the weakest topology that makes both the $i \colon \hat{\Lambda} \to \Lambda$, $i(\xi) = \xi$, and $\mathrm{d} \colon \hat{\Lambda} \to \Lambda$, $\xi \to \mathrm{d}\xi$ continuous.)

By Hahn–Banach we see in particular that the dual is

$$\left(\hat{\Lambda}\right)^* = \Lambda' + b\Lambda' = \Lambda' + \mathrm{d}^*\Lambda' \tag{12.36}$$

for the dual mapping $b = \mathrm{d}^*$ of (12.26). (Exercise: verify this.)

Similarly, the natural topology for $\check{\Lambda}$ is the strongest topology that renders continuous the two mappings $\Lambda \xrightarrow{i} \check{\Lambda}$ (i is the identical inclusion) and $\Lambda \xrightarrow{\mathrm{d}} \check{\Lambda}$.

Exercise 12.14 For this topology we can identify the dual $(\check{\Lambda})' = \Lambda' \cap b^{-1}\Lambda'$. No direct use of this will be made but doing this exercise will allow the reader to brush up on topological vector spaces! (see Bourbaki, 1953).

Observe incidentally that the density of $\mathscr{D}$ in $\check{\Lambda}$ for this topology is automatic because $\mathrm{d}\mathscr{D} \subset \mathscr{D}$. The proof that $\mathscr{D}$ is dense in $\hat{\Lambda}$ usually has to be done in an ad hoc manner (see Exercises 12.25 and 12.26 below).

Remark 12.15 Related to the above, one should note a subtle but largely irrelevant point. The identification of the dual to (12.36) is just abstract Hahn–Banach and makes sense whether $\mathscr{D}$ is dense in $\hat{\Lambda}$ or not. But unless we have that density, the sum in (12.36) is not the sum in $\mathscr{D}'$ and the dual of $\hat{\Lambda}$ cannot be identified to a subspace of $\mathscr{D}'$. In concrete terms, $(T_1, bT_2) \in \Lambda' \oplus b\Lambda'$ always gives rise to an element of $(\hat{\Lambda})^*$.

12.4.5 Examples and further remarks

(i) We shall consider $\mathscr{C} = \mathscr{C}(M)$ the space of differential forms with continuous coefficients. The topology is given by the seminorms

$$p_K(\omega) = \sup_K |\text{coefficients of } \omega|; \quad K \subset\subset M \text{ compact subsets.} \tag{12.37}$$

Changing the local charts and the coefficients clearly causes no problem. (Exercise: prove this.) This is a Fréchet space, or at least this will be the case for all the manifolds that we shall be considering. This means that it is a metrisable complete, locally convex TVS. A countable number of seminorms are needed to define the topology (see Bourbaki, 1953; Grothendieck, 1958 and §12.14.1 below).

The dual $\mathscr{C}'$ of $\mathscr{C}$ is the space of compactly supported forms with coefficients that are Radon measures (see Bourbaki, 1953; Bourbaki, 1963).

(ii) If we take $\mathscr{K}(M) \subset \mathscr{C}(M)$ the subspace of compactly supported continuous forms with the appropriate topology, the dual ($(\mathscr{K}(M))'$; see Bourbaki, 1963) is the space of all integration currents on M, that is, all forms with coefficients that are Radon measures.

(iii) The complex associated to $\mathscr{C}$ as in (12.35) is $\mathscr{C} \cap \mathrm{d}^{-1}\mathscr{C}$ and the natural Fréchet topology is given by the seminorms

$$q_K(\omega) = \sup_K [|\text{coefficients of } \omega| + |\text{coefficients of } \mathrm{d}\omega|]; \\ K \subset\subset M \text{ compact subsets,} \tag{12.38}$$

and this gives the topology defined in (12.35). For these topologies $\mathscr{D}$ is dense in both $\mathscr{C}$ and $\mathscr{C} \cap \mathrm{d}^{-1}\mathscr{C}$. (Exercise: prove this; see Exercises 12.25 and 12.26 for a more difficult case.) Similarly, a differential complex structure and a topology can be assigned on $\mathscr{K} \cap \mathrm{d}^{-1}\mathscr{K}$ and the dual of this is the space of currents T that admit the decomposition

$$T = S_1 + bS_2; \quad S_1, S_2 \text{ integration currents on } M. \tag{12.39}$$

These types of currents are important in a number of issues and we shall come back to these in §12.7 below. Equation (12.39) is very close to the more precise notion of *flat* currents developed in $\mathbb{R}^n$ in Federer (1969). Similarly, $\mathscr{K}' \cap \mathrm{d}^{-1}\mathscr{K}'$ is the complex of *normal* currents (see §10.2.6).

(iv) As a final point observe that we could have replaced the continuous forms in $\mathscr{C}$ and $\mathscr{C} \cap \mathrm{d}^{-1}\mathscr{C}$ by forms that have coefficients in L^∞_{loc}. Otherwise the topologies and the seminorms are defined as before in (12.37) – the sup being of course replaced by 'ess sup'. This generalisation has the following advantage over $\mathscr{C}$. Let $f\colon M_1 \to M_2$ be some mapping. Then if f is smooth the pullback works in $\mathscr{C}$, that is, $f^*\omega \in \mathscr{C}(M_1)$ for all $\omega \in \mathscr{C}(M_2)$, but this property breaks down if f is only locally Lipschitz. To rescue the property, L^∞_{loc} coefficients have to be considered.

Another feature of L^∞_{loc} is that, although its dual is a very big space, L^∞_{loc} itself is the dual of $L^1_{\mathrm{comp.}}$, which is the compactly supported L^1 functions. This we shall see will be used in §12.13 below.

Remark 12.16 (A final remark) The intermediate spaces Λ in (12.34) can very rarely be made reflexive $(\Lambda')' = \Lambda$. As we shall see, when this reflexivity holds it is a great help, but since it is not there in general we have to do without. The elaborate technical contortions that we shall indulge in §12.12 below are designed to give the correct substitute. This also gives an unexpected twist to the proofs.

12.5 The Heart of the Matter. Forms of Polynomial Growth

12.5.1 Riemannian norm on differential forms

Throughout this section a Riemannian structure will be assigned on the C^∞ manifold. For every local chart $\Omega \subset M$ we can then apply the Gram–Schmidt process on $dx_1, \dots, dx_n$ for the local coordinates and obtain $\omega_1, \dots, \omega_n$ smooth differential 1-forms that are orthonormal for the Riemannian structure on M, that is, the dual of the Riemannian structure of TM. Now any differential form $\omega \in \mathscr{C}(M)$ (or even the L^∞_{loc} variant of §12.4.5) can be written uniquely (this is only almost everywhere in the L^∞_{loc} variant):

$$\varphi = \sum_{p \geqslant 0} \sum_{|I|=p} a_I(m)\omega_I; \quad I = (i_1 < \cdots < i_p) \text{ is a } p\text{-multi-index},$$

$$\omega_I = \omega_{i_1} \wedge \cdots \wedge \omega_{i_p},\ a_I(m) \in \mathbb{R},\ m \in \Omega. \qquad (12.40)$$

One then defines

$$|\varphi(m)| = \sum |a_I(m)|; \quad m \in \Omega, \qquad (12.41)$$

and this up to equivalence is independent of the coordinates $x_1, \ldots, x_n$ and the choice of $\omega_1, \ldots, \omega_n$. The square root of the sum of the squares would have been a better choice in (12.41) and then the strict invariance is clear; this point, however, is not important. For a more formal presentation of these facts see Warner (1971, Exercise 2.13, p. 79). Also up to equivalence, $|\varphi|$ in (12.41) is invariant when we change the Riemannian metric on M to a quasi-isometric equivalent Riemannian metric. In (12.41) only forms of degree $\geqslant 1$ have been considered. If the degree is 0 then φ is a scalar function and in (12.41) we set the ordinary absolute value.

Lie groups Our main object of study will be connected Lie groups G assigned with their left-invariant Riemannian metric. In that case the orthonormal basis $\omega_1, \ldots, \omega_n$ of T^*M can be chosen globally and is obtained by the left translation of a set of orthonormal basis vectors $\omega_1^{(0)}, \ldots, \omega_n^{(0)}$ of the cotangent space T_e^*G at the identity point $e \in G$.

12.5.2 Spaces of differential forms on M

The basic space will be the space of bounded $\mathscr{C}(M)$ forms, that is, the forms φ of §12.4.5 for which

$$\|\varphi\|_0 = \sup_m |\varphi(m)| < +\infty. \tag{12.42}$$

For the L^∞_{loc} coefficients the supremum is of course replaced by the essential supremum (ess sup). Although $\| \ \|_\infty$ would conform better with the traditional terminology, the notation (12.42) is preferable as we can see in the notation below.

In what follows, $\mathrm{O} \in M$ will be a fixed base point and as before we shall denote $|m| = d(\mathrm{O}, m)$ with $m \in M$. The following notation and definitions will then be used throughout:

$$\|\varphi\|_p = \|\varphi(1+|m|)^{-p}\|_0; \quad p \geqslant 0,\ \varphi \in \mathscr{C}(M) \text{ (or in the } L^\infty_{\text{loc}} \text{ variant)},$$
$$\mathscr{C}_p = \mathscr{C}_p(M) = [\varphi \in \mathscr{C}(M);\ \|\varphi\|_p < +\infty],$$
$$\mathscr{C}(\text{pol}) = \mathscr{C}(M;\text{pol}) = \bigcup_{p \geqslant 0} \mathscr{C}_p \quad (\cdots \mathscr{C}_p \subset \mathscr{C}_{p+1} \subset \cdots).$$

Sometimes the notation $\mathscr{C}(M;\text{pol}, L^\infty_{\text{loc}})$ will also be used for the L^∞_{loc} variant.

12.5.3 The complex of polynomial forms

This will be the complex that we can construct from $\mathscr{C}(M;\mathrm{pol})$ by the construction of $\hat{\Lambda}$ in (12.35). We shall denote

$$\Lambda_P(M) = \Lambda(M;\mathrm{pol}) = \big[\varphi \in \mathscr{C}(M,\mathrm{pol});\ \mathrm{d}\omega \in \mathscr{C}(M,\mathrm{pol})\big]. \qquad (12.43)$$

The L^∞_{loc} variant can be defined analogously and we write $\Lambda(M;\mathrm{pol},L^\infty_{\mathrm{loc}})$.

The other way $\check{\Lambda}$ of turning $\mathscr{C}(M;\mathrm{pol})$ into a complex could have been used but we shall not use this here. Observe, however, that the theory that we shall build around $\Lambda_P(M)$ in the rest of the chapter, we can also build for this $\check{\Lambda}$ larger variant with essentially only formal changes in the arguments.

12.6 Statement of the Homology Theorems

12.6.1 Simply connected groups

The first result in that direction concerns the simply connected groups that are soluble and which as before constitute the building block of the theory.

Theorem 12.17 (Simply connected soluble groups) *Let G be some simply connected soluble group. Then the following conditions on G are equivalent*:

(i) *G is an NC-group*;
(ii) $\Lambda_P(G)$ *is acyclic, that is,* $H_n = 0$ *for* $n \neq 0$ (*see* §*12.4.1*);
(iii) *the homology of* $\Lambda_P(G)$ *is finite-dimensional*.

Recall that the complexes $\Lambda = \Lambda_P, \mathscr{E}, \ldots$ for any *connected* manifold U or topological space for that matter, we have $H_0(\Lambda) = \mathbb{R}$. The proof of the theorem will not be given before the next chapter. But it requires quite a bit of preparation and this will be carried out in the remainder of this chapter. Note, however, that the proof of the implication (i)⇒(ii) is contained in Chapter 8 because of §12.9 below.

It will be important to reformulate this basic result in terms of the classical Poincaré equation $\mathrm{d}\theta = \omega$ without the use of currents or even the formalism of homology theory. This is done in the following proposition, the proof of which will be given in §12.10.3 below and has nothing to do with the NC-condition. The proposition also holds for more general manifolds but the proof in this more general case requires more machinery and will not be given.

Proposition 12.18 *Let G be some connected Lie group. Then the following two properties are equivalent*:

(i) *The complex* $\Lambda_P(G)$ *is acyclic*.

(ii) *Let* $\varphi = \varphi_1 + \varphi_2 + \cdots \in \Lambda_P(G) \cap \mathscr{E}(G)$ *be some smooth differential form that grows polynomially, has no constant term and is closed, that is,* $\varphi_n \in \mathscr{E}_n$ *as in §12.4.2, and* $\mathrm{d}\varphi = 0$*. Then there exists* $\theta \in \Lambda_P(G) \cap \mathscr{E}(G)$ *such that* $\mathrm{d}\theta = \varphi$.

Similarly, the following two properties are equivalent.

(iii) *The homology of* $\Lambda_P(G)$ *is finite-dimensional.*
(iv) *There exist* $\omega^{(1)}, \ldots, \omega^{(p)} \in \Lambda_P(G) \cap \mathscr{E}(G)$*, finitely many closed smooth differential forms of polynomial growth such that for all* $\varphi \in \Lambda_P(G) \cap \mathscr{E}(G)$ *that is closed and smooth, there exist* $\theta \in \Lambda_P(G) \cap \mathscr{E}(G)$ *and scalars* $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ *such that*

$$\varphi = \mathrm{d}\theta + \lambda_1 \omega^{(1)} + \cdots + \lambda_p \omega^{(p)}. \tag{12.44}$$

One then usually says that the $\omega^{(1)}, \ldots$ form a basis of the homology.

The situation is illustrated well by compact groups that are NB-groups. Just about everything is known on the homology of compact Lie groups (e.g. Greub et al., 1973 Chapter IV; Dubrovin et al., 1990). But one thing that is obvious and certain is that this homology is finite-dimensional, as is the homology of any compact manifold, because among other things such a manifold admits a finite triangulation; see, for example, Hilton and Wylie (1960) and Bott and Tu (1982, §5.3.1). More to the point, we have the following theorem.

Theorem 12.19 (Simply connected groups) *Let G be some simply connected Lie group. Then the following two conditions on G are equivalent*:

(i) *G is an* NB-*group*;
(ii) *the homology of* $\Lambda_P(G)$ *is finite-dimensional.*

This result is in fact contained in the previous theorem. This can be seen in the following exercise.

Exercise 12.20 When G is simply connected, by §11.1.3 it is smoothly quasi-isometric to $Q \times K$, where Q is simply connected soluble and K is compact. The two canonical mappings

$$Q \underset{i}{\longrightarrow} G = Q \times K \underset{p}{\longrightarrow} Q \tag{12.45}$$

show that if $\omega^{(1)}, \ldots, \omega^{(p)} \in \Lambda_P(G)$, with $\mathrm{d}\omega^{(n)} = 0$, induce a basis of the homology of $\Lambda_P(G)$ then $\left[i^*\left(\omega^{(1)}\right)\right], \left[i^*\left(\omega^{(2)}\right)\right], \ldots$ (see (12.32) and §12.8 below) span the homology of $\Lambda_P(Q)$. This means that if $\Lambda_P(G)$ is finite then so is $\Lambda_P(Q)$. Theorem 12.17 therefore applies and shows that Q is NC. But this together with §11.1.4 shows that G is NB and proves the implication (ii)⇒(i) in Theorem 12.19.

The implication the other way round, (i)⇒(ii), in Theorem 12.19 follows from the polynomial retract of Q to a point and therefore of G to the compact submanifold K (see §12.2). By general considerations that we shall explain in §12.9 below, this retract gives an isomorphism between the homology of $\Lambda_P(G)$ and $H(K)$, the homology of K. This homology is therefore finite-dimensional.

Remark (i) One clearly sees from the above the necessity of using condition (iii) in Theorem 12.17.

(ii) The correct way to exploit the product in (12.45) is through the generalisation of Künneth's formula in the appendix to Chapter 14. This formula is rather sophisticated but once there, the equivalence of (i) and (ii) in the above theorem is automatic.

12.6.2 General connected Lie groups

What we did in the previous subsection was to use the structure theorems of Chapter 11 to reduce the proof of Theorem 12.19 to the proof of Theorem 12.17, that is, the case of a simply connected Lie group to the special case of a soluble simply connected Lie group. This reduction, at least in one direction, can also be done in a more general situation. To be specific, the following result holds.

Theorem 12.21 (General Lie groups) *Let G be some connected Lie group. Then the following two conditions on G are equivalent*:

(i) *G is an* NB-*group*;
(ii) *the homology of $\Lambda_P(G)$ is finite-dimensional*.

Because of §12.9 the implication (i) $\Longrightarrow$ (ii) is a consequence of Appendix F. In the next few lines, however, we shall show how this can be achieved without using the full thrust of that difficult appendix.

Our claim is that for the implication (i)⇒(ii) we can reduce the proof to the special case and assume that G is soluble (but not necessarily simply connected).

Exercise 12.22 To prove this, as in Exercise 12.10 we start from a general connected Lie group G and construct a smooth quasi-isometry $\simeq$ where

$$\widetilde{G} \simeq Q \times K, \qquad \widetilde{G}/F = G, \qquad \widetilde{G} \underset{\pi}{\longrightarrow} G, \tag{12.46}$$

Q is soluble, K is compact, both are connected and F is a finite central subgroup of the new Lie group $\widetilde{G}$. Furthermore, if G is assumed to be NB the group Q is NC. The above construction was done in §11.1.5.

Now let $\omega_1, \dots, \omega_n \in \Lambda_P(G) = \Lambda_P(\widetilde{G}/F)$ closed differential forms and let us pull them back to $\omega_j^* = \pi^*\omega_j \in \Lambda_P(Q \times K)$. If $[\omega_1^*] \cdots [\omega_n^*]$, the homology classes of these forms, are linearly dependent, then we can find scalars $\lambda_1, \dots, \lambda_n$, not all 0, and some form $\theta^* \in \Lambda_P(Q \times K)$ such that

$$\mathrm{d}\theta^* = \lambda_1 \omega_1^* + \cdots + \lambda_n \omega_n^*. \tag{12.47}$$

By taking averages with respect to the group action by F we can assume that $\theta^* = \pi^*\theta$ for some $\theta \in \Lambda_P(G)$ and therefore the original homology classes $[\omega_1], \dots, [\omega_n]$ are linearly dependent also.

What we have thus shown is that if the homology of $\Lambda_P(Q \times K)$ is finite, so is the homology of G. Incidentally, one can also easily prove this the other way round (see §14.3.2 below) but this fact is not needed here.

It remains to be seen that if Q is a soluble connected NC-group and if K is a compact connected group then the homology of $\Lambda_P(Q \times K)$ is finite.

Now, when Q is a connected soluble NC-group then Q is polynomially homotopically equivalent to a torus $\mathbb{T}^k$, $Q \simeq \mathbb{T}^k$ (see Appendix F). It follows therefore that $Q \times K \simeq$ some compact Lie group. Our final assertion therefore follows from the general results of §12.9 below.

We see therefore that modulo some relatively easy additional facts (much easier than the full thrust of Appendix F) we have a proof of the implication (i)⇒(ii) in the theorem.

Exercise Verify that Appendix F becomes much easier when we assume that the group is soluble. The additional algebraic results from Appendix A and Appendix B are then not needed. And although one uses the same geometric strategy explained in §§F.4, F.5, 8.5.2, the only additional fact that is needed is the easy algebraic result in (3.50) and the lemma in §C.1 (and even that lemma is not really essential; see Varopoulos, 2000b, §5). The bottom line is that the algebra in Appendix A and Appendix B is only essential for the homotopy of Theorem 12.9, and not for Theorem 12.21.

12.6.3 The scope and return on investment of Part III

Theorem 12.21 gives a characterisation of the B–NB condition in terms of homology theory. The homology used has to be defined, especially when seen in the light of the Poincaré equation. The bottom line is that this theorem is very much in the *nature of things*. Its main drawback is that it takes about 150 pages to give a complete proof of the implication (ii) $\Longrightarrow$ (i) of this theorem: in other words, essentially the whole of Part III. (*Added in proof*: In the Epilogue we describe a different approach to this implication. When and if this is written out, it will for sure give shorter proofs all round.)

The author hesitated in writing the details of that proof, because the effort seemed disproportionate, before finally deciding to do so. The following reasons were decisive in making that decision, and the reader may ponder these to decide whether to read Part III or not.

First, without the implication (ii) $\Longrightarrow$ (i), we do not have a general proof of Theorem 12.9 either, and this weakens the geometric theory considerably.

But not only that: a variety of new tools, or rather variations of classical tools in geometry and algebraic topology, had to be developed. These are interesting in their own right and also point to further developments in the subject. It is also these that take most of the space!

12.7 Banach Spaces of Currents and Their Duals

Notation The letter M that was used up to now to indicate a smooth connected manifold will now be changed to $U, V, \dots$. This conforms with de Rham's notation and, more to the point, will avoid confusion in the formulas that follow. As before, the manifolds that will be considered are assigned with a base point and a Riemannian metric. Here are samples of the notation that we shall be using:

$$m, x, \dots \in U; \quad \varphi \in \mathscr{E}(U),\ T \in \mathscr{D}'(U).$$

12.7.1 The total mass norm

We shall adopt the notation

$$M(T) = \sup\big\{|(T,\varphi)|;\ \varphi \in \mathscr{D}(U),\ \|\varphi\|_0 \leqslant 1\big\}; \qquad \| \ \|_0 \text{ as in (12.42)},\ T \in \mathscr{D}'(U), \tag{12.48}$$

and we shall also write, more generally,

$$M_p(T) = M(T(1+|m|)^p); \quad T \in \mathscr{D}'(U),\ p \geqslant 0. \tag{12.49}$$

The space of currents $T \in \mathscr{D}'(U)$, for which $M(T) < +\infty$, is the space of integration currents of finite total mass (see Federer, 1969). In fact, (12.49) is a slight abuse of notation because unless we already know that T is an integration current, then $T(1+|m|^p)$, where $|m|$ is the distance from the base point, may not make sense! However, we feel such pedantic points are best ignored.

Example 12.23 Radon measures on U can be identified to currents of dimension 0. For such a measure μ, $M(\mu) = \|\mu\|$ is its total mass norm (see Bourbaki, 1963). For $a \in U$ the standard notation δ_a will be used to indicate the Dirac δ-mass at a. We shall write (see §12.5.2)

$$\mathscr{C}_p^0(U) = \left\{ \varphi \in \mathscr{C}(U);\ |\varphi(m)|(1+|m|)^{-p} \underset{m\to\infty}{\longrightarrow} 0 \right\}. \tag{12.50}$$

Exercise 12.24 Show that $\mathscr{C}_p^0(U) \subset \mathscr{C}_p(U)$ is the closure of $\mathscr{D}(U)$ in $\mathscr{C}_p$ for the norm $\|\ \|_p$ of (12.5.2). See §12.7.2 below where a more intricate situation is considered.

As a consequence the dual can be identified to a space of currents

$$\left(\mathscr{C}_p^0(U)\right)^* = \left[T \in \mathscr{D}'(U);\ M_p(T) < +\infty\right]. \tag{12.51}$$

Exercise 12.25 Show that (12.51) follows from the definitions.

We shall also write

$$\mathscr{C}^*(U, \mathrm{pol}) = \bigcap_{p \geqslant 0} \left(\mathscr{C}_p^0(U)\right)^*, \tag{12.52}$$

which is the dual of $\mathscr{C}(U, \mathrm{pol})$ for the appropriate topology on this space. More of this topology and duality will be discussed in §12.13 and (12.63) below (see also the appendix to this chapter). However, no essential use of (12.52) or the topologies in question will be made. This space of currents consists of all the superpolynomially decaying integration currents. More precisely, T belongs to (12.52) if and only if

$$M_p(T) < +\infty; \quad p \geqslant 0 \tag{12.53}$$

(see also (12.67), (12.68) below). Inequality (12.53) can be expressed equivalently as

$$M\big[T\mathbb{I}[m \in U;\ |m| > R]\big] = \mathrm{O}(R^{-p}) \text{ as } R \longrightarrow \infty, \quad \text{for all } p > 0, \tag{12.54}$$

where $\mathbb{I}$ denotes the indicator function for the set outside the R-ball.

12.7.2 Banach spaces of complexes

We shall write

$$\Lambda_{p,q}(U) = [\varphi \in \mathscr{C}_p(U);\ d\varphi \in \mathscr{C}_q(U)]. \tag{12.55}$$

This is the space of currents for which the norm (cf. (12.35))

$$\|\varphi\|_{p,q} = \|\varphi\|_p + \|\mathrm{d}\varphi\|_q < +\infty. \tag{12.56}$$

The closure of $\mathscr{D}(U)$ in $\Lambda_{p,q}$ for the norm (12.55) is

$$\Lambda^0_{p,q}(U) = \left[\varphi \in \mathscr{C}^0_p(U);\ d\varphi \in \mathscr{C}^0_q(U)\right] \tag{12.57}$$

provided that $p \leqslant q$.

Exercise 12.26 This is a little more subtle. To see this we can consider say, locally Lipschitz, cut-off functions χ that have compact support and

$$\|\chi\|_0 + \|\mathrm{d}\chi\|_0 \leqslant 10^{10} \tag{12.58}$$

and are identically $\equiv 1$ in larger and larger compact subsets that saturate U. Such functions can be constructed by composing with the distance function $|m|$: $\chi(m) = \chi_0(|m|)$ for some χ_0 function on $\mathbb{R}$.

Having such a function we can replace any $\varphi \in \Lambda^0_{p,q}$ by $\chi\varphi \in \Lambda^0_{p,q}$ which is compactly supported. But to prove that this approximates φ we still have to consider

$$\mathrm{d}(\chi\varphi) = \mathrm{d}\chi \wedge \varphi + \chi \mathrm{d}\varphi \tag{12.59}$$

and to make this work we need the continuous inclusion $\mathscr{C}_p \underset{\rightarrow}{\subset} \mathscr{C}_q$. It is this that forces us to restrict ourselves to the case $p \leqslant q$.

Exercise 12.27 (The smoothing out) We still have to smooth out the form $\chi\varphi \in \Lambda^0_{p,q}$ to obtain an element in $\mathscr{D}(U)$. Since the support of $\chi\varphi$ is compact this is easy; for example, a partition of unity can be used to write $\chi\varphi = \sum \varphi_j$ with the φ_j having small support and then convolution for each φ_j in $\mathbb{R}^n$ to smooth out each φ_j.

The above regularisation can also be done by the method that is developed in §12.10 below in the case when U is a Lie group. This more precise density is, however, not needed for the identification below in (12.60) of the dual space to a space of currents (see Remark 12.15). Explicitly, the following can be asserted. From this density of $\mathscr{D}$, when $p \leqslant q$, we see, as in (12.39), that we can identify the dual space

$$\left(\Lambda^0_{p,q}(U)\right)^* = \left[T = T_1 + bT_2;\ T_1, T_2 \in \mathscr{D}',\ M_p(T_1) + M_q(T_2) < +\infty\right]. \tag{12.60}$$

Using the spaces of currents defined in (12.57) we have (see (12.43))

$$\Lambda_P = \bigcup_{p,q} \Lambda_{p,q} = \bigcup \Lambda^0_{p,q} = \mathscr{C}(U;\ \text{pol}) \cap \mathrm{d}^{-1}\mathscr{C}(U;\ \text{pol}) \tag{12.61}$$

because we always have

$$\Lambda_{p,q} \subset \Lambda^0_{p_1,q_1}; \quad p < p_1,\ q < q_1, \tag{12.62}$$

and if, as we shall briefly explain in §12.13.2 and the appendix to this chapter, the appropriate (inductive limit) topology (see e.g. Bourbaki, 1953) is assigned on this space, its dual can be identified with

$$\Lambda_P^* = \bigcap_{p\leqslant q} \left(\Lambda_{p,q}^0\right)^*. \tag{12.63}$$

This space of integration currents is in many ways as natural to consider in its own right as the original space Λ_P. It is, however, less simple to describe explicitly. This is done as follows. First, $T \in \Lambda_P^*$ if and only if for all $p,q > 0$ with $q \geqslant p$, we can rewrite $T \in \mathscr{D}'$ (see Remark 12.15),

$$T = T_1 + bT_2; \quad T_1 \in \left(\mathscr{C}_p^0\right)^*, \ T_2 \in \left(\mathscr{C}_q^0\right)^*, \tag{12.64}$$

with $M_p(T_1) < +\infty$, $M_q(T_2) < +\infty$. Given that $\left(\mathscr{C}_{p_1}^0\right)^* \subset \left(\mathscr{C}_p^0\right)^*$ for all $p \leqslant p_1$ we can dispense with the condition $q \geqslant p$ in (12.64) and we can write $T \in \Lambda_P^*$ in the form (12.64) for any p, q.

Remark 12.28 The space in (12.63) is not a priori the same space as $\mathscr{C}^*(U,\mathrm{pol}) + b\mathscr{C}^*(U,\mathrm{pol}) = \overline{\Lambda}_P$. On the other hand, we may be hard-pushed to produce an example where $\overline{\Lambda}_P \neq \Lambda_P^*$. (The author confesses that he has not spent any time on this issue!)

The flat seminorms To clarify these definitions further we shall define flat seminorms (possibly $+\infty$; cf. §10.2.6)

$$F_p(T) = \inf\left[M_p(T_1) + M_p(T_2);\ T = T_1 + bT_2\right]; \quad T \in \mathscr{D}'. \tag{12.65}$$

With this definition, $T \in \Lambda_P^*$ if and only if $F_p(T) < +\infty$, $p \geqslant 0$.

Remark and exercise A more transparent way of presenting (12.53) in the case of a Lie group goes as follows. With the notation of §12.5.1 we can represent any integration current T on G as

$$T = \sum T_I \omega_I, \tag{12.66}$$

where the T_I can be defined and identified globally (once the orthonormal basis $\omega_1, \ldots, \omega_n$ in §12.5.1 has been fixed) with Radon measures on G. (Strictly speaking, as in de Rham, 1960, §8, p. 42, each T_I is a current of degree 0. To make this a measure μ_I, we set $\langle \mu_I, \varphi\rangle = \pm T[\varphi\, \mathrm{d}\omega_J]$, with $\varphi \in C_0^\infty$, for the complementary multi-index J.) Let us denote $|T| = \sum |\mu_I|$ (see Bourbaki, 1963) which is then a positive Radon measure on G. Condition (12.53) then reads

$$\int |T|\,(1+|m|)^p < +\infty; \quad p \geqslant 1. \tag{12.67}$$

Alternatively and equivalently,

$$\text{total mass of}[G \setminus \text{ball of radius } R] = \mathrm{O}(R^{-p}), \quad \text{for all } p \geqslant 0, \qquad (12.68)$$

that is, that total mass decays superpolynomially.

Suggestion for the reader Use of the dual space (12.63) and Remark 12.28 will be non-essential and marginal. Those readers not familiar with this more sophisticated theory of TVS need not worry therefore.

12.8 Geometric Properties of $\mathscr{C}(U,\mathrm{pol})$ and a 'Technical' Review of Currents

In this section we shall present several technical lemmas, properties and definitions on currents of polynomial growth. Then, as we did in §12.3, we shall help the reader to do some more 'surfing' of de Rham (1960) and highlight some additional facts about the theory of currents. These will be used in the course of this and the next chapter. The reader is advised to skim through this section and then refer back as necessary.

12.8.1 Images by polynomial mappings

Let $f\colon (U,\mathrm{O}) \to (U_1,\mathrm{O}_1)$ be a polynomial mapping between Riemannian manifolds with base points as in §12.2.4. We shall denote by f^* the pullback mapping on differential forms and by f_* the dual pushforward mapping whenever defined, for example on $\mathscr{E}'(U)$ (see de Rham, 1960, §11). What counts is that we then have

$$f^*\mathscr{C}(U_1,\mathrm{pol}) \subset \mathscr{C}(U,\mathrm{pol}), \quad f^*\mathscr{C}_p(U_1) \subset \mathscr{C}_{cp+c}(U); \quad p \geqslant 0, \qquad (12.69)$$

$$f_*\mathscr{C}^*(U,\mathrm{pol}) \subset \mathscr{C}^*(U_1,\mathrm{pol}), \quad M_p(f_*T) \leqslant CM_{cp+c}(T); \quad T \in \mathscr{D}', \quad (12.70)$$

for $c, C > 0$ independent of p. From this, since f^* and f_* commute with d and b, we deduce that $f^*\Lambda_P \subset \Lambda_P$ and $f_*\Lambda_P^* \subset \Lambda_P^*$.

Exercise 12.29 Prove the second assertion in (12.69). The rest easily follows. The verification there is easy enough because it suffices to make it on forms α of degree 0, that is, functions, and there it suffices to use $|f(m)|_1 \leqslant C(1+|m|)^C$ in (12.8), and for forms $e_1, e_2, \ldots \in \mathscr{C}(U_1,\mathrm{pol})$ of degree 1 and use the definition for df in (12.7). More explicitly, $\langle f^*e_i, \xi\rangle = \langle e_i, \mathrm{d}f\xi\rangle$, $|\mathrm{d}f\xi| \leqslant C(1+|m|)^C|\xi|$ for $\xi \in T_m(U)$. The verification for general forms is then done by expressing any such form φ (locally) as a linear combination of forms that are simple monomials of the form $\alpha e_{i_1} \wedge e_{i_2} \wedge \cdots$ (cf. (12.40)).

The above also holds when f is not smooth but only locally Lipschitz, but then, in the target space of f^*, we have to allow the larger space $\mathscr{C}(U;\mathrm{pol}, L^\infty_{\mathrm{loc}})$ of §12.5.3, and no use will be made of this aspect of things.

12.8.2 Polynomial mappings in Lie groups

Here G is some connected Lie group assigned with the left-invariant Riemannian metric and $e \in G$ as the base point.

(i) The simplest polynomial mappings are the left and right actions

$$s_a, \tau_a \colon G \longrightarrow G; \quad a \in G, \quad \tau_a g = ag, \quad s_a g = ga; \quad g \in G. \tag{12.71}$$

The constants of these polynomial mappings of course depend on a.

Sometimes it will be essential to keep track of these constants. The best way to do that is to introduce new notation and set

$$M_p^a(T) = M_0\left(T(g)(1+|a^{-1}g|)^p\right), \quad a \in G,\ T \in \mathscr{D}'(G),\ p \geqslant 0, \tag{12.72}$$

which essentially amounts to changing the base point. Then it is clear that for the left translation we have

$$M_p^a\big((\tau_a)_* T\big) = M_p(T); \quad M_p\big((\tau_a)_* T)\big) \leqslant C\big(1+|a|\big)^p M_p(T), \tag{12.73}$$

because we are using a left-invariant Riemannian structure. The second estimate in (12.73) follows from the first and from (12.72).

Nothing like this can of course be done for right translation and there, in general, the constants that would appear in (12.73) grow exponentially in $|a|$.

Exercise 12.30 (i) Use §8.1 to elaborate on this. For right translations, for the constants to stay bounded, in general a has to stay in some compact subset of G.

(ii) The product mapping is

$$G \times G \longrightarrow G, \quad (g_1, g_2) \longrightarrow g_1 g_2; \quad g_1, g_2 \in G. \tag{12.74}$$

If we go back to §8.1 again we see that the restriction of this mapping,

$$(g_1, g_2) \longrightarrow g_1 g_2; \quad g_1 \in G,\ g_2 \in K,\ K \subset\subset G,\ K \text{ compact}, \tag{12.75}$$

is polynomial but the constants depend on K.

(iii) A mapping used in homotopies. Let

$$\exp \colon \mathfrak{g} \longrightarrow G; \quad \exp(\xi) \in G,\ \xi \in \Omega \subset \mathfrak{g} \tag{12.76}$$

be the exponential mapping (Varadarajan, 1974, §2.10) that is defined and is a bijective diffeomorphism in some compact neighbourhood $\Omega \subset \mathfrak{g}$ of 0 in the

Lie algebra $\mathfrak{g}$. This mapping has already played an important role in several places in the analytic theory (e.g. §§3.4 or 3.5).

For every $g \in \exp(\Omega) \subset G$ such that $g = \exp(\xi)$ for some $\xi \in \Omega$ we can then define a homotopy:

$$F(x,t) = xg_t \in G; \quad g_t = \exp(t\xi),\ 0 \leqslant t \leqslant 1,\ x \in G,\ 0 \leqslant t \leqslant 1. \qquad (12.77)$$

This is clearly a polynomial mapping (from $G \times [0,1]$ to G) with constants that are uniform in g as long as Ω stays fixed.

Exercise 12.31 Use §8.1 to prove this. Observe that the same thing cannot be asserted in general for $g_t x$, because there the second factor runs through the generally non-compact G (see §14.3.2 below).

12.8.3 Double forms and double currents

Here, until the end of the section, we shall pick up again our review of the theory of currents that we started in §12.3, and explain some of the technical aspects of the theory that will be needed in the proofs. We shall conform closely to the notation of de Rham (1960) and consider two differential manifolds V, W. De Rham uses the notation $\mathscr{E}(V \times W)$ (resp. $\mathscr{D}(V \times W)$), *not* to indicate $\mathscr{E}(M)$ (resp. $\mathscr{D}(M)$) for the manifold $M = V \times W$, but the space of *double C^∞* forms and the space of such forms that are compactly supported. We then define $\mathscr{E}(V \times W)$ as the space of differential forms on V with coefficients in $\mathscr{E}(W)$ or, equivalently, vice versa, differential forms on W with coefficients on $\mathscr{E}(V)$. A more symmetric definition is also given in de Rham (1960, §§7,12) and similarly for $\mathscr{D}(V \times W)$.

The duals of these spaces for the appropriate topologies are defined to be $\mathscr{D}'(V \times W)$ the space of double currents, and $\mathscr{E}'(V \times W)$ the subspace of compactly supported *double* currents.

To make an essential clarification of the notion we can identify $\varphi \in \mathscr{D}(M)$, that is, an ordinary form on $M = V \times W$, with a double form 'simply by bringing in the product structure in the definition'. By this cryptic statement we simply mean the following.

If $U \subset V$ and $U' \subset W$ are local charts in V and W respectively (this is the notation in de Rham, 1960, pp. 36, 58) then coordinate changes in U and U' put together make up a coordinate change in $U \times U' \subset V \times W = M$. This simply means that if $\tilde{\gamma} \in \mathscr{D}(M)$ we can look at it as a double form γ. In de Rham's notation (de Rham, 1960, §13) this is written out as $\gamma = \mathscr{A}^* \tilde{\gamma}$. By duality a double current $L \in \mathscr{D}'(V \times W)$ corresponds to a current in $M = V \times W$ that is denoted $\mathscr{A}L$ and is defined by $\mathscr{A}L[\tilde{\gamma}] = L[\mathscr{A}^* \tilde{\gamma}]$.

This set of notions and notation may at first sight appear rather 'heavy', especially if one is to take this seriously and follow in detail the presentation of de Rham (1960, §13). The point, however, is that sometimes it is necessary to distinguish between the variables in a product manifold; for example, in considering homotopies $F(m,t)$, $(m,t) \in M \times \mathbb{R}$ in (12.3), the variables m and t play very different roles on this product manifold. The ordinary forms on this manifold are then best considered as double forms. This will be examined in more detail in §12.9 below. And a full detailed, but also more energy-consuming, exposition can be found in de Rham (1960, §14).

12.8.4 Application of double currents

Here we shall specialise the set-up of the previous subsection to $V = W = G$ some Lie group, and, for two currents $T_1, T_2 \in \mathscr{D}'(G)$, define a double current $T_1 \times T_2 \in \mathscr{D}'(G \times G)$ because the algebraic tensor product $\mathscr{D}(G) \otimes \mathscr{D}(G) \subseteq \mathscr{D}(G \times G)$ can be identified with a dense subspace of the space of double forms; see de Rham (1960, p. 59). Here, if $\varphi_1, \varphi_2 \in \mathscr{D}(G)$, then $\psi = \varphi_1 \cdot \varphi_2$ is a double form (see de Rham, 1960, §7, p. 36, and (12.81) below) and we set $\langle T_1 \times T_2, \psi \rangle = T_1[\varphi_1] T_2[\varphi_2]$. Note that the use of '$\times$' that we make here deviates from de Rham's notation.

Exercise 12.32 This fact is well known (see Schwartz, 1957) if $V = \mathbb{R}^n$, $W = \mathbb{R}^m$. Therefore this tensor product is dense in the space of double forms with their support in $U \times U'$ for two charts U, U' in V and W respectively because of the representation (12.81) below. For the general case we use finite partitions of unity to take care of the compact support.

We shall consider now the product mapping

$$\pi\colon G \times G \longrightarrow G; \quad \pi(g_1, g_2) = g_1 g_2. \tag{12.78}$$

It follows that as long as the support of one of the two currents T_1 or T_2 is compact, the direct image

$$\pi_* \mathscr{A}(T_1 \times T_2) = T_1 \cdot T_2 \subset G \tag{12.79}$$

can be defined because the mapping is proper on the support of $T_1 \times T_2$ (see de Rham, 1960, pp. 56, 58). The convolution sign $*$ is normally used for measures.

Example 12.33 With the notation of (12.71) and §12.7.1,

$$(\tau_a)_* T = \delta_a \cdot T, \quad (s_a)_* T = T \cdot \delta_a; \quad T \in \mathscr{D}'(G). \tag{12.80}$$

We shall elaborate further on the above definitions in §12.8.5.2 below.

12.8.5 The boundary operators on double currents and the 'produit tensoriel'

The same notation as before will be used and, once more, the notation of de Rham (1960) will be maintained when possible. For the local charts $U \times U' \subset V \times W$, the local coordinates in U and U' will be $x^1, \dots, x^n$ and $y^1, \dots, y^m$, so that $\dim V = n$, $\dim W = m$. With the notation of de Rham (1960, p. 36), a double form can then be represented as

$$\gamma(x,y) = \sum_{I,J} C_{I,J}(\mathrm{d}x^I) \cdot (\mathrm{d}y^J); \quad I = i_1 < \cdots < i_p,\ J = j_1 < \cdots < j_q. \quad (12.81)$$

The degree of this double form is said to be p in x and q in y if, in the summation (12.81), only the coefficients $C_{I,J}$ for which $|I| = p$ and $|J| = q$ are non-zero. Similarly (de Rham, 1960, pp. 57–58), a double current L is said to be of dimension p in x and q in y (equivalently of degree n–p in x and m–q in y) if $L[\varphi] = 0$ for all double forms φ that are not of degree p in x and q in y. On such (homogeneous) double currents one then defines the operators

$$w_x L = (-1)^p L, \quad w_x^* L = (-1)^{n-p} L; \quad p = \deg L \text{ in } x, \qquad (12.82)$$

and then extends to the whole space of double currents by linearity.

On double forms one defines the two exterior differentials d_x and d_y in the x and y variables and these commute (de Rham, 1960, p. 37). One then defines on currents the dual operators by $b_x L[\gamma] = L[\mathrm{d}_x\gamma]$ and $\mathrm{d}_x = w_x b_x$ (de Rham, 1960 p. 58). Similar definitions are given for w_y, w_y^*, d_y, b_y on the same space of double currents (de Rham, 1960 p. 58).

All this keeping track of the $\pm$ signs in the definitions of the boundary may seem unduly complicated at first sight. The reader should not take all this too seriously in a first reading of de Rham (1960) but should nevertheless bear in mind that it is important and that this problem is also intrinsic in homology theory in general (see Massey, 1991, Chapter VIII). The reader should rather concentrate on formula (12.83) below and on the illustrations of this formula that are given next. There the geometric meaning of the operators b and w should become apparent.

12.8.5.1 The formula for the boundary Let L be some double current on $V \times W$ with which we can associate an ordinary current on the manifold $M = V \times W$. This current is denoted by $\mathscr{A}L$ in de Rham (1960, §13) and then $\mathscr{A} : L \to \mathscr{A}L$ becomes a linear operator. The formula that relates b, the boundary operator on currents on M, with the boundary operators on the currents on V and W can, with our previous notation, be written out as the following identity between operators (de Rham, 1960, p. 62, §13):

$$b\mathscr{A} = \mathscr{A}\left(b_x + w_x^* b_y\right). \tag{12.83}$$

This says, *grosso modo*, that to take the boundary in the product we have to combine the two boundaries in the x-variables and the y-variables. The additional twist is that in the second component a $\pm$ crops up, depending on the dimension of the first component. This is very much inherent in homology theory and among other things it guarantees that the identities $b^2 = 0$, $b_x^2 = 0$, $b_y^2 = 0$ are compatible (see Example 12.37 below). We hope that the geometric illustration in the next subsection will help to clarify matters.

12.8.5.2 The commutative produit tensoriel The space $\mathscr{D}(V \times W)$ of double forms γ of (12.81) is defined in de Rham (1960, p. 36) as the space of differential forms on V with coefficients in $\mathscr{D}_W$ (this is de Rham's notation for $\mathscr{D}(W)$). Similarly, $\mathscr{D}(W \times V)$ can be defined, and the obvious identification $J\colon \mathscr{D}(V \times W) \to \mathscr{D}(W \times V)$ is used tacitly and throughout (de Rham, 1960, p. 36 again).

Now let $\gamma(x,y) \in \mathscr{D}(W \times V)$, and also let $T(x) \in \mathscr{D}'(V)$, $S(y) \in \mathscr{D}'(W)$ (de Rham, 1960, p. 59). The coefficients of γ lie in $\mathscr{D}_V$; we can then act by $T(x)[\gamma(x,y)] \in \mathscr{D}_W$ and on this can apply S and set $L[\gamma] = S(y)\left[T(x)[\gamma(x,y)]\right]$ so that $L \in \mathscr{D}'(W \times V)$. This procedure can be done the other way round and we obtain $L' \in \mathscr{D}'(V \times W)$ by $L'[\gamma] = T(x)\left[S(y)[\gamma(x,y)]\right]$. Testing on the special double forms $\gamma = \alpha(x)\beta(y)$ we see that, modulo the identification J, these are identical.

The notation $L = ST = S(y)T(x)$, $L' = TS = T(x)S(y)$ is used in de Rham (1960, p. 59), where this product is called *le produit tensoriel*, and is commutative. With this commutative product we clearly have $b_x(TS) = (bT)S$, $b_y(TS) = T(bS)$. Examples of this are used in the homotopy formulas (see §12.9 below and in de Rham, 1960, p. 67) where the product $IT = I(t)T(y) \in \mathscr{D}'(\mathbb{R} \times W)$ is considered for the current $I = [0,1] \in \mathscr{D}'(\mathbb{R})$ of the next subsection. In §12.8.4 we prefer to denote $L' = T \times S$ and $L = S \times T$ and say that $J^*L = L'$ for the dual identification.

We shall now stick with de Rham's notation and, with T, S as above we shall define the double current $R = TS = ST$ on the product of V with W. We can use the operator $\mathscr{A}$ on the two manifolds $M = V \times W$ and $N = W \times V$ to define the two currents $\mathscr{A}_M R$, $\mathscr{A}_N R$ on M and N respectively. Notice that if $\theta\colon M \to N$ is given by $\theta(x,y) = (y,x)$, then $\theta\mathscr{A}_M R = \mathscr{A}_N R$ (to see this test on some forms $\varphi(x) \wedge \psi(y)$). Now let U be a third manifold and let $\alpha\colon U \to V$, $\beta\colon U \to W$ be diffeomorphisms. The two diffeomorphisms $\alpha \times \beta\colon U \times U \to M$, $\beta \times \alpha\colon U \times$

$U \to N$ can then be used to define the two currents $(\alpha \times \beta)^{-1}\mathscr{A}_M R = A$ and $(\beta \times \alpha)^{-1}\mathscr{A}_N R = B$ on the manifold $U \times U$.

Now let $U = G$ be a Lie group and let π denote the group product of (12.78). In the above terms, definition (12.79) now becomes

$$(\alpha^{-1}T)\cdot(\beta^{-1}S) = \pi A, \quad (\beta^{-1}S)\cdot(\alpha^{-1}T) = \pi B.$$

This should put the reader at ease and explain why, despite the fact that the produit tensoriel is commutative, the product (12.79) on the group in general is not. We shall return to this point in (13.99) below.

Example 12.34 Let $V = W = G$ be some Lie group and let π be the product mapping of (12.78). If $T_1, T_2 \in \mathscr{D}'(G)$ are such that one of them has compact support, then the product $T_1 \cdot T_2 = \pi_* \mathscr{A}(T_1 \times T_2) \in \mathscr{D}'(G)$ can be defined. To simplify notation we shall also assume that T_1 is homogeneous of dimension r. Then if in formulas (12.83) and (12.82) we apply the pushforward mapping π_* we obtain the formula

$$b(T_1 \cdot T_2) = (bT_1)\cdot T_2 + (-1)^r T_1 \cdot (bT_2). \tag{12.84}$$

12.8.6 Examples of currents in $\mathbb{R}^n$

One item of notation that we shall use systematically and that is also used in de Rham (1960, §14, p. 67) is the current $I \in \mathscr{E}'(\mathbb{R})$ in the real line defined formally by

$$I[\varphi] = \int_0^1 \varphi; \quad \varphi = \varphi_1\, dx \in \mathscr{D}_1(\mathbb{R}),\ \varphi_1 \in C_0^\infty. \tag{12.85}$$

Here the notation is as in §12.4.2 and I vanishes on $\mathscr{D}_0(\mathbb{R})$. The more general definition of $I^n \in \mathscr{E}'(\mathbb{R}^n)$ of a current of dimension n is given by the chain $c = I^n = [0,1]^n \subset \mathbb{R}^n$ (with the canonical embedding) as in §12.3.4 where the orientation $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n =$ Lebesgue measure on $\mathbb{R}^n$ will be fixed. Similarly, the subspaces $\mathbb{R}^p = (0,\ldots,x_{i_1},\ldots,0,x_{i_2},\ldots)$, where all but p coordinates are 0, can be oriented by $\mathrm{d}x_{i_1} \wedge \cdots =$ Lebesgue measure on $\mathbb{R}^p$. If $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q$, where $\mathbb{R}^p$ involves the first p coordinates and $\mathbb{R}^q$ involves the last $q = n - p$ coordinates, and if I^p and I^q are the corresponding currents on $\mathbb{R}^p$ and $\mathbb{R}^q$, then the product double current $I^p \times I^q \in \mathscr{E}'(V \times W)$ can be defined as in §§12.8.4, 12.8.5.2 with $V = \mathbb{R}^p$, $W = \mathbb{R}^q$. By the definitions we then have $I^n = \mathscr{A}(I^p \times I^q)$; that is, I^n is the current on $\mathbb{R}^n$ that corresponds to this double current.

Remark 12.35 We can also consider $\mathbb{R}^n$ as a Lie group; the above two subspaces are then subgroups (and $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q = \mathbb{R}^p \oplus \mathbb{R}^q$ is the direct product group). By the pushforward of the natural injections $\mathbb{R}^p, \mathbb{R}^q \to \mathbb{R}^n$ we can then identify $I^p, I^q \in \mathscr{E}'(\mathbb{R}^n)$ as currents on $\mathbb{R}^n$. Then, with the notation of (12.79), we have $I^n = I^p \cdot I^q$.

Now, the boundary bI^n is a current whose support is the geometric boundary $\partial[0,1]^n$ and corresponds to $(n-1)$-Lebesgue integration on the affine hyperplanes with $\pm$'s so that Stokes' formula holds (see de Rham, 1960, §6, p. 31):

$$\int_{I^n} \mathrm{d}\varphi = \int_{bI^n} \varphi, \qquad (bI^n, \varphi) = (I^n, d\varphi); \quad \varphi \in \mathscr{D}(\mathbb{R}^n). \tag{12.86}$$

The boundary of I in (12.86) is simple enough to work out. With the notation of (12.85) and §12.3.6 we have

$$bI = \delta_1 - \delta_0; \quad I[\mathrm{d}\varphi] = \int_0^1 \mathrm{d}\varphi = \varphi_1(1) - \varphi_1(0). \tag{12.87}$$

We can use formula (12.83) and the product structure of I^n to work out the signs needed for the Lebesgue measure on the affine hyperplanes to make formula (12.86) hold good.

Example 12.36 Suppose $n = 2$ and let $I^2 = I_1 \times I_2$ be the unit square in the plane where $\operatorname{supp} I_1 = [0 \leqslant x_1 \leqslant 1,\ x_2 = 0]$ and vice versa for I_2 for the coordinates $(x_1, x_2) \in \mathbb{R}^2$. Here the orientation is such that $\mathrm{d}x_1 \wedge \mathrm{d}x_2$ is the Lebesgue measure. The current I^2 is thus defined and the geometric definition of the orientation on ∂I^2 is given by the inner normal and the usual rule. This orientation on ∂I^2 makes (the geometric) Stokes' theorem

$$\int_{I^2} \mathrm{d}\varphi = \int_{\partial I^2} \varphi \tag{12.88}$$

work (see Warner, 1971) and defines a current in $\mathscr{E}'(\mathbb{R}^2)$ which has to be bI^2 in the general version of Stokes' theorem in (12.86). This of course amounts to the definition of the b-operator for currents.

So here (12.83) is what it takes to specify algebraically the orientations on the boundary that make Stokes' theorem work.

Example 12.37 Connections with the cubical homology in algebraic topology (see e.g. Hilton and Wylie, 1960 or Massey, 1991). Another equivalent and very closely related way of defining the singular homology ($\mathscr{S}$) of a topological space is by considering the complex spanned by all continuous mappings, not of the standard simplex Π (as in §12.4.3 or again in the next section, §12.9) but by the mappings of the unit cubes I^n. The boundary operator $\partial-$ in this complex ($\mathscr{S}_{\text{cub}}$) is then defined by adding with $+$ or $-$ the images of the faces of the cubes in ∂I^n. And of course these ± 1 are none other than the ones obtained by applying formula (12.83) repeatedly on the coordinate. This is the way things should be if we are to have the conclusion $(\mathscr{S}_{\text{cub}}, \partial) \subset (\mathscr{E}', b)$. (Of course, for this to hold, exactly as in §12.4.3, we must also make the construction of the cubical homology with *Lipschitz* mappings of the unit cubes into the manifold.)

Exercise 12.38 Check things out.

12.9 The Use of Polynomial Homotopy in the Complexes $\Lambda_P(U), \Lambda_P^*(U)$

12.9.1 The abstract chain homotopies. The literature

Even with rudimentary exposure in algebraic topology one knows that one cannot go very far in homology theory without the systematic use of homotopy (see Hilton and Wylie, 1960; Massey, 1991). For the homology in the complex of currents one finds a systematic exposition of this in de Rham (1960, §14) and also Federer (1969, §4.1.9). Here we shall give a brief introduction to this, enough to show how things work. For more details one will have to refer to the above references.

The use of homotopy that we shall make in §12.9.7 is to show that when the Riemannian manifold U is polynomially retractable to a point then the complexes $\Lambda_P(U)$ and $\Lambda_P^*(U)$ are acyclic. (The second interpretation of the complex $\mathscr{E}'$ in §12.4.2 is then used!) The way this is done consists in constructing in these complexes a chain homotopy. Let us recall what this means for a general complex as in §12.4.1. A chain homotopy is then a sequence of mappings $h = (h_n\colon \Lambda_n \to \Lambda_{n-1};\ n \in \mathbb{Z})$ that satisfies

$$d_{n-1} \circ h_n + h_{n+1} \circ d_n = \text{Identity if } n \neq 0. \tag{12.89}$$

Certainly, for $n \neq 0$ if $\xi \in \Lambda_n$ is a cycle, that is, $\mathrm{d}\xi = 0$, then $\xi = \mathrm{d}h_n\xi$ is a boundary and $[\xi] = 0$. This means that $H_n(\Lambda) = 0, n \neq 0$ and Λ is acyclic. The diagram is as follows:

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & \Lambda_{n-1} & \stackrel{\mathrm{d}_{n-1}}{\longrightarrow} & \Lambda_n & \stackrel{\mathrm{d}_n}{\longrightarrow} & \Lambda_{n+1} & \longrightarrow & \cdots \\ & & & {}^{h_n}\swarrow & & {}^{h_{n+1}}\swarrow & & & \\ \cdots & & \Lambda_{n-1} & \stackrel{\mathrm{d}}{\longrightarrow} & \Lambda_n & \longrightarrow & \cdots . & & \end{array} \tag{12.90}$$

This is a special case of a chain homotopy between two chain maps $f, g\colon (\Lambda, d) \to (\Lambda', d')$ between two complexes (i.e. f, g intertwine d and d'). The maps that are induced on the homologies $f_*, g_*\colon H(\Lambda) \to H(\Lambda')$ are then identical. (Exercise: check or work this out; see the end of the next section or Cartan and Eilenberg, 1956, §IV.3.)

We shall illustrate this in the concrete situation of singular homology in the next subsection.

12.9.2 Brief overview of chain homotopy in singular homology

The complex $\mathscr{S} = (\mathscr{S};X)$ of the singular homology on a topological space X is the vector space (over $\mathbb{R}$) whose basis consists of all singular simplexes, that is, the continuous mappings

$$\varphi\colon \Pi \longrightarrow X; \quad \Pi = (0, e_1, \dots, e_n) \subset \mathbb{R}^n,\ n \geqslant 0,\ e_j = (0, \dots, 1, 0, \dots) \quad (12.91)$$

for the standard simplexes $\Pi = \Pi_n$ (see §12.4.3). The boundary operator $\partial\varphi = \varphi_0 - \varphi_1 + \cdots$ for $\varphi_j = \varphi|_{\Pi_j}$ = the restriction on the boundary simplexes (see Hilton and Wylie, 1960, or Dubrovin et al., 1990, for details).

Let $f, g\colon X \to Y$ be two continuous mappings that are homotopic as in §12.2.1 and

$$F\colon X \times [0,1] \longrightarrow Y; \quad F(x,0) = f(x),\ F(x,1) = g(x),\ x \in X. \quad (12.92)$$

Then f, g induce chain mappings on the corresponding singular complexes and these induce mappings

$$f_*, g_*\colon H(\mathscr{S};X) \longrightarrow H(\mathscr{S};Y) \quad (12.93)$$

on the singular homologies by setting $f \circ \varphi$, $g \circ \varphi$ for the singular simplexes. The fundamental fact about the homotopy F is that it guarantees that $f_* = g_*$. The verification of this fact is well known (see Hilton and Wylie, 1960) but it is also very instructive in what we shall do for currents in the next subsection. In the next few lines we shall therefore recall how this is done (see also Dubrovin et al., 1990, §5.2, Figure 35).

In Figure 12.1 we show how in the cases where the dimension $n = 1, 2$ the prism $\Pi \times [0,1]$ and the boundary prisms of $\partial\Pi \times [0,1]$ can be decomposed into simplexes. Geometrically, we clearly have for the boundaries of these prisms

$$\partial\,(\Pi \times [0,1]) = \partial\Pi \times [0,1] \cup (\Pi \times \{0\}) \cup (\Pi \times \{1\}). \quad (12.94)$$

Now let $(\Pi;\varphi)$ be the corresponding singular simplex in X (Figure 12.2). The mapping $f \circ \varphi$, $g \circ \varphi\colon \Pi \to Y$ can be extended to a mapping from $\Pi \times [0,1] \to Y$ that is given by

$$\Pi \times [0,1] \ni (\xi, t) \longrightarrow F\,(\varphi(\xi), t) \in Y. \quad (12.95)$$

In Figure 12.2 we indicate by f_*, g_* the mappings that are induced from the singular simplexes of X to singular simplexes of Y. By linearity this induces also a mapping from the singular complex of X to the singular complex of Y.

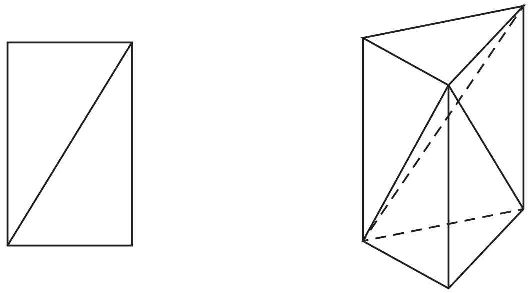


Figure 12.1 Decomposition of $\Pi_m \times [0,1]$ into simplexes for $m = 1,2$.

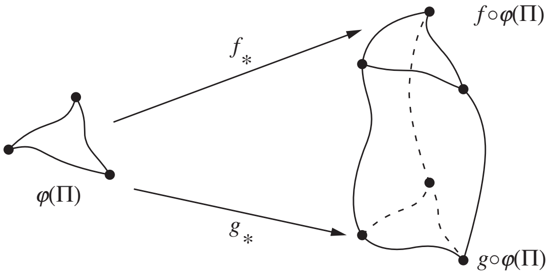


Figure 12.2 The mappings on Π_2 and on the prism.

Observe that for this the *appropriate* identification of the simplexes of Figure 12.1 with the corresponding standard simplex as in (12.91) has to be made before we actually obtain singular simplexes.

What follows from the above is that once $(\Pi;\varphi)$ is given we determine singular simplexes

$$f_*(\Pi;\varphi) = (\Pi; f \circ \varphi), \quad g_*(\Pi;\varphi) = (\Pi; g \circ \varphi) \tag{12.96}$$

as shown in Figure 12.2. We can also define a singular chain which is the sum of the singular simplexes of higher dimension that are determined by the images under the mapping (12.95) of the simplexes of the decomposition of $\Pi \times [0,1]$ of Figure 12.1 (the dimension is 3 in the right-hand figure). This chain will be denoted by $F_*(\Pi;\varphi)$.

This mapping extends by linearity to the whole singular complex

$$F_*\colon (\mathscr{S};X) \to (\mathscr{S};Y)$$

and increases the dimension by 1. This is like the mapping (12.90) for two different complexes. The claim is that, if the identification of the standard simplexes (12.91) with simplexes that decompose $\Pi \times [0,1]$ and $\partial\Pi \times [0,1]$ is done properly, we have the fundamental identity that characterises chain homotopies

$$F_* \circ \partial + \partial \circ F_* = f_* - g_*. \tag{12.97}$$

This, in less precise but perhaps more transparent notation, can also be written as

$$\partial F_* \Pi + F_* \partial \Pi = f \circ \varphi \Pi - g \circ \varphi \Pi. \tag{12.98}$$

Seen like this (12.98) appears simply as the result of applying the mapping F of (12.95) to Figure 12.1. Finally, by the last statement in the previous subsection, we conclude that the identity (12.97) implies that the induced mappings f_*, g_* on the singular homologies are identical (see Cartan and Eilenberg, 1956).

Exercise 12.39 Prove this. If $c \in (\mathscr{S}, X)$ is a cycle, that is, if $\partial c = 0$ then $f_* c - g_* c = \partial F_* c$ and therefore the difference is a boundary.

12.9.3 Chain homotopy on $\mathscr{E}'$. Heuristics

Let W, V be C^∞ manifolds. We saw in §12.4.3 how the Lipschitz singular chains of $(\mathscr{S}; W)$ in §12.9.2 (note here we assume that the mapping φ in (12.91) is Lipschitz) can be identified with currents in $\mathscr{E}'(W)$. This identifies this 'Lipschitz' singular simplex $(\mathscr{S}; W, \mathrm{Lip})$ to a subcomplex of $\mathscr{E}'(W)$. One can easily see that the way the definition of ∂ was done was designed to make Stokes' theorem work so that the boundary operators ∂ of $\mathscr{S}$ and b as in (12.26) coincide in this identification. One can also prove (cf. §12.4.3) that in the definition of singular homology we can restrict ourselves to Lipschitz singular simplexes.

Now let $f, g \colon W \to V$ be Lipschitz mappings that are homotopic by a Lipschitz homotopy F as in (12.3). Here, to be as close as possible to the notation of de Rham (1960, §14), we shall denote this homotopy by

$$\mu(t, y) = \mu_t y = x, \ \mu_1 = f, \ \mu_0 = g; \quad y \in W, \ x \in V, \ t \in \mathbb{R}. \tag{12.99}$$

The aim is to construct a chain homotopy $M \colon \mathscr{E}'(W) \to \mathscr{E}'(V)$ that satisfies the formula

$$\mu_1 T - \mu_0 T = f_* T - g_* T = bMT + MbT; \quad T \in \mathscr{E}'(W), \tag{12.100}$$

where the $*$ in the direct image of currents is inserted to highlight the analogy with (12.97). The mapping M increases by 1 the dimension of homogeneous currents.

Insofar that $\mu_1 = f_*$ and $\mu_0 = g_*$ is the natural extension from the space of

Lipschitz chains of $(\mathscr{S}, W, \mathrm{Lip})$ to $\mathscr{E}'(W)$, the natural way of viewing – and possibly also proving – (12.100) is that M is an extension of the mapping F_* of the previous subsection.

The above idea could probably be carried out as long as the currents $T \in \mathscr{E}'(W)$ are integration currents because then the space of Lipschitz chains in $(\mathscr{S}, W, \mathrm{Lip})$ could possibly be seen to be dense in $\mathscr{E}'(W)$. For instance, if $\dim T = 0$ and T is a Radon measure in W this is clear enough, because we can approximate T by a linear combination of δ-masses. Things, however, get messy in higher dimensions and even messier for currents that are not integration currents. So something else has to be done.

That 'something else' is what is done in de Rham (1960, §14). We shall briefly outline this construction in the next few lines but for the details the reader will have to fall back to de Rham's exhaustive/ing exposition.

Remark 12.40 A raison d'être of flat currents is formula (12.100). Indeed, for a normal current T (i.e. T, bT integration current) the current on the right-hand side of (12.100) is not necessarily normal because of the term bMT, but it is flat as long as in our construction MT is an integration current. For more on this and other subtle related points in geometric measure theory, see Federer (1969, §4.1.13).

12.9.4 The construction of the chain homotopy on $\mathscr{E}'$

Figures 12.1 and 12.2 should be kept in mind and we shall recall the notation that was used in §§12.3, 12.8. As in §12.8.6, $I \in \mathscr{E}'(\mathbb{R})$ is the integration current on $[0,1]$, and $T \in \mathscr{E}'(W)$ and $IT \in \mathscr{E}'(\mathbb{R} \times W)$ is the product (produit tensoriel) double current as in §12.8.5.2. The current on the manifold $\mathbb{R} \times W$ that will be considered is $\mathscr{A}(IT)$. The notation $\mathscr{A}$ is used in de Rham (1960, §14) to indicate that a double current is considered as an ordinary current.

As explained in §12.8.5 and (12.87),

$$b\mathscr{A}(IT) = \mathscr{A}\big((bI)T\big) - \mathscr{A}(IbT) \in \mathscr{D}'(\mathbb{R} \times W), \tag{12.101}$$

$$bI = \delta_1 - \delta_0 \in \mathscr{D}'(\mathbb{R}). \tag{12.102}$$

From (12.102) we see that if we apply the mapping μ of (12.99) to $\mathscr{A}\big((bI)T\big)$ we obtain $\mu_1 T - \mu_0 T$. Therefore the same μ applied to (12.101) gives

$$\mu b\mathscr{A}(IT) + \mu\mathscr{A}(IbT) = \mu_1 T - \mu_0 T, \tag{12.103}$$

where μ (of a current) is the direct image of that current (see §12.3.7). Now μ and b commute and we shall set

$$MT = \mu\mathscr{A}(IT); \quad T \in \mathscr{E}'(W), \tag{12.104}$$

where we use de Rham's notation. With these we see that (12.103) is exactly the required formula (12.100) which is also de Rham (1960, §14, p. 68, formula (3)).

This derivation of formula (12.100) is considerably more direct and clear than for the corresponding formula (12.97) in singular homology where we had to 'chase around' the various triangulations. This is but one among several examples where the formalism of currents can indeed be used with profit.

12.9.5 The chain homotopy on $\mathscr{E}$

Having defined the homotopy operator M on $\mathscr{E}'$, the easiest way to define the corresponding homotopy operator M^* is to use the duality between $\mathscr{E}$ and $\mathscr{E}'$ and write

$$M^*\colon \mathscr{E}(V) \longrightarrow \mathscr{E}(W); \quad (M^*\varphi, T) = (\varphi, MT),\ \varphi \in \mathscr{E}(V),\ T \in \mathscr{E}'(W), \tag{12.105}$$

where the notation is as in (12.100).

The formula for the pullback mappings on forms by the two mappings then follows at once:

$$\mu_1^* - \mu_0^* = dM^* + M^*d. \tag{12.106}$$

This is de Rham (1960, §14, p. 69, formula (4)), and just before, one finds the explicit expression

$$M^*\varphi = I(t)[\mathscr{A}^*\mu^*\varphi], \tag{12.107}$$

where $I(t)$ stands for the current that we denoted by I in §12.9.4. The meaning of the product in (12.107) is explained in de Rham (1960, §12, Theorem 9, p. 59), and is exactly as in the definition of the produit tensoriel that we recalled in §12.8.5.2. We shall not elaborate further on de Rham's §14.

Rather, we shall rewrite (12.107) in a simple-minded way as is done in many of the elementary presentations of de Rham cohomology, and in doing this we shall use the local coordinates $(t, y_1, \dots, y_n)$ on $\mathbb{R} \times W$ (see (12.106)). We can then write

$$\mu^*\varphi = \sum_I a_I\, \mathrm{d}y_I + \sum_J b_J\, \mathrm{d}y_J \wedge \mathrm{d}t$$

for the usual notation $\mathrm{d}y_I = \mathrm{d}y_{i_1} \wedge \cdots$ for increasing multi-indices $I = (i_1 < \cdots)$. The explicit expression of (12.107) then becomes

$$M^*\varphi = \pm \sum_J \left(\int_0^1 b_J\, \mathrm{d}t \right) \mathrm{d}y_J, \tag{12.108}$$

for an appropriate choice of the $\pm$. We shall leave it as an exercise (consulting Dubrovin et al., 1990, §§1.3–1.4, if required) for the reader to verify that (12.106) holds.

A by-product of this construction is that it gives an illustration of the notion of double differential forms. Indeed, the construction is invariant under coordinate changes of the special kind $t \to t'$ and $(y_1, \ldots) \to (y_1', \ldots)$, but one is not allowed to mix the two. Note also that the above 'coordinate approach' to the problem is what is done at the end of de Rham (1960, §14), but there it comes out 'less simple minded'.

All in all, the moral of the story is that the two operators M and M^* are dual to each other; it suffices therefore to spell out the definition of just one and the other follows automatically. When choosing how to define M our primary motivation was the close connection that this has with classical homotopy (see §§12.9.2, 12.9.3) in algebraic topology.

12.9.6 Application to the homology theory of Lie groups

The main application of the chain homotopies of the previous section is that if, say, two manifolds $M_1 \underset{\beta}{\overset{\alpha}{\rightleftarrows}} M_2$ are homotopically equivalent as in §12.2.1 then their homologies $H(M_1) \cong H(M_2)$ are isomorphic. This of course is a very general fact and works for just about every homology theory (simplicial, singular, etc.) but to be specific let us assume that α, β are C^∞ and that the homotopy is also C^∞. Then the homologies of the two complexes $\mathscr{E}(M_1)$ and $\mathscr{E}(M_2)$ are isomorphic. This here of course amounts to the fact that $\dim H_n(\mathscr{E}(M)) = \beta_n(M)$ the Betti numbers (possibly $+\infty$) are the same for the two complexes.

In our context this applies to a Lie group G and $K_0 \subset G$ some maximal compact subgroup and the homotopy $G \simeq K_0$ of §12.1.3. This gives the very important and well-known fact on the Betti numbers $\beta_n(G) = \beta_n(K_0) < +\infty$.

For the same reason this also applies to the homologies of the complexes $\mathscr{E}'(M)$. This homology should be compared with the compactly supported cohomology H_c of the manifold (see de Rham, 1960, §19; Bott and Tu, 1982).

12.9.7 The polynomial homotopy and the complexes Λ_P, Λ_P^*

Let $f, g\colon (M, \mathrm{O}) = W \to (M_1, \mathrm{O}_1) = V$ be polynomial smooth (or just locally Lipschitz) mappings that are homotopic by a homotopy F, as in §12.2.1, (12.3) that is also polynomial. From §12.8 these induce mappings

$$f^*, g^*: \Lambda_P(V) \longrightarrow \Lambda_P(W), \tag{12.109}$$

$$f^*, g^*: H(\Lambda_P(V)) \longrightarrow H(\Lambda_P(W)). \tag{12.110}$$

Some care is needed here if we wish to insist that these should be defined for f, g locally Lipschitz. As we pointed out in §7.6.3 for such mappings, we need to consider the complexes $\Lambda_P(M; L^\infty_{\text{loc}})$ that come from $\mathscr{C}(M; \text{pol}, L^\infty_{\text{loc}})$ (cf. §12.5.2), the forms with L^∞_{loc} coefficients. This aspect of things will not be essential in what follows and we shall always assume f and g to be smooth. Be that as it may, the constructions of the chain homotopies that we gave in §12.9.4 extend and give chain homotopies

$$\begin{aligned} M&: \Lambda_P^*(W) \longrightarrow \Lambda_P^*(V), \\ M^*&: \Lambda_P(V) \longrightarrow \Lambda_P(W). \end{aligned} \tag{12.111}$$

Exercise 12.41 Use §12.8.1 to verify this. For M in (12.111) use the fact that $T \in \left(\mathscr{C}_p^0(W)\right)^*$ of (12.51) implies that $\mathscr{A}(IT) \in \left(\mathscr{C}_p^0(\mathbb{R} \times W)\right)^*$. To see this, (12.67) could be used. More formally, we can start by showing that in general $M_0(\mathscr{A}(TS)) = M_0(T)M_0(S)$, for two Riemannian manifolds with a base point and the corresponding Riemannian product structure and the notation TS for de Rham's produit tensoriel. (The notation of (12.49) and §12.8.5.2 are used here.) We apply this to (12.104). Then use §12.8.1 and (12.104). For M^* use the duality. Alternatively, of course this can also be seen from the elementary coordinate definition of M^* given at the end of §12.9.5. Not surprisingly this elementary verification is more tedious. Notice also that because of the homotopy formula, say (12.106), in effect, for (12.111), we only need to verify that M^*, say, maps $\mathscr{C}(U, \text{pol})$ into itself (see §12.5.2).

From (12.106), (12.111) we conclude, as in Exercise 12.39, that $f^* = g^*$ on the corresponding homologies in (12.110). We sum up:

If we make the assumption that W, V are polynomially homotopically equivalent (see §12.2.4) then $H(\Lambda_P(W)) \cong H(\Lambda_P(V))$ are isomorphic and the same for the Λ_P^* complexes.

12.9.8 Applications to the NC-condition

We have the following applications of the above.

(i) If Q is simply connected NC-soluble then $Q \simeq \{e\}$ is polynomially homotopically equivalent to a point (Theorems 7.10, 12.2). It follows that the complexes $\Lambda_P(Q)$ and $\Lambda_P^*(Q)$ are acyclic; that is, all the $H_n = 0$ for $n \neq 0$.

(ii) If G is simply connected NB then $G \simeq K$ is polynomially homotopically equivalent to a compact group K (see Theorem 12.6). It follows that the homologies of $\Lambda_P(G)$ and $\Lambda_P^*(G)$ are finite-dimensional.
(iii) The conclusion of (ii) holds for *all* connected groups (see Theorem 12.8). But the proof uses Appendix F and is more difficult.

The notation that is adopted here and throughout is of course the homological notation for the complex $\Lambda_P^*(G)$ and the differential decreases the dimension (cf. §12.4.2) and the index n in $H_n(\Lambda_P^*)$ indicates the *dimension*.

12.10 Regularisation

12.10.1 The setting of the problem

The smoothing operator we introduce below is very important in the homology theory of manifolds. But from our point of view it only becomes essential in Chapter 14 where we shall systematically be using the simplex $\Lambda_P \cap \mathscr{E}$ of smooth forms of polynomial growth. In concrete terms, we have already encountered this complex in Proposition 12.18 and, as promised, a proof of this proposition will be given in this section. But for those readers not wishing to surf de Rham (1960) more than they have to, this section will not be essential before, and if, they get as far as Chapter 14. We therefore suggest that the reader skips this section in a first reading.

Explicitly, using the homotopy formulas of the previous section, we shall explain a procedure that allows us to regularise currents by a linear operator:

$$\mathscr{D}'(U) \ni T \to RT \in \mathscr{E}(U) \tag{12.112}$$

on the C^∞ manifold U. This smoothing operator has the basic property that when T is closed then RT is too, and furthermore RT is homologous to T; that is, $T - RT \in b\mathscr{D}'(U)$. In what follows we shall restrict ourselves to the case where the manifold is a connected Lie group G. Note that this special case is exactly what is done in de Rham (1960, §15, p. 77), as a first step towards the general construction that is given by de Rham.

12.10.2 The construction of the regularising operator

We return to §12.8.2, and from the homotopy of (12.77) we shall define the homotopy operator S_g^*, with $g \in G$, as in §12.9.5. This satisfies

$$s_g^*\varphi - \varphi = dS_g^*\varphi + S_g^* d\varphi; \quad \varphi \in \Lambda_P \tag{12.113}$$

(see de Rham, 1960, p. 75), where we recall that $s_g\colon x \to xg$, with $x \in G$, is a right translation, and the basic property is that $S_g^*\colon \Lambda_P \to \Lambda_P$. This is a consequence of (12.111) and of the fact that the homotopy is polynomial (see §12.8.2). Note also that we could not have done the same thing with the left action τ_g on the group.

Now the left action on G induces an action on $\mathscr{D}'(G)$ that commutes with S_g^*. This implies that $S_g^*\colon \mathscr{E}(G) \to \mathscr{E}(G)$.

Exercise To see this, observe that every ξ in the Lie algebra induces a right-invariant Lie derivative on $\mathscr{D}', \mathscr{E}', \ldots$ (i.e. the derivative with respect to the flow $g \to e^{t\xi} g$; see Warner, 1971, §2.24) that commutes with S_g^*. See also de Rham (1960, Theorem 12(3), p. 80).

Now any smooth measure $\mathrm{d}\mu = f\,\mathrm{d}g$, with $f \in C_0^\infty(G)$ and $\int \mathrm{d}\mu = 1$, can be used as a mollifier (see Schwartz, 1957; Katznelson, 1968) and we can integrate (12.113) to obtain (see de Rham, 1960, §15, p. 75)

$$\begin{aligned} R^*\varphi - \varphi = \mathrm{d}A^*\varphi + A^*\mathrm{d}\varphi; \quad R^*\varphi &= \int s_g^*\varphi\,\mathrm{d}\mu(g), \\ A^*\varphi &= \int S_g^*\varphi\,\mathrm{d}\mu(g),\ \varphi \in \Lambda_P, \end{aligned} \tag{12.114}$$

and by what we just said,

$$A^*\colon \Lambda_P \to \Lambda_P; \quad A^*\colon \mathscr{E} \to \mathscr{E}. \tag{12.115}$$

Moreover, by the basic properties of the mollifier (these were used in §10.3.6), we have $R^*\colon \Lambda_P \to \Lambda_P \cap \mathscr{E}$.

All the properties of the regularising operator R^* that are needed have thus been verified. In the next subsection we shall give an important application of this regularisation and to simplify notation we shall drop the $*$ and denote the corresponding operators by R and A.

12.10.3 Proof of Proposition 12.18

Refer back to the proposition for conditions (i)–(iv).

If $\varphi \in \Lambda_P$ is closed, by (12.114) we see that $[\varphi] = [R\varphi]$ are in the same cohomology class and since $R\varphi \in \Lambda_P \cap \mathscr{E}$ this shows that (ii) $\Longrightarrow$ (i).

Similarly, let $\omega^{(1)} \cdots \omega^{(p)} \in \Lambda_P \cap \mathscr{E}$ be closed forms as in (12.44), and let $\theta \in \Lambda_P$ be some closed form. By (12.114), we have $\theta = R\theta + \mathrm{d}A\theta$, and since condition (iv) says that $R\theta = \mathrm{d}\theta_1 + \lambda_1\omega^{(1)} + \cdots + \lambda_p\omega^{(p)}$, with $\theta_1 \in \mathscr{E}(G) \cap \Lambda_P$, it follows that $\theta = \mathrm{d}\theta_1 + \mathrm{d}A\theta + \lambda_1\omega^{(1)} + \cdots$. In other words, (iv) $\Longrightarrow$ (iii).

Conversely, assume that condition (i) holds. Let $\varphi \in \Lambda_P \cap \mathscr{E}$ be some closed form without constant term. We can then write $\varphi = \mathrm{d}\theta$ for some $\theta \in \Lambda_P$. But by

(12.114) we can also write $\theta = R\theta + \mathrm{d}A\theta + A\varphi$. Therefore $\varphi = \mathrm{d}(R\theta + A\varphi) = \mathrm{d}\theta_1$ where $\theta_1 \in \Lambda_P \cap \mathscr{E}$. In other words, we have shown that (i) $\Longrightarrow$ (ii).

Similarly, assume condition (iii) holds and let $\varphi \in \Lambda_P \cap \mathscr{E}$ be some closed form. We can then write

$$\varphi = \mathrm{d}\theta + \lambda_1 \omega^{(1)} + \cdots + \lambda_p \omega^{(p)}; \quad \lambda_1, \ldots \in \mathbb{R},\ \theta \in \Lambda_P, \tag{12.116}$$

and where $\omega^{(1)}, \ldots, \omega^{(p)}$ is some basis of the cohomology of Λ_P. By replacing, if necessary, $\omega^{(j)}$ by $R\omega^{(j)}$, we can also assume that $\omega^{(1)}, \ldots \in \Lambda_P \cap \mathscr{E}$. Now use (12.114) to get

$$\begin{aligned} \theta &= R\theta + \mathrm{d}A\theta + A\mathrm{d}\theta \\ &= R\theta + \mathrm{d}A\theta + A\big(\varphi - \lambda_1\omega^{(1)} - \cdots - \lambda_p\omega^{(p)}\big) \\ &= \theta_1 + \mathrm{d}A\theta, \end{aligned} \tag{12.117}$$

where $\theta_1 \in \Lambda_P \cap \mathscr{E}$. By substituting θ from (12.117) in (12.116) we obtain

$$\varphi = \mathrm{d}\theta_1 + \lambda_1 \omega^{(1)} + \cdots + \lambda_p \omega^{(p)}$$

as needed in condition (iv). This means that (iii) $\Longrightarrow$ (iv).

Exercise The above is a simple-minded but direct way of proceeding! The reader is invited to ponder the following equivalent, but more sophisticated, formulation: the canonical inclusion $H(\mathscr{E} \cap \Lambda_P) \to H(\Lambda_P)$ is an isomorphism. Compare with de Rham (1960, §18, p. 94).

12.11 Duality Theory for Complexes

12.11.1 Notation and definitions

Let us go back to §12.4 and consider the chain complex of topological vector spaces and continuous linear mappings

$$\Lambda:\ \cdots \longrightarrow \Lambda_{n-1} \xrightarrow{\mathrm{d}} \Lambda_n \xrightarrow{\mathrm{d}} \Lambda_{n+1} \longrightarrow \cdots; \quad \mathrm{d}^2 = 0.$$

We shall also consider the chain complex of dual spaces (i.e. the spaces of continuous linear functionals; see e.g. Bourbaki, 1953; Grothendieck, 1958) and the dual mappings

$$\Lambda^*:\ \cdots \longleftarrow \Lambda^*_{n-1} \xleftarrow{d^*} \Lambda^*_n \xleftarrow{d^*} \Lambda^*_{n+1} \longleftarrow \cdots.$$

One can, as in §12.4.1, consider the homologies of these complexes $H(\Lambda)$ and $H(\Lambda^*)$. The issue is to find out how they relate to each other. This question is usually formulated more simply by setting

$$\dim H_n(\Lambda) = \beta_n, \quad \dim H_n(\Lambda^*) = \beta_n^*; \quad n \in \mathbb{Z}, \tag{12.118}$$

the corresponding Betti numbers possibly $+\infty$ and trying to decide how these are related.

The classical example All the spaces are finite-dimensional (their topology is tacitly taken to be unique, e.g. the Euclidean topology). These are the complexes that arise from a finite geometric simplicial complex in algebraic topology (see e.g. Hilton and Wylie, 1960; Dubrovin et al., 1990). The answer to our question is then very simple and very well known and we have $\beta_n = \beta_n^*$ for $n \in \mathbb{Z}$, (and of course then, as pointed out in §12.4.1, only finitely many are non-zero). This can be treated as an elementary exercise in linear algebra, which in fact will be carried out in detail in a more general setting later on. Or the reader could find this explicitly in the above references. In that case the dual spaces Λ_n^* are finite-dimensional and are also assigned with their natural topology.

Natural topologies can also be given to the duals in all the natural examples of complexes that we have considered, for example $\mathscr{D}$, $\mathscr{D}^* = \mathscr{D}'$, $\mathscr{E}$, $\mathscr{E}^* = \mathscr{E}'$, etc. in §12.4.2. When this can be done, the bidual complex, $\Lambda^{**} = (\Lambda^*)^*$, can be considered and we have the natural identification $\Lambda_n \subset \Lambda_n^{**}$, with $n \in \mathbb{Z}$. In the finite-dimensional case we have $\Lambda_n^{**} = \Lambda_n$ but this is also the case for $(\mathscr{D}')^* = \mathscr{D}$ and $(\mathscr{E}')^* = \mathscr{E}$. We are then in the case of reflexive spaces (see de Rham, 1960, §§9, 10, 17); recall that $*$ and 'prime' are used interchangeably to indicate the dual. For finite-dimensional complexes, the easiest proof that $\beta_n = \beta_n^*$ (as, say, in Dubrovin et al., 1990) uses this reflexivity (see Example 12.43 below).

For general complexes, if we are in a reflexive situation our problem about the relation between β_n and β_n^* is, again, usually easier to answer. Unfortunately, with the exception of the examples that we have just described, the reflexive situation does not occur among the natural complexes that we shall need to consider.

Be that as it may, in the remainder of this chapter we shall examine this problem in a systematic way.

12.11.2 Notation and standard identifications

Here Λ will denote a topological vector space (TVS; see §12.4.1 and footnote) and Λ^* will denote its dual. For $x \in \Lambda$, $x^* \in \Lambda^*$ we denote by $(x^*, x) = (x, x^*) = x^*(x)$ the natural scalar product. For a subspace $B \subset \Lambda$ we write

$B^{\perp} = [x^* \in \Lambda^*;\ (x^*, b) = 0,\ b \in B]$, and clearly $B^{\perp} = (\overline{B})^{\perp}$, for the closure $\overline{B}$ of B in Λ. Let us now consider two subspaces that will be assigned with their induced topologies

$$\Lambda \supset Z \supset B, \qquad \Lambda^* \supset B^{\perp} \supset Z^{\perp}, \tag{12.119}$$

and let us assume that B is closed so that we consider the following TVS and their duals where we make the standard canonical identifications:

$$\begin{gathered}(\Lambda/B) \supset (Z/B), \quad (\Lambda/B)^* \supset (Z/B)^{\perp}, \quad (\Lambda/B)^* = B^{\perp} \subset \Lambda^*; \\ (Z/B)^* = (\Lambda/B)^* \Big/ (Z/B)^{\perp}, \end{gathered} \tag{12.120}$$

where the quotient topology is assigned on the quotient spaces. The subspace $(Z/B)^{\perp} \subset (\Lambda/B)^*$ consists of the elements $\xi \in (\Lambda/B)^*$ that vanish on all $\dot{z} \in Z/B$. This happens if, in the identification (12.120), ξ not only belongs to $B^{\perp}$ but to the smaller space $Z^{\perp}$ and we can therefore identify canonically

$$(Z/B)^{\perp} = Z^{\perp} \subset B^{\perp} = (\Lambda/B)^*, \qquad (Z/B)^* = B^{\perp}/Z^{\perp}. \tag{12.121}$$

Consider now two TVSs, their duals, a continuous linear mapping between them and its dual mapping:

$$\Lambda \xrightarrow{d} \tilde{\Lambda}, \quad \Lambda^* \xleftarrow{d^*} \tilde{\Lambda}^* \colon \left(d^* \tilde{\xi}^*, \xi\right) = \left(\tilde{\xi}^*, d\xi\right); \quad \xi \in \Lambda,\ \tilde{\xi}^* \in \tilde{\Lambda}^*. \tag{12.122}$$

The standard notation $\operatorname{Im} d, \operatorname{Ker} d, \ldots$ for the image space and the kernel of a mapping are used below and we have

$$\operatorname{Im} d^* \subset (\operatorname{Ker} d)^{\perp} \subset \Lambda^*, \qquad \operatorname{Ker} d^* = (\operatorname{Im} d)^{\perp} \subset \tilde{\Lambda}^*. \tag{12.123}$$

Exercise 12.42 Prove this. For the first relation, $(d^* f, z) = (f, dz) = 0$, $f \in \tilde{\Lambda}^*$, $z \in \operatorname{Ker} d$. For the second relation,

$$\begin{aligned} z \in \operatorname{Ker} d^* \Leftrightarrow \left(d^* z, f\right) = 0, \text{ for all } f \in \Lambda \Leftrightarrow (z, df) = 0, \\ \text{for all } f \in \Lambda \Leftrightarrow z \in (\operatorname{Im} d)^{\perp}. \end{aligned}$$

We shall use these facts and identifications to the complex Λ and to its dual in §12.11.1:

$$\cdots \xrightarrow{d_{n-1}} \Lambda_n \xrightarrow{d_n} \Lambda_{n+1} \longrightarrow \cdots, \quad \cdots \xleftarrow{d^*_{n-1}} \Lambda^*_n \xleftarrow{d^*_n} \Lambda^*_{n+1} \longleftarrow \cdots, \tag{12.124}$$

$$\begin{gathered} B = \overline{\operatorname{Im} d_{n-1}} \subset Z = \operatorname{Ker} d_n \subset \Lambda_n, \\ \operatorname{Ker} d^*_{n-1} = (\operatorname{Im} d_{n-1})^{\perp} = B^{\perp} \subset \Lambda^*_n. \end{gathered} \tag{12.125}$$

Without assigning a topology to the dual spaces Λ^*_n we obtain therefore a surjective linear mapping

$$\begin{aligned} H_n(\Lambda^*) &= \operatorname{Ker} d_{n-1}^* / \operatorname{Im} d_n^* \\ &\longrightarrow B^\perp/Z^\perp = (Z/B)^* = \big(\operatorname{Ker} d_n \,/\, \overline{\operatorname{Im} d_{n-1}}\big)^*. \end{aligned} \tag{12.126}$$

Under the assumption that the subspaces $\operatorname{Im} d_{n-1}$ are closed we finally have a surjective linear mapping

$$H_n(\Lambda^*) \longrightarrow \big(H_n(\Lambda)\big)^*; \quad n \in \mathbb{Z}. \tag{12.127}$$

Example 12.43 When all the spaces Λ_n are finite-dimensional then (12.127) holds and in (12.118) we have $\beta_n^* \geqslant \beta_n$. Furthermore, we are then in a reflexive situation and $\Lambda^{**} = \Lambda$ and therefore we also have $\beta_n \geqslant \beta_n^*$. The proof we gave is essentially the standard proof that $\beta_n = \beta_n^*$ in that case (see Dubrovin et al., 1990, §2.9).

Example 12.44 (The classical complexes $\mathscr{D}$, $\mathscr{E}$ of §12.4.2 on an orientable manifold) As we have already pointed out for the natural topologies we are again in a reflexive situation with $(\mathscr{D})^* = \mathscr{D}'$, $(\mathscr{E})^* = \mathscr{E}'$. We shall use the notation $\mathrm{d}_{\mathscr{D}}, \mathrm{d}_{\mathscr{D}'}, \ldots$ for the differentials of these complexes. One important property of these complexes is the fact that (12.123) can be strengthened here and we have

$$\operatorname{Im} d_{\mathscr{D}'} = (\operatorname{Ker} d_{\mathscr{D}})^\perp, \qquad \operatorname{Im} d_{\mathscr{E}'} = (\operatorname{Ker} d_{\mathscr{E}})^\perp. \tag{12.128}$$

This statement is the celebrated Poincaré duality for de Rham cohomology (see de Rham, 1960, §22 and Bott and Tu, 1982, §12.15).

This type of duality *fails* for the complexes Λ_P, Λ_P^*, that are the subject matter of Part III of the book. In the classical situation $\Lambda = \sum_{i\in I} \mathbb{R}_i$, $\Lambda^* = \prod_{i \in I} \mathbb{R}_i$, with $\mathbb{R}_i \cong \mathbb{R}$, and these spaces are assigned with their natural topologies (see Grothendieck, 1958, §IV.1, no. 6, p. 273). Among the other things that one uses for the proof of (12.128) when the manifold is not compact is the fact that every linear functional on Λ is continuous (see de Rham, 1960, §22, p. 207 or §G.4.1 below). This 'strange' property of course does not hold on Λ_P.

Remark (Not used in what follows.) More generally, if Λ, d is a TVS with $\mathrm{d}\colon \Lambda \to \Lambda$ a linear mapping such that $\mathrm{d}^2 = 0$, a differential, and Λ^*, d^* is the dual space with the dual differential, then we have a canonical vector space isomorphism $\operatorname{Ker}\mathrm{d}^* / \overline{\operatorname{Im}\mathrm{d}^*} \cong \big(\operatorname{Ker}\mathrm{d}/\overline{\operatorname{Im}\mathrm{d}}\big)^*$, where the closure in Λ^* is taken for the weak $\sigma(\Lambda^*, \Lambda)$ topology. This follows from (12.126) and Hahn–Banach, which says that $\overline{\operatorname{Im}\mathrm{d}^*} = (\operatorname{Ker}\mathrm{d})^\perp$ (see Grothendieck, 1958).

12.12 The Use of Banach's Theorem

12.12.1 Banach's theorem

The key new ingredient that will be used is Banach's theorem (see e.g. Grothendieck, 1958, Bourbaki, 1953, I, §3). This fundamental fact in analysis says that if $i\colon B_1 \to B_2$ is a one-to-one onto continuous linear mapping from one Banach space to another, then i is actually a Banach space isomorphism; that is, $i^{-1}\colon B_2 \to B_1$ is also continuous and we write $B_1 \cong B_2$. This fact also holds if we assume that B_1, B_2 are Fréchet spaces, that is, spaces that are metrisable and complete (cf. footnote 1 in §12.4.1). An equivalent way of stating this is to say that if $\alpha\colon B_1 \to B_2$ is a continuous surjective mapping between Fréchet spaces, that is, if $\operatorname{Im}\alpha = B_2$, then $B_2 \cong B_1 / \operatorname{Ker}\alpha$ (for the quotient topology).

This fact will now be incorporated into the considerations of our previous section. The notation of §12.11 will be preserved throughout and Λ, $\tilde{\Lambda}$ in (12.122) will be assumed to be Banach spaces or more generally Fréchet spaces. This new condition will be moulded in the identifications of §12.11 and the additional information that we shall obtain will be best expressed in a sequence of diagrams.

Exercise Let $\alpha\colon B_1 \to B_2$ be as above but now we make the weaker assumptions that $\operatorname{Im}\alpha$ is of finite codimension. Then $\operatorname{Im}\alpha$ is a closed subspace. For the proof, by the above observation, α can be assumed one-to-one and if V is the finite-dimensional algebraic complement of $(\operatorname{Im}\alpha)$ in B_2 then $V \oplus B_1$, with the natural topology, goes surjectively on B_2 by $(\text{Identity})_V \oplus \alpha$ on B_2.

12.12.2 The diagrams

The notation and the conditions are as explained above and in §12.11, and in this subsection we always *assume* for the mapping d of (12.122) that $\operatorname{Im} d = B$ is *closed*.

In diagram 1 of Figure 12.3, d_1 is the induced mapping from Λ to $B = \operatorname{Im} \mathrm{d}$ and i is the canonical injection of B in $\tilde{\Lambda}$.

In diagram 2, d_1^* and i^* are the dual mappings. The dual B^* of B is taken with the dual Banach topology and i^* is surjective by Hahn–Banach.

In diagram 3, $Z = \operatorname{Ker} d$, π is the canonical projection to the quotient space with the quotient topology and $\dot{d}_1$, by Banach's theorem, is a TVS *isomorphism* between Banach spaces (or more generally Fréchet spaces).

In diagram 4, $\dot{d}_1^*$ is a vector space isomorphism, that is, one-to-one and onto. But observe that no topologies are a priori given to these spaces; π^* is the dual of π, π^* is one-to-one and it factors as explained in diagram 5.

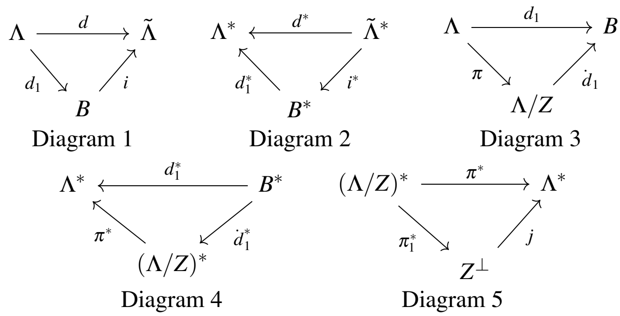


Figure 12.3

In diagram 5, π_1^* is one-to-one and onto, that is, a vector space isomorphism, and j is the one-to-one identification of the subspace $Z^\perp$ in Λ^*, where we use the notation of §12.11.2.

By combining these diagrams we obtain the following composition of mappings:

$$\tilde{\Lambda}^* \xrightarrow[\text{onto}]{i^*} B^* \xrightarrow[\cong]{\dot{d}_1^*} (\Lambda/Z)^* \xrightarrow[\cong]{\pi_1^*} Z^\perp \xrightarrow[\text{one-to-one}]{j} \Lambda^*. \tag{12.129}$$

By the diagrams we have

$$\pi^* = j \circ \pi_1^*, \quad \pi^* \circ \dot{d}_1^* = d_1^*, \quad d_1^* \circ i^* = d^*, \tag{12.130}$$

and (12.130) implies

$$j \circ \pi_1^* \circ \dot{d}_1^* = d_1^*, \quad j \circ \pi_1^* \circ \dot{d}_1^* \circ i^* = d^*, \tag{12.131}$$

that is, (12.129) is a factorisation of d^*. Finally, we conclude the required improvement of (12.123), namely

$$\operatorname{Im} d^* = (\operatorname{Ker} d)^\perp. \tag{12.132}$$

The above will be summed up in the following result.

Lemma 12.45 *Let us assume that the complex of §12.11.1,*

$$\Lambda : \ \cdots \Lambda_{n-1} \xrightarrow{d_{n-1}} \Lambda_n \xrightarrow{d_n} \Lambda_{n+1} \longrightarrow \cdots,$$

is a chain complex of Banach or, more generally, of Fréchet, spaces. Let us further assume that all the subspaces $\operatorname{Im} \mathrm{d}_n \subset \Lambda_{n+1}$ *are closed. We then have*

$$\operatorname{Im} d_n^* = (\operatorname{Ker} d_n)^\perp; \quad n \in \mathbb{Z}. \tag{12.133}$$

This implies that the canonical mapping of (12.126)*,*

$$H_n(\Lambda^*) = \operatorname{Ker} d^*_{n-1}/\operatorname{Im} d^*_n \underset{\cong}{\longrightarrow} (\operatorname{Ker} d_n/\operatorname{Im} d_{n-1})^*$$
$$= \text{topological dual of the Banach space } H_n(\Lambda),$$

is a (*vector space*) *isomorphism*.

Observe that this lemma is an alternative way of proving the identity of Betti numbers $\beta_n = \beta^*_n$ for the complexes with finite-dimensional spaces Λ_n. This proof is less simple but it works for Banach spaces that are not necessarily reflexive provided that these Betti numbers are finite.

This last point will be used explicitly below. We shall therefore state it as a corollary.

Corollary 12.46 *Let us assume that* Λ *in* §*12.11.1 is a chain complex of Fréchet spaces*. *Then*

(i) *if we assume that* Λ *is acyclic then the dual complex* Λ^* *in* §*12.11.1 is acyclic*;
(ii) *if we assume that* Λ *is finite* (*i.e*. *the homologies are finite-dimensional as in* §*12.4.1*) *then* Λ^* *is finite*.

For (ii) we use the exercise in §12.12.1.

Notice finally that (12.133) is an abstract version of (12.128). So in particular, in Example 12.44 when the manifold is compact, $\mathscr{D}(M)$ is a Fréchet space and we have a proof of Poincaré duality! This should not come as a surprise because we started with the condition that $\operatorname{Im} d$ is closed. This condition is thus seen to be very strong (see the exercise in §12.12.1).

12.13 The Use of More Sophisticated Topological Vector Spaces

12.13.1 The scope of this section

The reader who does not feel comfortable with the general theory of topological vector spaces and who is not prepared to dip into the classical references on TVS (e.g. Grothendieck, 1958, Chapter IV) could skip reading this section altogether.

The results from functional analysis that will be essential for the proof of Theorem 12.17 will in fact be given in §12.15 below and there only Banach spaces are used. However, the arguments in §12.15 will appear rather ad hoc unless one is guided by a more global set-up that we shall explain here and in §12.14. What makes things more ambiguous is that the ideas in this section work only up to a point for the proof of Theorem 12.17, and anyway the content of §12.15 is essential.

12.13.2 The natural topologies on Λ_P, Λ_P^*

We saw in §12.7 that the spaces Λ_P, Λ_P^* can be represented as the subspaces of $\mathscr{D}'(U)$ for the manifold U as follows:

$$\Lambda_P = \bigcup_m \Lambda_{m,m}, \quad \Lambda_P^* = \bigcap_m \left(\Lambda_{m,m}^0\right)^*. \tag{12.134}$$

From the definitions it is easy to verify that the single index m suffices (see §12.7).

The natural topologies to assign to these spaces are essentially imposed by the way they are defined (see Bourbaki, 1953; Schwartz, 1957 or de Rham, 1960, §9). On Λ_P we give the inductive limit topology of Banach spaces and on Λ_P^* the projective limit topology of the dual spaces. The standard notation (used in these references) is

$$\Lambda_P = \varinjlim \Lambda_{m,m}, \qquad \Lambda_P^* = \varprojlim \left(\Lambda_{m,m}^0\right)^*. \tag{12.135}$$

The inductive limit topology is the strongest topology that makes all the injections $\Lambda_{m,m} \subsetneq \Lambda_P$ continuous and the projective limit topology is the weakest topology that makes all the 'projections' (which here are in fact one-to-one) $\Lambda_P^* \subsetneq \left(\Lambda_{m,m}^0\right)^*$ continuous.

The topology on Λ_P^* is a complete metrisable topology, that is, a Fréchet topology. This is simply because it is the countable projective limit of Banach spaces (see Bourbaki, 1953). This topology, as is always the case for Fréchet spaces, is given by an increasing sequence of seminorms which here are (see (12.65))

$$p_1 \leqslant p_2 \leqslant \cdots; \quad p_n(T) = F_n(T),\ T \in \mathscr{D}'(M). \tag{12.136}$$

Notice that abstract inductive limits are not in general Hausdorff. Here we have no problem because $\Lambda_P \subsetneq \mathscr{D}'(U)$.

12.13.3 An illustration

The following proposition shows how we can pass from the Fréchet complex Λ_P^* to the original complex Λ_P. This proposition illustrates both the regularisation of §12.10 and the use of the Banach theorem that we apply to Fréchet spaces in §12.12. We shall make explicit use of the dual complex Λ_P^* in the next chapter, but the use of the next proposition will not be essential. The proof of the proposition will be given as a series of exercises and the reader can skip it if they so wish (together with the rest of the section for that matter).

Proposition 12.47 *Let G be some connected Lie group and let Λ_P, Λ_P^* be the corresponding complexes as in §12.5.3. Then*

(i) *if Λ_P^* is acyclic Λ_P is acyclic also*;
(ii) *if the homology of Λ_P^* is finite-dimensional then so is the homology of Λ_P.*

It may well be true that the above implications work the other way round and even for a general manifold. In other words, it is possible that for a general manifold the two complexes Λ_P and Λ_P^* have to be acyclic *simultaneously*. This question presents some independent interest.

The proof illustrates well many of the ideas that we have developed. We shall outline the proof of (i) in the following exercises.

Exercise 12.48 Let us modify the spaces of §12.7 as follows:

$$\mathscr{C}_p^*(L^1) = \left[T \in L^1_{\mathrm{loc}}(G);\ M_p(T) < +\infty\right], \tag{12.137}$$
$$\Lambda_{p,q}^*(L^1) = \left[T = T_1 + bT_2;\ T_1, T_2 \in L^1_{\mathrm{loc}},\ M_p(T_1) + M_q(T_2) < +\infty\right].$$

These are exactly the spaces of (12.51), (12.60) with the additional requirement that the coefficients of the corresponding forms are locally L^1 functions with respect to Lebesgue measure. The space $\Lambda_P^*(L^1) = \bigcap_{p,q} \Lambda_{p,q}^*(L^1)$ is then defined as in (12.63).

The first step in the proof consists in showing that the acyclicity of Λ_P^* implies the acyclicity of $\Lambda_P^*(L^1)$. The proof is done by the regularisation procedure of §12.10. The details, however, are not entirely trivial and to write the whole thing properly takes a bit of doing! One must show in particular that $A\Lambda_p^*(L^1) \subset \Lambda_p^*(L^1)$ where A is as in §12.10 or de Rham (1960, p. 75).

Exercise 12.49 As in (12.135) we assign on $\Lambda_P^*(L^1)$ the projective limit topology. This is a Fréchet topology and its dual is the space $\Lambda_P(G;\mathrm{pol}, L^\infty_{\mathrm{loc}})$ of §12.5.3. This is a routine exercise in TVS.

Exercise 12.50 Now use Corollary 12.46 to show that $\Lambda_P(G;\mathrm{pol}, L^\infty_{\mathrm{loc}})$ is acyclic.

Exercise 12.51 The regularisation of §12.10 finally gives the acyclicity of Λ_P.

Remark 12.52 The same exercise can be carried out with the complex $\overline{\Lambda}_P$ of Remark 12.28 instead.

Remark 12.53 All these are interesting exercises but do not have great incidence in the proof of our theorem. This could of course change if one could produce a proof of the following:

$$\text{acyclicity } \Lambda_P \Longleftrightarrow \text{acyclicity of } \Lambda_P^* \Longleftrightarrow \text{acyclicity of } \overline{\Lambda}_P \qquad (12.138)$$

for all manifolds, or at least find the natural condition on a manifold for (12.138) to hold.

12.13.4 An unsuccessful but instructive attempt to prove Proposition 12.47 the 'other way round'

Let us examine closely the inductive limit topology that is given in (12.135) for the space Λ_P. This topology is not Fréchet but by its definition is what in the specialised literature is call a limit Fréchet or $\mathscr{L}\mathscr{F}$-topology. A key property that these spaces admit is that Banach's theorem as explained in §12.12.1 holds for these spaces (see Grothendieck, 1958, §IV.1, no. 5, p. 271).

Things therefore look promising and we seem to be in 'good shape' to reverse all the arguments in the proof of Proposition 12.47 and prove the implications the other way round. In carrying this program out we are stopped by a rather unexpected and intriguing obstacle. Banach's theorem has to be applied in §12.12 not only to the space Λ but also to the closed subspace B, and the subspaces of $\mathscr{L}\mathscr{F}$-spaces are not in general $\mathscr{L}\mathscr{F}$ (see Jarchow, 1981, pp. 270, 281; Grothendieck, 1958, p. 263). Nor is there any guarantee that Banach's theorem holds for these subspaces. The relevant references for the reader who wishes to pursue this matter further are as above.

The question of whether acyclicity of Λ_P implies acyclicity of Λ_P^* for a general manifold remains therefore open and seen in this light it is a question that presents some independent interest. An effort to tackle this general problem is made in §12A.2. For the experts or enthusiasts in TVS note that, related to the above, there exists a famous example due to G. Köthe which shows that subspaces of bornological spaces need not be bornological (see Bourbaki, 1953, §IV.5, Exercise 21 or Grothendieck, 1958, §IV.4, Exercise 4).

12.13.5 An exercise in topological vector spaces

We describe here one of the tricks we have to use in the proof of the $\mathscr{L}\mathscr{F}$-version of Banach's theorem. Explicit use of this will be made in §12.15 later.

Exercise Let Λ_r, with $r \geqslant 1$, and B be Banach spaces, let $\delta_r \colon \Lambda_r \to B$ be continuous linear mappings and let $U = \bigcup(\delta_r \Lambda_r) \subset B$. Then if $U = B$ there exists some r_0 such that $\delta_{r_0}\Lambda_{r_0} = B$. To see this let $D_r \subset \Lambda_r$ denote the unit ball of the Banach space. By our hypothesis we have $\bigcup_{r,n} \overline{\delta_r(nD_r)} = B$. The Baire category theorem implies therefore that there exists r such that $\overline{\delta_r(D_r)}$ is a neighbourhood of 0 in B. This implies $\delta_r\Lambda_r = B$ by one of the standard lemmas that one proves on the way to Banach's theorem (see e.g. Grothendieck, 1958, §1.14.2, p. 69 or Bourbaki, 1953 or even any other text on elementary functional analysis).

The above argument can give a slightly stronger conclusion. Let the notation be as above but make the weaker assumption that $U \subset B$ has finite, or even more generally, countable, codimension. Then the conclusion is that there exists some r such that $\delta_r U_r$ is closed and of finite codimension in B.

Only the finite codimension will be needed (in §12.16 below) and the proof is as in §12.12.1: for V a finite-dimensional subspace such that $B = U \oplus V$ (the algebraic sum) we consider the corresponding mappings $\Lambda_r \oplus V \to B$; then we are in the previous case.

The countable case is almost identical: we let $V = \bigcup V_j$ be the union of finite-dimensional subspaces. We then consider the corresponding mappings $\Lambda_r \oplus V_j \to B$ and argue as before.

12.14 The Acyclicity of Λ_P^* and $\overline{\Lambda}_P$ of §12.7.2

This section only uses Fréchet spaces and as a result it should be less forbidding than the previous one to non-experts in the theory of TVS. Nonetheless, like the previous subsection it is not essential for the proof of the main theorem and the reader who so wishes could skip it. The fact remains, however, that the methods and the patents that are developed here are very instructive for understanding the proofs that will be given in the crucial §12.15.

12.14.1 Fréchet spaces and their quotients

It follows easily from the definition that the topology on any metrisable (locally convex) TVS E is defined by an increasing sequence of seminorms (see e.g. Bourbaki, 1953). Let

$$p_1 \leqslant p_2 \leqslant \cdots, \quad q_1 \leqslant q_2 \leqslant \cdots \tag{12.139}$$

be two such sequences that define on E the same topology. In concrete terms, this happens if and only if the two sequences of seminorms satisfy the following property.

There exists $\mathbb{N} \ni i \to \theta(i) \in \mathbb{N}$ *a mapping from* $\mathbb{N} = (1,2,\dots)$ *into itself and constants* $\lambda_i > 0$, $i \geqslant 1$ *such that*

$$p_i \leqslant \lambda_i q_{\theta(i)}, \quad q_i \leqslant \lambda_i p_{\theta(i)}; \quad i \geqslant 1. \tag{12.140}$$

Exercise 12.54 Prove this. Think of normed spaces. Use the definitions (see Grothendieck, 1958, p. 38).

When $H \subset E$ is a closed subspace then the quotient topology on E/H is given by the seminorms $\dot{p}_1 \leqslant \dot{p}_2 \leqslant \cdots$ that are defined by

$$\begin{gathered} \dot{p}_j(\dot{x}) = \inf[p_j(x);\, x \in E,\, \pi(x) = \dot{x}]; \quad \dot{x} \in E/H, \\ \pi\colon E \longrightarrow E/H \text{ is the canonical projection.} \end{gathered} \tag{12.141}$$

This is also an easy exercise on the definitions.

If we combine the above two observations and use Banach's theorem we conclude the following.

Let $f\colon E \to \widetilde{E}$ *be some continuous linear mapping from the Fréchet space E* onto *the Fréchet space* $\widetilde{E}$. *Further, let*

$$p_1 \leqslant p_2 \leqslant \cdots \quad \textit{and} \quad \tilde{p}_1 \leqslant \tilde{p}_2 \leqslant \cdots \tag{12.142}$$

be seminorms that give the corresponding topologies on E and $\widetilde{E}$ *respectively. Then there exist a mapping* θ *and constants as in* (12.140) *such that for all* $i = 1,2,\dots$ *and all* $\tilde{x} \in \widetilde{E}$ *there exists* $x \in E$ *such that*

$$\pi(x) = \tilde{x}; \quad p_i(x) \leqslant \lambda_i \tilde{p}_{\theta(i)}(\tilde{x}). \tag{12.143}$$

Notice that the x depends on $\tilde{x}$ and on i (see Bourbaki, 1953, I, §3).

12.14.2 The acyclicity of Λ_P^*

The above observations will be applied to the Fréchet complex Λ_P^* of §12.7, which will be assumed to be acyclic, and to the mapping $b\colon \Lambda_P^* \to \Lambda_P^*$ given by the boundary operator. We shall apply (12.143) to

$$E = \Lambda_P^*; \quad \widetilde{E} = [S \in \Lambda_P^*;\, bS = 0,\, S = S_1 + S_2 + \cdots,\, \dim S_j = j] \tag{12.144}$$

and to the mapping $b\colon E \to E$, which by our hypothesis satisfies $\widetilde{E} \subset bE$. The definition of $\widetilde{E}$ says that the component S_0 of zero dimension is 0. The conclusion from (12.143) that is needed is stated in the following result.

Proposition 12.55 *Assume that the complex* Λ_P^* *is acyclic. Then there exists a mapping* $\mathbb{N} \ni i \to \theta(i) \in \mathbb{N}$ *and a sequence of positive constants* $\lambda_1, \lambda_2, \dots$

such that for all $S \in \Lambda_P^*$ that is closed and has zero-dimensional component $S_0 = 0$, and all $q \in \mathbb{N}$, there exists $T \in \Lambda_P^*$ such that

$$bT = S; \quad M_q(T) \leqslant \lambda_q M_{\theta(q)}(S). \tag{12.145}$$

To see how this follows from the general fact (12.143) observe the following. For all $S \in \Lambda_P^*$ we have $F_p(S) \leqslant M_p(S)$ $(p \geqslant 0)$ for the flat seminorms of Λ_P^* in (12.65). And also observe that when $S = bT_1$ for some $T_1 \in \Lambda_P^*$ with $F_q(T_1) \leqslant 1$, then $T_1 = T + bT'$ as in (12.60) with $T, T' \in \mathscr{D}'(U)$, $M_q(T)$, $M_q(T') \leqslant 2$ and that we also have $S = bT$.

Remark 12.56 There is a drawback in (12.145). We could start from some closed current $S \in \mathscr{C}^*(U,\text{pol})$ and construct $T \in \mathscr{D}'$ that satisfies (12.145). It would be nice to be able to assert that the T that satisfies (12.145) is $T \in \mathscr{C}^*(U,\text{pol})$. This stronger assertion is easily seen to follow from the acyclicity of the smaller complex $\overline{\Lambda}_P = \mathscr{C}_P^*(U;\text{pol}) + b\mathscr{C}_P^*(U,\text{pol})$ of Remark 12.28. This complex with its natural topology as in (12.36) is also Fréchet. To prove, under the acyclicity of $\overline{\Lambda}_P$, the stronger assertion, observe that now if $S \in \overline{\Lambda}_P$ is closed and as in the proposition, we can write $S = bT_1$ with $T_1 \in \overline{\Lambda}_P$, and with $\overline{M}_q(T_1) \leqslant \lambda_q \overline{M}_{\theta(q)}(S)$ where $q \to \theta(q)$ and the λ_q are as before in (12.143) and now $\overline{M}_q$ are the seminorms that define the topology on $\overline{\Lambda}_P$. The difference with the previous case is that now

$$\overline{M}_q(T) = \inf[M_q(T_1) + M_q(T_2);\ T = T_1 + bT_2,\ T_1, T_2 \in \mathscr{C}^*(\text{pol})] \tag{12.146}$$

and not just $T_i \in \mathscr{D}'$. The argument finishes as before.

Observe also that $\overline{\Lambda}_P$ is the dual of the same space Λ_P assigned with a topology that is a priori different from the inductive limit topology of (12.135). This topology is given on $\Lambda_P = \mathscr{C}(\text{pol}) \cap d^{-1}\mathscr{C}(\text{pol})$ by the procedure explained in (12.35), (12.36). One 'superficial' advantage that $\overline{\Lambda}_P$ has over Λ_P^* is indeed the above remark that makes the Poincaré equation easier to state. (Whether these two topologies of Λ_P are the same is of course an interesting issue, and one that is clearly related to the abstract problem on $\mathscr{L}\mathscr{F}$-spaces that we alluded to in §12.13.4: it is easy to see that one (which?) is stronger than the other.)

12.15 The Acyclicity of Λ_P

12.15.1 Comments on the proof

We finally come to the explicit description of the acyclicity (see §12.4.1 for the definition) of the complex Λ_P of §12.7 and to the proof of the proposition that will be essential for our main theorem.

This proposition is directly inspired from Proposition 12.55 on Fréchet spaces. But it is more subtle to prove because the fundamental use of the Baire category argument, which is the main tool for proving Banach's theorem, has to be built into the proof. Herein lies the additional difficulty. On the upside, only Banach spaces are used in the proof and none of the more sophisticated TVS notions and results of §§12.13 or 12.14 will be needed. As a result the reader who has skipped the last two sections could pick up the argument here. But readers that feel really comfortable with TVS should start with the appendix at the end of this chapter where an abstract result involving only functional analysis is explained and which in turn contains essentially all that will be done in this section.

The only thing that will be needed in this proof will be the definitions and the notation of §§12.7 and 12.12. This notation will be reorganised in (12.147)–(12.150) below.

12.15.2 The diagram and the use of Baire category

We shall assume that the complex $\Lambda_P(U)$ is acyclic and we shall fix some $p \geqslant 0$. We shall go back to diagram 1 of Figure 12.3 and specialise the spaces and the mappings as follows (see §12.7):

$$\begin{array}{ccc} \Lambda & \xrightarrow{\ \mathrm{d}\ } & \tilde{\Lambda} \\ & \searrow \quad \nearrow & \\ & B; & \end{array}$$

$$\tilde{\Lambda} = \tilde{\Lambda}_p = \Lambda^0_{p,0} = [\omega \in \mathscr{C}^0_p, \mathrm{d}\omega \in \mathscr{C}^0_0], \tag{12.147}$$
$$B = B_p = \left[\omega \in \mathscr{C}^0_p;\ \mathrm{d}\omega = 0,\ \omega \text{ homogeneous of degree } \nu + 1\right] \subset \tilde{\Lambda}_p.$$

The space Λ will be chosen as

$$\Lambda = \Lambda^{(\nu)}_{q,p} = \left[\omega \in \mathscr{C}^0_q;\ \mathrm{d}\omega \in \mathscr{C}^0_p,\ \omega \text{ homogeneous of degree } \nu\right], \tag{12.148}$$

where the index q will be chosen presently and $\nu \geqslant 0$, the degree of the homogeneous forms, will be fixed throughout. To simplify notation, in what follows we shall also drop the exponent and write $\Lambda^{(\nu)}_{q,p} = \Lambda_{q,p}$. Going back to §12.7.2, note that this is the component of degree ν in (12.55) – not to be confused with (12.57). For these spaces of currents we have

$$\begin{array}{ccccc} \bigcup_r \Lambda_{r,q}^{(\nu-1)} & \xrightarrow{\ \delta\ } & \Lambda_{q,p} & \xrightarrow{\quad \mathrm{d} \quad} & \tilde{\Lambda}_p \\ & & \searrow & & \nearrow \\ & & & B_p; & \end{array} \qquad \begin{array}{l} \delta \text{ is induced by d}, \\ \nu \geqslant 1, \end{array} \tag{12.149}$$

$$\Lambda_{q,p}^* \xleftarrow{\quad \mathrm{d}^* \quad} \left(\tilde{\Lambda}_p\right)^*, \tag{12.150}$$

where the union on the left-hand side is the union of the subspaces of $\Lambda_{r,q}^{(\nu-1)} \subset \mathscr{D}'$ and d is the differential of the complex Λ_P and it gives a continuous mapping between the corresponding Banach spaces in (12.149). In (12.150) we consider the dual spaces and the dual mapping.

Since $\mathrm{d}^2 = 0$ and by the acyclicity of Λ_P we clearly have

$$\mathrm{d}\left(\Lambda_{q,p}\right) \subset B_p, \quad \mathrm{d}\left(\bigcup_q \Lambda_{q,p}\right) = B_p. \tag{12.151}$$

Now comes the pivotal use of Baire category (see §12.13.5). This ensures that there exists some q_0 such that

$$\mathrm{d}\left(\Lambda_{q,p}\right) = B_p; \quad q \geqslant q_0 \quad \left(\text{recall } \Lambda_{q_1,p} \supset \Lambda_{q,p},\ q_1 \geqslant q\right). \tag{12.152}$$

In what follows, some index q as in (12.152) will be fixed and then we are in the situation where the argument of §12.12.2 and (12.132) holds. Explicitly, in (12.149), (12.150) we have $(\mathrm{Ker}\,\mathrm{d})^{\perp} = \mathrm{Im}\,\mathrm{d}^*$.

We shall denote

$$H_s = [S \in \mathscr{D}';\ M_s(S) < +\infty,\ bS = 0,\ S \text{ homogeneous of dimension } \nu];\ s > q.$$

The scalar product $S[\omega]$, with $S \in H_s$, $\omega \in \Lambda_{q,p}$, then identifies H_s with a subspace $\Lambda_{q,p}^*$ and in that identification we have

$$\|S\|_{\Lambda_{q,p}^*} \leqslant C M_s(S); \quad S \in H_s. \tag{12.153}$$

Let us now return to the first part of the diagram and use the acyclicity and the Baire category argument of §12.13.5 again. We conclude that there exists some index r such that $d\Lambda_{r,q}^{(\nu-1)} = \mathrm{Ker}\,\mathrm{d}$. With this we shall now fix some new index $s > r+q+p$. The claim is that $H_s \subset (\mathrm{Ker}\,\mathrm{d})^{\perp} \subset \Lambda_{q,p}^*$ for the duality (12.149), (12.150).

To see this, let $f \in \mathrm{Ker}\,\mathrm{d}$ and $\phi \in \Lambda_{r,q}^{(\nu-1)}$ such that $\mathrm{d}\phi = f$. Because of Exercises 12.26 and 12.27, there then exists $\mathscr{D} \ni \phi_n \to \phi$ as $n \to \infty$ for the topology of $\Lambda_{s,s}$. It follows that

$$\langle S, f \rangle = \lim \langle S, \mathrm{d}\phi_n \rangle = 0; \quad S \in H_s, \tag{12.154}$$

and because $f \in \operatorname{Ker} d$ is arbitrary, our assertion follows. And the bottom line is that $H_s \subset \operatorname{Im} d^*$ because of (12.132).

Now since $\operatorname{Im} d^*$ is a closed subspace we can use Banach's theorem to conclude that for all $S \in \operatorname{Im} d^*$ there exists $T_0 \in (\tilde{\Lambda}_p)^*$ such that

$$d^* T_0 = S, \quad \|T_0\|_{(\tilde{\Lambda}_p)^*} \leqslant C \|S\|_{\Lambda^*_{q,p}}, \tag{12.155}$$

where the constant is independent of S.

For the definition of the norm in $(\tilde{\Lambda}_p)^*$ we go back to §12.7 and taking into account (12.153) we can reinterpret (12.155) as follows. For all $S \in H_s$ there exists $T_0 = T + bT'$ such that (12.155) holds and for which $M_p(T), M_0(T') \leqslant CM_s(S)$. But then also $S = bT$.

This argument works for all degrees $\nu = 1, 2, \ldots$. We therefore have proved the following proposition.

Proposition 12.57 *Let us assume that $\Lambda_P(U)$ is acyclic and let $p \geqslant 0$. Then there exists $q_0 = q_0(p) \geqslant p$ and for all $q \geqslant q_0$ there exists $C = C_{q,p} > 0$ such that for all*

$$S \in \mathscr{D}', \quad M_q(S) < +\infty, \quad bS = 0, \quad S_0 = 0 \tag{12.156}$$

we can solve with bounds the equation

$$T \in \mathscr{D}', \quad bT = S, \quad M_p(T) \leqslant CM_q(S). \tag{12.157}$$

In (12.156) S_0 denotes the component of dimension 0 of the current S. Note also that the q_0 here is not the same as in (12.152) and that, if we use the notation of the proof, we can take $q_0(p) \approx r + q + p$. As a matter of fact, the notation would have been more consistent if in the statement of the proposition I had used the letter s of the proof rather than q. For later purposes, however, it is better to leave things as they are.

12.15.3 Comment on Propositions 12.57 and 12.55

Proposition 12.57 is of an algebraic nature, but it is basic for the proofs in Chapter 13 and a number of comments are in order.

12.15.3.1 Comparison of Propositions 12.57 and 12.55 These propositions and the variant in Remark 12.56 about the acyclicity of $\overline{\Lambda}_P$ read almost identically. The only differences, and of course this changes everything, are these (in what follows we shall simplify notation and write $\mathscr{C}_r^*$ for the dual space $(\mathscr{C}_r^0)^*$ of (12.51)):

(1) In Proposition 12.57, S is given and lies in $\mathscr{C}_q^*$ and T is constructed in $\mathscr{C}_p^*$.

(2) In the $\overline{\Lambda}_P$ variant of Proposition 12.55, S lies in $\mathscr{C}^*(\text{pol})$ and T is constructed in $\mathscr{C}^*(\text{pol})$.
(3) In the original Proposition 12.55 on Λ_P^*, S was, say, in $\mathscr{C}^*(\text{pol})$ but the T that we construct a priori only lies in Λ_P^*.

It is (1) that is essential for our theorem but the proofs of (2) and (3) are much easier because the hypothesis there is stronger (see Proposition 12.47). Also, this circle of arguments does not prove (12.138) because the current T in (12.157) is not shown to belong to Λ_P^*.

12.15.3.2 The choice procedure in Proposition 12.57 There is something *asymmetric* about the way we have to make our choices (in informal language one could qualify the choice procedure as 'lopsided'). Indeed,

$$\begin{array}{c} \text{given } p \text{ we choose } \longrightarrow q = q_0(p) \gg p \\ \swarrow \\ \text{but then, given } S \in \mathscr{C}_q^* \text{ we choose } T \in \mathscr{C}_p^*. \end{array} \tag{12.158}$$

All in all, the situation is quite 'subtle' and one can easily go wrong. In the appendix to this chapter we shall return to this strange choice procedure and examine what it means from the point of view of general TVSs.

In the next chapter the way of choosing the index q from the index p in (12.158) will be used systematically and that choice will be iterated. We shall give the details when we come to it (see §13.6.1) but *grosso modo* what we do is the following.

We shall construct inductively $s+2$ positive indices $p_{s+1}, p_s, \dots, p_1, p_0$ (and should think $10^{10} \leqslant p_{s+1} \ll p_s \ll \cdots$). This construction is done *backwards* and assuming that p_{j+1} has been chosen then p_j is arbitrary but is larger than $q_0(p_{j+1})$ for the $q_0(\cdot)$ of (12.158) and of Proposition 12.57. We start with a more or less arbitrary but large p_{s+1} and the length $s+2$ will be determined by the geometry of the group.

Once p_0 has been reached we stop and fix $S^0 \in \mathscr{C}_{p_0}^*$. Since p_0 could be very large and certainly 'out of control' this S^0 will have to be compactly supported.

From this we shall switch and move *forward* and construct *successively* $S^0, S^1, \dots,$ such that $S^i \in \mathscr{C}_{p_i}^*$, until we reach S^{s+1}. With this, we close the loop, so to speak.

This successive construction of the currents $S^0, \dots$ uses Proposition 12.57 and the choice of the indices. How this is done is a long story that will be explained in Chapter 13. The above schematic idea of 'closing the loop' will, on the other hand, be good to keep in mind although at this point it need not be taken too seriously.

12.16 The Case Where the Homology of $\Lambda_P(U)$ Is Finite-Dimensional

Let us now assume that, say, the complex $\Lambda_P(U)$ has finite-dimensional homology rather than being acyclic. Then we have the following version of Proposition 12.57.

Proposition 12.58 *Assume that the homology of complex $\Lambda_P(U)$ is finite-dimensional and that* $\dim H(\Lambda_P) = h < m < +\infty$ *for some integer m. Then for all $p \geqslant 0$ there exist $q_0 = q_0(p) \geqslant p$ such that for every $q \geqslant q_0$ there exists $C = C_{p,q} > 0$ with the following property.*

Let $S_1, \ldots, S_m \in \mathscr{D}'(U)$ be such that

$$\|\{S\}\| = \sum_{j=1}^{m} M_q(S_j) < +\infty; \quad bS_j = 0,\ 1 \leqslant j \leqslant m. \tag{12.159}$$

Then there exist scalars $\lambda_i, \ldots, \lambda_m \in \mathbb{R}$ with $\sup_j |\lambda_j| = 1$ and $T \in \mathscr{D}'(U)$ that solves with bounds the equation

$$bT = \sum \lambda_j S_j; \quad M_p(T) \leqslant C\|\{S\}\|. \tag{12.160}$$

Notice that here we do not have to worry about the 0-dimensional component of the currents. The proof is a very easy adaptation of Proposition 12.57.

Exercise As an exercise, the reader should fill in the details of the outline proof below. To simplify and use the notations of §12.15.3, we shall assume that $S_1, \ldots$ are homogeneous of dimension $\nu \geqslant 1$. This will suffice for the applications in §13.7. The general case is similar but new notations are needed.

First, by §12.13.5 as in §12.15.2, with p given, we shall fix q so that in the diagram (12.149) $\mathrm{Im}\,\mathrm{d} = B$ is closed and therefore $\mathrm{Im}\,\mathrm{d}^* = (\mathrm{Ker}\, d)^\perp$.

Then, as in §§12.15.2, 12.13.5, we can find an r large enough so that $\delta\left(\Lambda_{r,q}^{(\nu-1)}\right) = \mathrm{Im}\,\delta_r$ (here δ_r is the restriction of d on $\Lambda_{r,q}^{(\nu-1)}$) is a closed subspace of $\mathrm{Ker}\,\mathrm{d}$ and of codimension $\leqslant h = \dim H(\Lambda_P)$, the dimension of the homology. If $s > 0$ is large enough, by the same argument as in §12.15.2 this implies that if $S \in \mathscr{D}'$ is such that $bS = 0$ and $M_s(S) < +\infty$ then $\langle S, f\rangle = 0$ for all $f \in \mathrm{Im}\,\delta_r$ and this says that $S \in (\mathrm{Im}\,\delta_r)^\perp$. But $\mathrm{Im}\,\mathrm{d}^* = (\mathrm{Ker}\,\mathrm{d})^\perp \subset (\mathrm{Im}\,\delta_r)^\perp$ is a subspace of codimension $\leqslant h$. From this, the existence of scalars such that $\sup_j |\lambda_j| = 1$ and $\sum \lambda_j S_j \in \mathrm{Im}\,\mathrm{d}^*$ follows as long as $m > h$. From this we argue as in (12.155) and the few lines that follow it, to conclude.

A refinement of this will be needed in §13.7 (see Exercise 13.7): there, instead of one family of currents, several such families $S_{\alpha,1},\dots,S_{\alpha,m}$, with $1 \leqslant \alpha \leqslant \alpha_0$, will be given and then, provided that m is large enough, the scalars $\lambda_1,\dots,\lambda_m$ can be chosen to work simultaneously for all these families.

Another refinement of the proposition consists in weakening the hypothesis. We assume instead that the dimension of $H(\Lambda_P)$ is *countable*. Then we can use the refinements of §12.13.5 to obtain the same conclusion. The reader can rewrite the proof under this weaker hypothesis if so wished. The only use that one makes of this refinement is that with it we can prove that when G is a B-group then not only is $H(\Lambda_P)$ infinite-dimensional but even uncountably so! This fact, however, is neither important nor surprising and we shall not return to it.

The analogue of this proposition holds under the assumptions that one or the other of the complexes $\Lambda_P^*, \overline{\Lambda}_P$ that we considered in §12.14 have finite homology. The only differences are the ones that we pointed out in §12.15.3 and also, of course, that now the exploitation of the key relation (12.132), $\operatorname{Im} d^* = (\operatorname{Ker} d)^{\perp}$, in §12.12 is not needed, and everything is much easier.

12.17 The Partial Acyclicity of the Complexes

Let $\nu = 1,2,\dots$. We then say that the complex Λ_P is acyclic at the level ν if $H_\nu(\Lambda_P) = 0$. A similar definition can be given for the general complex Λ of §12.4.1 and in particular for the complexes Λ_P^* and $\overline{\Lambda}_P$ of §12.7.

Similarly, we say that a complex Λ is finite at level ν if $\dim H_\nu(\Lambda) < +\infty$. Observe that for the geometric complexes $\Lambda_P, \Lambda_P^*, \dots$ we have $H_\nu = 0$ for every ν that is not in the range $0 \leqslant \nu \leqslant \dim U$.

The statements of Propositions 12.55 and 12.57 that we gave admit a variant in terms of this more refined definition. The proofs are verbatim identical and one only has to keep track in the formulas in the proofs of the additional index ν for which the homology vanishes. The only reason why these propositions were stated in terms of global acyclicity was because we did not wish to overload their statements, at least not in a first reading, with yet another index ν.

However, here is this more refined version of Proposition 12.57 and the thing to observe is that here it is preferable to consider homogeneous currents in order to see exactly how the index ν comes into play.

Proposition 12.59 *Let us assume that $\Lambda_P(U)$ is acyclic at levels ν and $\nu+1$, for some $1 \leqslant \nu \leqslant \dim U$, and let $p \geqslant 1$. Then there exists $q_0 \geqslant p$ and for all $q \geqslant$*

q_0 there exists $C > 0$ such that for all homogeneous currents of dimension ν,

$$S \in \mathscr{D}', \quad \dim S = \nu, \quad M_q(S) < +\infty, \quad bS = 0, \tag{12.161}$$

we can solve with bounds by a homogeneous current T the equation

$$T \in \mathscr{D}', \quad bT = S, \quad \dim T = \nu + 1, \quad M_p(T) \leqslant C M_q(S). \tag{12.162}$$

Remark 12.60 The reason why q_0 and C in the proposition are not shown to depend on ν is that ν only takes a finite number of values that are not trivial: $\nu = 1, \ldots, \dim U$.

The proof of this will be left as an exercise for the reader. It will involve going through the proof that we gave in §12.15.2 and verifying that the acyclicity of Λ_P was used twice at two different levels: first $\nu + 1$, then ν. Similarly, for the analogue of Proposition 12.58, if we know that $\dim H_i(\Lambda_P) < \infty$, with $i = \nu, \nu + 1$, and if m is large enough in that proposition, then the conclusion (12.159), (12.160) can be drawn provided that the currents S_j are ν-dimensional.

12A Appendix: Acyclicity in Topological Vector Spaces

12A.1 The position of the problem

Much of the second part of this chapter was about the following problem. Let E be some topological vector space assigned with a differential, that is, some continuous linear operator $\mathrm{d}\colon E \to E$ such that $\mathrm{d}^2 = 0$. The dual space E^* is then assigned with the dual differential d^* but no topology is assigned to E^*. The corresponding homology $H(E) = \operatorname{Ker} d / \operatorname{Im} \mathrm{d}$ (and similarly for $H(E^*)$) is defined. The issue was thus: assume that $\dim H(E) < +\infty$; can we then assert the same thing for $H(E^*)$? In this appendix we shall treat this problem as a general one on topological vector spaces. We shall simplify the problem here and only consider $H(E) = 0$ because this case already contains all the ideas that are used in the proof of the general situation. Notice that the situation here is more general than what we considered in §§12.14–12.17 insofar that we do not demand the *grading* of a simplicial complex.

We saw in §12.12 that our problem admits a positive answer when E is a Banach (Fréchet) space and the case of $\mathscr{L}\mathscr{F}$ spaces was alluded to in §12.13. Let us finally make it clear that this appendix is meant for readers that are familiar with (and like!) the theory of topological vector spaces.

12A.2 A class of topological vector spaces and ersatz acyclicity

We shall consider a special class of topological vector spaces that is natural and contains all the examples considered in this chapter. Let E_n be a sequence of Banach spaces and let E_n^* denote the dual spaces and let $\| \, \|_n$, $\| \, \|_n^*$ denote the norms and the dual norms. We shall assume that there exists a sequence of dense injections $i\colon E_n \to E_{n+1}$ (i.e. linear, continuous one-to-one, and $i(E_n)$ dense in E_{n+1}); the dual mappings $\pi\colon E_n^* \leftarrow E_{n+1}^*$ are then also one-to-one. The example of $E_n = \Lambda^0_{n,n}(U)$ for a manifold U (see §12.7) is what we have considered in this chapter, and then $\mathscr{D}(U)$ is dense in all these spaces. The class of topological vector spaces that we shall consider are the corresponding inductive limits $E = \lim_{\to} E_n$ under the assumption that E is Hausdorff (note that the Hausdorff property is not something that comes for free with inductive limits: cf. the final remark in §12.13.2). The dual E_n^* is then the projective limit of the E_n^* (see in particular Grothendieck, 1958, §IV.1).

The situation is summarised as follows:

$$E_0 \subset \cdots E_n \subset E_{n+1} \subset \cdots \subset E = \varinjlim E_n, \tag{12A.1}$$

$$E_0^* \leftarrow \cdots E_n^* \leftarrow E_{n+1}^* \leftarrow \cdots \leftarrow E^* = \varprojlim E_n^*, \tag{12A.2}$$

where the E_j are identified with subspaces of E, and by our hypothesis all the arrows in the above definition of the projective limit are one-to-one continuous linear mappings. We can then assert that the differential d that is defined on E has the following property. For all n there exist m and $\mathrm{d}_{n,m}\colon E_n \to E_m$ some continuous linear mapping such that the restriction of d on E_n can be factored as $i_m \circ \mathrm{d}_{n,m}$ for i_m, the natural inclusion of E_m in E. (This holds for any linear mapping $\mathrm{d}\colon E_n \to E$; see Grothendieck, 1958, §IV.1, no. 5.)

The dual mapping $\mathrm{d}^* = b$ can certainly be defined on E^*. By analogy with §12.7 we shall call this the boundary mapping: it can also be defined on the E_n^* as follows.

Let n_0 be such that d maps E_0 continuously in E_{n_0} as above. Then for each $n \geqslant n_0$ the dual mapping $b\colon E_0^* \leftarrow E_n^*$ can be defined. More generally, when d maps E_p continuously in E_q the boundary mapping $b\colon E_p^* \leftarrow E_q^*$ can be defined. Concretely, if $n \geqslant n_0$ and $S \in E_n^*$ then $bS \in E_0^*$ and we say that S is *closed* if $bS = 0$.

In the above general set-up, it may or may not be true that $H(E^*, b) = 0$ follows; but we know of no counterexample. Nonetheless, we can prove the following ersatz version of this fact which implies all the results of §§12.15–12.17.

Theorem (The ersatz acyclicity of E^*) *Let $p \geqslant n_0$ be given*; *then there exists*

$q_0 = q_0(p) \geqslant n_0$ *such that for all* $q \geqslant q_0$ *there exists* $C = C_{p,q} > 0$ *with the following property. For all closed* $S \in E_q^*$ *there exists* $T \in E_p^*$ *such that*

$$bT = S; \quad \|T\|_p^* \leqslant C\|S\|_q^*.$$

In concrete terms, both bT and S can be identified with elements of E_0^* and our condition says that they are identical. This can be expressed, without the definition of the boundary operator b, by saying that $q \geqslant n_0$ and

$$\langle T, d\phi \rangle = \langle S, \phi \rangle; \quad \phi \in E_0.$$

One recognises here an abstract version of Proposition 12.57 on integration currents of superpolynomial decay. Notice also that what is subtle in the above statement is this: if $S \in E^*$ and $p \geqslant 1$ we can for sure solve $bT = S$ with $T \in E_p^*$ but not necessarily in E^* as the acyclicity of E^* would have implied.

12A.3 The proof

We consider $B = \operatorname{Im} \mathrm{d} \subset E$ which is closed because of the acyclicity, and identify E_n with a subspace of E. Let p be as given in the theorem and let $B_p = E_p \cap B$ a closed subspace of E_p.

Let $H = E_q \cap \mathrm{d}^{-1} E_p$ for some q. Clearly, $\mathrm{d}H \subset B_p$ and if we assign H with the norm $\|x\|_q + \|\mathrm{d}x\|_p$, then H is a Banach space. (For if (x_n) is Cauchy in H then $x_n \to x$ in E_q implying $\mathrm{d}x_n \to \mathrm{d}x$ in E. Also $\mathrm{d}x_n \to y$ in E_p implies $\mathrm{d}x_n \to y$ in E. Since E is Hausdorff, it follows that $y = \mathrm{d}x$.)

By §12.13.5 and the same argument as in §12.15.2 we deduce that there exists some q such that $\mathrm{d}H = B_p$. We shall fix such a q and fix the corresponding space H. Notice that for this step we have not used the full thrust of the acyclicity but only that $\operatorname{Im} \mathrm{d}$ is closed.

We then have the same diagram as in §12.15.2:

$$\begin{array}{ccccc} E_r \cap \mathrm{d}^{-1} E_q & \xrightarrow{\ \delta\ } & E_q \cap \mathrm{d}^{-1} E_p & \xrightarrow{\quad \mathrm{d}_0 \quad} & E_p, \\ & & & \searrow \quad \nearrow & \\ & & & B_p, & \end{array}$$

$$H^* \xleftarrow{\ \mathrm{d}_0^*\ } E_p^*.$$

Both δ and d_0 are induced by d; the dual map is d_0^*. Now we shall use the acyclicity and §12.13.5 to choose some $r \geqslant q$ so that $\operatorname{Im} \delta = \operatorname{Ker} \mathrm{d}_0$. We shall also use §12.12 to deduce that $\operatorname{Im} \mathrm{d}_0^* = (\operatorname{Ker} d_0)^{\perp}$.

Now let $s \geqslant r$ be such that d maps E_r continuously into E_s as explained. The following are natural inclusions (i.e. one-to-one):

$$E_s^* \subsetneq E_r^* \subsetneq E_q^* \subsetneq \left(E_q \cap \mathrm{d}^{-1}E_p\right)^* = H^*.$$

For the last observe that $E_0 \subset H$ because $p \geqslant n_0$.

Now let $S \in E_s^*$. Then S can be identified with an element of H^* and we have $\|S\|_{H^*} \lesssim \|S\|_s^*$. But also, if $bS = 0$, this element in H^* belongs to $\mathrm{Im}\,\mathrm{d}_0^*$. Indeed, for all $f \in \mathrm{Ker}\,\mathrm{d}_0 = \mathrm{Im}\,\delta$ we have $f = \mathrm{d}\phi$ for some $\phi \in E_r$ and therefore $\langle S, f\rangle = \langle S, \mathrm{d}\phi\rangle = \langle bS, \phi\rangle = 0$ (to see this use $\mathrm{d}\colon E_r \to E_s$, $b\colon E_r^* \leftarrow E_s^*$). This means that $S \in (\mathrm{Ker}\,\mathrm{d}_0)^\perp = \mathrm{Im}\,\mathrm{d}_0^*$.

On the other hand, it is clear, for the boundary operator, that we can factorise $b = i^* \circ \mathrm{d}_0^*\colon E_p^* \to E_0^*$ because

$$E_0 \xrightarrow{i} E_q \cap \mathrm{d}^{-1}E_p = H \xrightarrow{\mathrm{d}_0} E_p$$

for the canonical inclusion i.

We can now use Banach's theorem once more because $\mathrm{Im}\,\mathrm{d}_0^* = (\mathrm{Ker}\,\mathrm{d}_0)^\perp$ is closed in H^*. From this it follows that for all $S \in \mathrm{Im}\,\mathrm{d}_0^*$ there exists $T \in E_p^*$ such that $bT = S$ and $\|T\|_p^* \leqslant C\|S\|_{H^*}$. If we apply this to a closed element $S \in E_s^*$ as explained above, we deduce that $\|T\|_p^* \leqslant C\|S\|_s^*$ as needed in the ersatz acyclicity.

12A.4 The $\mathrm{Im}\,b$

In the above argument we saw that the first part of the proof works under the condition that $\mathrm{Im}\,\mathrm{d} \subset E$ is closed. What is actually used is that for every $p \geqslant 0$, $\mathrm{Im}\,\mathrm{d} \cap E_p$ is closed in E_p. Under that condition it is not clear that we can conclude for the mapping $b\colon E^* \to E^*$ that $\mathrm{Im}\,b \subset E^*$ is closed (probably not in general!) but we *can* conclude a slightly weaker condition on $\mathrm{Im}\,b$. Indeed, let us go back to the diagram of §12A.3 for some fixed p and q and the corresponding H. Also let $\pi\colon E^* \to H^*$ be the canonical (injective, i.e. one-to-one) mapping. Then for the closure $\overline{\mathrm{Im}\,b} \subset E^*$ we have $\pi(\overline{\mathrm{Im}\,b}) \subset \mathrm{Im}\,\mathrm{d}_0^*$ because $\mathrm{Im}\,\mathrm{d}_0^*$ is closed. The conclusion that we can also draw is therefore the following.

Let $S \in \overline{\mathrm{Im}\,b} \subset E^*$ and let $p \geqslant 0$. Then there exists $T \in E_p^*$ such that $bT = S$.

The importance of this observation lies in the fact that with this and the methods of Chapters 13, 14 (see §13A.1) we can prove a refinement of Theorem 12.21, namely the following.

Let G be some connected $\mathbb{C}$*-group and let (Λ_P, d) be the corresponding polynomial complex. Then* $\mathrm{Im}\,\mathrm{d}$ *is not a closed subspace of Λ_P.*

13

The Polynomial Homology for Simply Connected Soluble Groups

In this chapter we shall complete the proof of Theorem 12.17 and obtain thus the homological characterisation of the C–NC condition for soluble simply connected groups. This characterisation can be formulated directly, without the terminology of polynomial homology, by the Poincaré equation $d\theta = \omega$ on the group. The issue is whether we can solve this equation for a closed form ω of polynomial growth by a form θ of polynomial growth.

The NC-part of this theorem was proved in §12.9.8. So, unless otherwise stated, all the Lie groups G considered in this chapter will be assumed to be *soluble simply connected C-groups*.

We shall freely use the results and the notation of Chapter 12 and a comprehensive understanding of that chapter is essential for what follows, but unfortunately not only of Chapter 12. Many constructions and definitions from Chapters 7–10 will also be needed and will have to be reactivated as we go along. Therefore, in §13.1 we recall some of that material.

13.1 The Reductions and Notation of Chapters 8–10. The Organisation of the Proof

13.1.1 The basic reduction

The basic reduction from Proposition 8.3 combined with §9.1 consists in proving that every simply connected soluble group (C- or NC-) is polynomially homotopically equivalent to a group whose algebra is, in our case of a C-group, a special soluble algebra (SSA), as described in §9.1.7. We shall recall and fix the notation and use it freely in the rest of the chapter.

Such an SSA-group is of the form $G = N \leftthreetimes V$, where N is nilpotent and V is a Euclidean space. Furthermore, the action of V on N is semisimple and if

$\mathfrak{g}$, $\mathfrak{n}$, $\mathfrak{a}$ are the Lie algebras of G, N, V respectively all the roots of the action of $\mathfrak{a}$ on $\mathfrak{n}$ are real. All the groups here are simply connected: here V stands for 'vector space' and $\mathfrak{a}$ for 'Abelian'.

Since G is a C-group the constructions of §9.1.7 can be made and we have the two alternatives for these SSAC groups. The reader will have to go back to Chapter 9 for the details but here we shall recall the necessary notation and this will stay fixed for the rest of the chapter. The notation of §9.5 is

$$\begin{gathered} G = N \leftthreetimes V; \quad V = A' \oplus A, \quad N' \subset N, \\ G' = N' \leftthreetimes V = N' \leftthreetimes (A' \oplus A) = (N' \leftthreetimes A') \oplus A \subset G. \end{gathered} \tag{13.1}$$

Briefly, N' is a closed subgroup that is stabilised by the V-action and thus a subgroup $G' \subset G$ can be defined. The Euclidean space V splits into a direct sum with $A' \cong \mathbb{R}^{r-1}$, $A \cong \mathbb{R}^s$ with $r \geqslant 2$ and $s \geqslant 0$. The action of A on N' is trivial, hence the direct sum decomposition of G'.

We can distinguish two alternatives according to what $N' \leftthreetimes A'$ looks like:

The Abelian alternative In the case $N' = \mathbb{R}^r$ and the roots of the action in $N' \leftthreetimes A' = G_r$, with $r \geqslant 2$, have the special C-configuration of (9.33). The notation G_r was used for these groups in §9.2.1.

The Heisenberg alternative In the case $r = 2$, $A' = \mathbb{R}$ and $N' = \mathbb{H}$ is the Heisenberg group of Definition 9.5 and Example 11.20. In the group $N' \leftthreetimes A'$ the action of A' on N' is as explained there.

13.1.2 The $\mathrm{LL}(R) - \partial\,\square^r$ construction in the two alternatives

The distinction between the two alternatives is of course important, but once that distinction has been made, the proofs in Chapters 9 and 10 were essentially identical for the two cases.

In both cases the pivot was the construction of sets

$$S \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset N' \leftthreetimes A' \subset G' \subset G. \tag{13.2}$$

This is a short way of saying that there exists

$$f \in \mathrm{Lip}(\log R)^C, \quad f\colon \partial\,\square^r \longrightarrow N' \leftthreetimes A', \quad f(\partial\,\square^r) = S, \tag{13.3}$$

where f depends on the large parameter $R \gg 1$, but the constant C does not (cf. §9.3.2). In the Heisenberg case $r = 2$ and $\partial\,\square^2 \simeq S^1$, that is, the 1-sphere.

This summarises the first basic construction of Chapter 9, and when $s = 0$ that construction was sufficient for the proofs that we gave in Chapter 10.

13.1.3 The second basic construction

This first basic construction does not suffice when $s \geqslant 1$, and in §9.5 we gave the second basic construction that was needed.

For the latter, our starting point was the LL(R)-set constructed from the first construction. We proceeded from there to the second construction. This was done in a strictly identical manner in the two different cases of the *Abelian* and *Heisenberg* alternatives.

This second construction is simple enough when $s = 1$, but it becomes progressively more involved as $s = 2, 3, \ldots$.

13.1.4 The organisation of this chapter

The proofs of this chapter unwind in exactly the same way as in Chapters 9, 10. The difference lies in the fact that we have to be much more formal. The qualitative considerations for instance of §9.5 have to be replaced by algebraic relations between currents. Indeed, it is the systematic use of currents that makes the difference and allows us to pass from the mostly descriptive proofs of Chapters 9, 10 to the much more rigid proof of our theorem here. Two new geometric ideas will have to be incorporated in the process. The way that all this plays out will be described in the next few lines.

In §13.2 we shall revisit the first basic construction and reformulate it in terms of currents. Once this is done we shall in §13.3 prove the non-acyclicity for C-groups in Theorem 12.17 for the special case $s = 0$.

This special case $s = 0$ simplifies matters considerably since in that case only the first basic construction is needed. But even in this special case, to treat the case where the homologies are only finite and not zero, new ideas are needed. This problem did not arise in Chapters 9, 10. The new idea of bouquets of spheres will be introduced in §§13.4 and 13.4.3 and we shall illustrate these ideas by giving the proof that the homology for C-groups is infinite – shorthand for saying the dimension is infinite – for the simple case $s = 0$.

To treat the general case $s \geqslant 0$, the second basic construction is needed. But even in the simplest case $s = 1$ with an acyclic complex Λ_P a new geometric feature crops up. Furthermore, in the more general case of a finite homology for Λ_P yet another geometric problem has to be examined. To illustrate these two new geometric problems as clearly as possible we shall start from this special case $s = 1$ and complete the proof of the theorem there in §§13.5 and 13.6.4.1. This finishes the first part of this chapter and the reader who has understood this part well will already have a very good idea of the proof of Theorem 12.17.

The second part of the chapter for the general $s \geqslant 0$ could have been left as an exercise for the reader!

This second part may well be an exercise but it is an elaborate one. It takes the rather long §§13.6, 13.7 to complete. Nonetheless the task in hand merely consists of developing the necessary terminology and notation to push the ideas of §13.5 to higher dimensions. This terminology and notation have, not surprisingly, a strong flavour of homological algebra. Some readers may enjoy this, others on the other hand (like the author of this book) may not!

13.1.5 List of special cases

For the convenience of readers we shall make a list here that explains how we deal with the indices r, s and with $h = \dim\big(H(\Lambda_P)\big)$. When the complex is acyclic, $h = 1$ which comes from zero-degree cohomology, H_0. More generally, $1 \leqslant h < +\infty$ is the case of a complex of finite-dimensional homology. The implication (iii) $\Longrightarrow$ (i) in Theorem 12.17 can then be reformulated as follows:

$$[h < +\infty] \quad \text{and the C-condition are incompatible.}$$

First note that the index r is not an issue here. It certainly gave us a lot of trouble to pass from the easy case $r = 2$ to the general case in Chapters 9, 10 for the first basic construction. But this is now a done thing.

- The case $s = 0, h = 1$ is the easiest and is done in §13.3.
- The case $s = 0, h \geqslant 1$ is done in §13.4.3.
- The case $s = 1, h = 1$ is done in §13.5.
- The case $s \geqslant 1, h = 1$ is done in §13.6.
- The general case $s, h \geqslant 1$ is finally done in §13.7, but the special case $s = 1$, $h \geqslant 1$ can be found in §13.6.4.2.

By doing these special cases first we have made this chapter longer than it need be. But we also feel that this gives a much better chance for the reader to understand the ideas in the proofs.

13.2 The Currents Generated by the First Basic Construction

13.2.1 The definition of the currents

The notation for the first basic construction will be as in §§9.2–9.5 and more recently (13.2), (13.3) above:

$$S \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset N' \leftthreetimes A'. \tag{13.4}$$

This is the image of a $\mathrm{Lip}\,((\log R)^c)$ mapping $f\colon \partial\,\square^r \to N' \leftthreetimes A' \subset G'$. We have shown that we can even assume that $f\colon S^{r-1} \to N' \leftthreetimes A'$ can be smoothed out and be assumed to be C^∞ and in fact defines an $(r-1)$-dimensional embedded sphere $S^{r-1} \subseteq \mathbb{R}^{2r-1}$. Once we have smoothed f, it defines a chain and a current in $N' \leftthreetimes A'$ by the definition of de Rham (1960, pp. 27, 40), which was recalled in §12.3.3:

$$S[\omega] = \int_{\partial\square^r} f^*\omega. \tag{13.5}$$

For this definition the orientations have been fixed once and for all as follows. We fix an orientation on $\mathbb{R}^r$ for the Abelian alternative and this induces an orientation on the unit ball B^r or on $\square^r$ and thus also on the boundary S^{r-1} or the faces of $\partial\,\square^r$. For the Heisenberg alternative, S is just an embedded circle (i.e. the 1-sphere S^1) and thus we can orient clockwise or anticlockwise, and we do this once and for all.

If the function f is not assumed smooth but only Lipschitz (in fact piecewise affine) we can still use formula (13.5) to define a current because the pullback $f^*\omega$ of the form ω can again be defined and has L^∞-coefficients for the natural Lebesgue measure of $\partial\,\square^r$.

In what follows in this chapter we shall abuse notation and use the same letter S for the current defined in (13.5) and for the set S in (13.3) which is the support of this current. This notational convention and terminology will be convenient and will be adopted throughout without further mention; we shall, for instance, say that S is an LL(R) *current*. It should also be observed that if we use the notation of §12.8.6 for the current I^r then all the above can be summarised by saying that S is the direct image by the mapping f of the boundary of I^r. Formula (13.5) then says that

$$S = f_*(bI^r); \quad r \geqslant 2. \tag{13.6}$$

From (13.6), among other things it is clear that S is a closed current, that is, $bS = 0$. To define the direct image of these currents we shall always assume, as we may, that the function f is piecewise affine. Alternatively, the image of a normal current by a Lipschitz function can always be defined (see §§10.2.3, 10.2.6 or Federer, 1969).

13.2.2 The metric properties of the current S

For $S \in \mathscr{E}'(G') \subset \mathscr{E}'(G)$ as in (13.5) we shall normalise by left translation, as we may, and always assume that $f(0) = e \in G$ is the identity of G for 0 the origin of $\mathbb{R}^r$. This also applies in the Heisenberg case with $r = 2$. This means in both the Abelian and Heisenberg alternatives, that we have $e \in \operatorname{supp} S$ and therefore by the Lipschitz property of f we have

$$\operatorname{supp} S \subset \left[g \in G';\ |g|_{G'} \leqslant (\log R)^C\right] = \text{ball of radius } (\log R)^C;\ \ R \geqslant C, \quad (13.7)$$

for appropriate constants and the distance in the group G' from the origin $|g|_{G'} = d_{G'}(e, g)$.

From this and from the definition of the norms M_p in §12.7 we can state the essential properties of the current S:

$$\begin{aligned} S \in \mathscr{E}'(G') \subset \mathscr{E}'(G), \quad bS = 0, \quad M_j(S) \leqslant (\log R)^{c_j}; \\ R \geqslant C,\ j = 0, 1, 2, \ldots \end{aligned} \quad (13.8)$$

for constants $C, c_0, c_1, \ldots$ that are independent of R. It goes without saying that for the definition of the norms M_p we use left-invariant Riemannian structure and the neutral element $e \in G$ is the base point of the manifold.

Exercise 13.1 Verify this. Use (13.7) and also the fact that $f \in \operatorname{Lip}(\log R)^C$ implies that the Jacobians involved in the pullback of the definition (13.5) satisfy analogous $O((\log R)^C)$ uniform estimates. Notice also that $|g|_G \leqslant |g|_{G'}$ in (13.7) and that the M_j is (13.8) refers to (12.67) for the group G' or even for the group $N' \leftthreetimes A'$ because of (13.1). This, however, implies the same thing for the group G. (A similar implication the other way round for a group and a subgroup fails; see §2.14.)

13.3 The Special Case $s = 0$ and an Acyclic Complex

No new ideas will be needed in this section and it is only a matter of recycling the proofs of Chapter 10 using the new terminology of currents.

13.3.1 The contradiction and connections with Chapter 10

Here we shall assume that in the group $G = N \leftthreetimes (A' \oplus A)$ of §13.1.1, the group $A = \{0\}$: that is, $s = 0$; we shall also make the assumption that the polynomial homology of the complex Λ_P vanishes at the level $r-1, r$ (see §12.17):

$$H_j(\Lambda_P) = 0; \quad j = r-1, r. \tag{13.9}$$

This assumption will lead to a contradiction. In this way we shall complete the proof of Theorem 12.17 when Λ_P is acyclic and $s = 0$. The reason why we need the two dimensions r, $r-1$ will be the use of Proposition 12.59 in §13.3.2 below. Despite the formal resemblance, this is not directly related to the hypothesis of Proposition 9.34. It should rather be compared with (13.22) and with remark (iv) in §13.3.4.

At first sight it might appear awkward to give the proof by contradiction. This, however, both here and in all the other proofs, is the more convenient way to proceed. This, after all, is not very different from the attitude that was adopted in Chapters 9, 10 (especially §§9.5.2–9.5.5) but there, instead of the acyclicity properties (13.9), it is the filling $\mathscr{F}$-property that is used.

For the proof we shall use the chain and current S constructed from the first basic construction in §13.2. The transversality properties of the support of this current will be used (see §9.3.3). As already explained in §13.2.1, the support of the chain is also denoted by

$$S \text{ is an } \mathrm{LL}(R) - \partial\,\square^r \subset G'. \tag{13.10}$$

To fix ideas we shall assume that the group satisfies the Abelian alternative. The modifications needed for the Heisenberg alternative are automatic and they will be left as an exercise for the reader.

In the proof below we shall follow closely the proof given in §10.2. As there, the geometric considerations that follow are easier to visualise when the support S in (13.10) has been smoothed out first and is an S^{r-1}, that is, a C^∞ embedded sphere in G'. The smoothing out process in §10.3 was, however, quite elaborate. Furthermore, the proof we gave in §10.2.4 does not use this smoothness and relies only on the transversality. The smoothing process was thus avoided. It is this second variant of the proof that will be adapted here but it will certainly help the reader to think in terms of smooth embedded spheres $S^{r-1} \subset G'$.

13.3.2 The properties of the current S

The two main properties of the current S that are used are not surprising. First there is the metric condition of (13.8) that $M_p(S) = \mathrm{O}\left((\log R)^{c_p}\right)$ for any $p \geqslant 0$. Then the transversality property of the support has to be used.

This metric condition combined with the acyclicity condition (13.9) will be used in Proposition 12.59. From this it follows that we can construct some current $S^1 = \hat{S} \in \mathscr{D}'_{(r)}(G)$, that is, a current of dimension r that satisfies

$$b\hat{S} = S, \quad M_0\left(\hat{S}\right) \leqslant CM_p(S) = \mathrm{O}\left((\log R)^c\right), \tag{13.11}$$

where the constants C, c and p are independent of R. The notation $\hat{S}$ for this current is designed to keep the analogy with the notation of Chapters 9 and 10, although $\hat{S}^1 = S^1$ would be better since this is consistent with the notation that will be used systematically later on in this chapter.

We shall examine now the transversality condition for the current S and to do that we shall proceed as in §10.2 and on $G = N \leftthreetimes A'$ we shall use exponential coordinates of the second kind (cf. §7.3.1) so that with these coordinates the general element $g \in G$ is represented $g = (X,Y)$, with $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ being these coordinates for the nilpotent group N and $Y = (y_1, \dots, y_{r-1}) \in \mathbb{R}^{r-1}$ being the Euclidean coordinates of the Abelian $A' \cong \mathbb{R}^{r-1}$. In terms of these coordinates the transversality condition on S gives the following.

We can define a narrow slice N_c,

$$N \subset [g = (X,Y);\ |Y| \leqslant c] = N_c \subset G, \tag{13.12}$$

for some appropriately small but fixed (i.e. independent of R) constant c. This slice has the following property.

Let us denote by $\partial_0 \square_R^r \subset \square_R^r \subset N'$ the vertices of the cube $\square_R^r$ that was used for the construction of S. Let us also denote $V_c = [v \in A';\ |v| < c]$ for the same constant c as in (13.12) and for each $P \in \partial_0 \square_R^r$ write

$$E_P = \{P\} \times V_c \subset G. \tag{13.13}$$

We then have

$$\operatorname{supp} S \cap N_c = \bigcup \left[E_P;\ P \in \partial_0 \square_R^r\right]. \tag{13.14}$$

This is not all the information contained in the transversality condition of Chapters 9 and 10, but it is enough to remind the reader what it is all about and also to induce them to go back to these chapters for more details. Notice both here and in the analogous transversality conditions of this chapter, that $\operatorname{supp} S \subset G'$ and in (13.14) we cut it with a subset of G and that a slight abuse of notation is made because to do this we must consider $\operatorname{supp} S \subset G$.

13.3.3 The construction of the differential form and the contradiction

The pivot for the contradiction that will complete the proof is, as in Chapter 10, the differential form ω that we constructed in §10.2. This differential form depends on S and on R. The reader should go back to Chapter 10 for the details but for convenience we shall recall here the salient features of that differential form.

The exponential coordinates (X,Y) of the previous subsection for G are used and we define this C^∞ differential form by

$$\omega = \varphi(g)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{r-1}; \quad g = (X,Y) \in G. \tag{13.15}$$

We require that it has the following three properties:

(i) The support of ω is compact and also $\operatorname{supp}\omega \subset N_c$ for the constant c in (13.12) where the transversality condition holds.

(ii) There exist constants that are independent of R for which we have (notation of §12.5.1)

$$\|\mathrm{d}\omega\|_0 = \sup|\mathrm{d}\omega| \leqslant CR^{-c}; \quad R \geqslant C. \tag{13.16}$$

As usual, the Riemannian norm $|\mathrm{d}\omega|$ is taken with respect to the left-invariant Riemannian structure on G. What is important is that we can accommodate property (13.16) with the next property below.

(iii) We shall fix at will some $\varphi_0 \in C^\infty(\mathbb{R}^{r-1})$ such that

$$\operatorname{supp}\varphi_0 \subset [Y;\ |Y - Y_0| \leqslant c_0] \subset V_c = [|Y| \leqslant c] \tag{13.17}$$

for some small constant c_0 and some $Y_0 \in \mathbb{R}^{r-1} = A'$ sufficiently close to 0, where c and V_c are as in (13.13). The additional condition on $\omega = \varphi\,\mathrm{d}y_1 \wedge \cdots$ that will be imposed is that φ takes the following preassigned values on the sets E_P of (13.13):

$$\varphi(P,Y) = \varepsilon_P\varphi_0(Y); \quad Y \in V_c,\ P \in \partial_0\square_R^r \subset N. \tag{13.18}$$

The ε are $\varepsilon_P = 0, \pm 1$ and all but one will in fact be chosen to be 0. By the transversality condition we can then choose φ_0 and the ε_P so that

$$S[\omega] = \langle S, \omega\rangle \geqslant 1. \tag{13.19}$$

How this construction was done was explained in detail in §10.2. The presence of the localising function was also explained in §10.2.4. In the case when S is a C^∞ embedded S^{r-1} sphere, this additional localisation is unnecessary. In fact, this additional localisation is never necessary, even without having gone through the smoothing first. The actual raison d'être of this localisation is that it means we are spared from 'chasing' the global orientations of S near the vertices $P \in \partial_0\square_R^r$.

For more details on all this and on the compatibility of (13.16), (13.18) the reader should refer back to §10.2.1. For the reader's convenience, recall however, that for this we set $\varphi = \varphi_1(X)\varphi_0(Y)$ and for the norm we can use the direct product $N \times V$ Riemannian structure since the support of φ_0 is as in (13.17); cf. §10.1.3. We then have $\mathrm{d}\omega = \mathrm{d}\varphi_1 \wedge \varphi_0 \, \mathrm{d}y_1 \wedge \cdots$, $|\mathrm{d}\omega| \lesssim |\mathrm{d}\varphi_1|_N$ and this can be estimated by (13.16) as long as the mutual distances of the vertices $\partial_0 \square_R^r$ in N are $\gtrsim R^c$.

Having constructed the above differential form ω, the contradiction that we promised now follows on one line because (13.11), (13.19) give

$$1 \leqslant \langle S, \omega \rangle = \langle d\omega, \hat{S} \rangle \leqslant \|d\omega\|_0 M_0\left(\hat{S}\right) = \mathrm{O}\left(R^{-c}\right) \tag{13.20}$$

for some fixed $c > 0$ (independent of R). This completes the proof.

13.3.4 Additional comments and remarks

(i) First of all, let us observe that the proof that we gave above is no different from the proofs that we gave in §10.2 under the conditions $\mathscr{F}_{r-1} \ldots$. The proofs of that chapter did not go through a contradiction and they were formulated in terms of direct estimates, but that different presentation is purely window dressing. The similarity (in fact verbatim identical) of these proofs is not surprising because the $\mathscr{F}_{r-1}$ applied to the current S says that we can find a new current $\hat{S}$ such that $b\hat{S} = S$ and where

$$\hat{S} \text{ is an } \mathrm{LL}(R) - \square^r \subset G. \tag{13.21}$$

This is a special acyclicity condition that in terms of algebraic topology could informally be summarised as follows. All these spherical chains in $\pi_{r-1} \subset H_{r-1}(\Lambda_P)$ are boundaries and give trivial homology; better still, they are trivial in the homotopy group.

(ii) The next observation is that we can carry the same proof out if instead of the hypotheses (13.9) we assume that for the dual complex Λ_P^* of Chapter 12 we have

$$H_{r-1}\left(\Lambda_P^*\right) = 0. \tag{13.22}$$

Strictly, nothing changes and in particular $\hat{S}$ exists as in (13.11). But now the reason why such an $\hat{S}$ can be constructed is not Proposition 12.59 but Proposition 12.55 (cf. remark (iv) below).

(iii) Note that the way that this proof adapts to the Heisenberg alternative is automatic. In this case we have $G' = \mathbb{H} \rightthreetimes \mathbb{R}$, where now $r = 2$, and the current $S = f_*(bI^2)$ in G is constructed as in §§13.2.1, 13.2.2. The geometric details of the construction can be found in §§9.2.5, 9.2.6. The construction of the differential form ω on G in §13.3.3 easily adapts. We obtain, as before, analogues of (13.19), (13.20) and the required contradiction.

Exercise 13.2 Write the details out for this.

(iv) Finally, we come back to the two dimensions $r, r-1$ in the acyclicity condition (13.3). Similar conditions, for example (13.36), (13.79) and so on, will crop up right through the chapter. It is of course the use of Proposition 12.59 that makes these conditions necessary. The best attitude in reading this chapter is not to take this too seriously and make the stronger assumption that Λ_P is acyclic or that it has finite-dimensional homology as the case might be (i.e. for the homology of all dimensions). Then we only need to use Proposition 12.57 and can ignore the more specialised §12.17. Theorem 12.17 follows anyhow.

13.4 Bouquets of Currents

A new geometric construction will be introduced in this section. This will be a finite collection of currents S constructed by the first basic construction as in §13.2. The supports of these finitely many currents all contain the origin but apart from that they are 'far from each other'; see Figures 13.1, 9.2. These will be referred to as bouquets (of currents). If we go as far as to smooth out in the first basic construction then each support S is an embedded C^∞ sphere and the union of these supports really looks like a bouquet of spheres. This smoothing, however, will not be done in general.

13.4.1 Definition for the Abelian alternative

We recall the notation $G_r = \mathbb{R}^r \leftthreetimes \mathbb{R}^{r-1} = N' \leftthreetimes A' \subset G'$ and $\square_R^r \subset N'$ from §13.1 and Chapter 9. Here the construction will be made in G_r (see §9.2.1); $r \geqslant 2$ will stay fixed and $R \gg 1$ will as usual denote the large basic parameter on which the first main construction depends. The construction below will depend on $R \gg 1$ and we proceed as follows.

The first step is to consider not one cube but a finite collection of cubes

$$\square_j = \square_{c_jR}^r = [0, c_jR]^r \subset \mathbb{R}^r = N'; \quad 1 \leqslant j \leqslant m \tag{13.23}$$

of size c_jR with distinct constants c_j, that are fixed once and for all and are independent of R. For clarity one should think of these constants as being large and geometrically very far from each other, that is, that for $i \neq j$ either c_i/c_j or c_j/c_i is, say larger than 10^{10}. These cubes are contained in each other like Russian dolls and have the origin in common and the whole configuration depends on the parameter R (see Figure 13.1(a)).

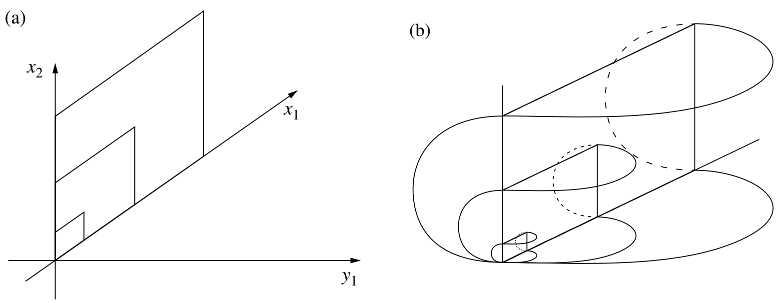
We now use each of these cubes $\square_j$ to make the first basic construction of §§9.3–9.4 so that we obtain

$$S_j \text{ is an } \mathrm{LL}(R) - \partial\square^r \subset G_r,$$
$$(\text{the vertices}) = \partial_0\square_j \subset \square_j \cap S_j; \quad 1 \leqslant j \leqslant m. \tag{13.24}$$

In the smooth case these are embedded spheres that go through the origin (see Figures 13.1(b), 9.2).

To each S_j we associate the corresponding current and we define the bouquet of currents

$$\mathscr{B} = (S_1, \ldots, S_m), \quad |\mathscr{B}| = \bigcup_j \operatorname{supp} S_j. \tag{13.25}$$



The cubes $\square^2$ that give the 'Russian dolls' in G_2.

A bouquet of (one-dimensional) currents.

Figure 13.1

The supports of these currents have in common the neutral element $e \in G_r$ and therefore also the set

$$\{0\} \times V_a \subset N' \leftthreetimes A'; \quad V_a = \left[v \in A';\ |v| \leqslant a\right] \tag{13.26}$$

for some constant a independent of R, where R is always assumed to be large. But apart from this, the intersections of these supports with the slice N_c of (13.12) are disjoint. This is expressed formally in (13.28). We could describe this informally by saying that these supports are 'as disjoint as they can be'.

By the properties of the first basic construction (13.8) there exist constants c'_p, $p = 1, \ldots,$ and $C > 0$ independent of R such that

$$M_p(S_j) \leqslant (\log R)^{c'_p}; \quad R \geqslant C,\ 1 \leqslant j \leqslant m,\ p = 1, 2, \ldots. \tag{13.27}$$

The transversality condition of the first basic construction will now give the following transversality property on the set $|\mathscr{B}|$ of (13.25).

We shall abuse notation somewhat by writing $\partial_0\mathscr{B} = \bigcup \partial_0 \square_j$ the set of all the vertices of the cubes (13.25) in the definition of the bouquet. We recall also the definition of the canonical projection $\pi\colon G_r = N' \leftthreetimes A' \to A' = \mathbb{R}^{r-1}$ and $V_c = [v \in A';\ |v| < c]$ for some appropriately small $c > 0$. The transversality for the bouquet now says, among other things, that

$$|\mathscr{B}| \bigcap \pi^{-1}(V_c) = \bigcup_P [\{P\} \times V_c;\ P \in \partial_0\mathscr{B}] = \bigcup_P E_P. \tag{13.28}$$

More information is of course contained in the transversality condition (see §§9.3, 9.4) and this information will be used in what follows.

13.4.2 The Heisenberg alternative

The modifications needed in the above definition of a bouquet in the Heisenberg alternative are rather obvious.

Here $N' \leftthreetimes A' = \mathbb{H} \leftthreetimes \mathbb{R}$ as in §13.1 and the first basic construction gives

$$S \text{ is an } \mathrm{LL}(R) - \partial \square_R^2 \subset N' \leftthreetimes A'. \tag{13.29}$$

The reader will have to go back to §9.2 to check out the notation (in §9.2 the Heisenberg group is denoted by H) and it will also be left as an exercise to give the analogous definition of a bouquet in this case. The construction is, in fact, if anything, easier to visualise here because $r = 2$ and S is an embedded circle. The additional complication is that $S \cap \mathbb{H}$ consists of six points and is not just the four vertices of $\square_R^2$. All this was explained in detail in §9.2 and it is easy to see that this makes no difference in the analogue of (13.28) and in the further use that we shall make.

Exercise Verify the above. The $\partial \square$ that are represented by the Russian dolls of Figure 13.1(a) are now replaced by the $c_j R$ $(1 \leqslant j \leqslant m)$ 'scalar multiples' of the segments of the word (9.44) of Figure 9.3. For this use the fact that the Heisenberg group admits a natural *dilation structure* $g \to \lambda \cdot g, g \in \mathbb{H}, \lambda > 0$.

13.4.3 The first illustration of the bouquet of currents

We shall preserve the notation of §13.3 and assume that $G = N \leftthreetimes A'$ and that $A = \{0\}, s = 0$. To fix ideas we shall assume that we are in the Abelian alternative.

Now, instead of the acyclicity condition (13.9), we shall assume that the homology is finite-dimensional:

$$\dim H_j(\Lambda_P) < m; \quad j = r, r-1. \tag{13.30}$$

Here we shall use the bouquets defined in §13.4.1 with m in (13.25) equal to the bound m for the dimension of the homology in (13.30). As before, from these we shall produce a contradiction. This will complete the proof of Theorem 12.17 for the special case $s = 0$.

In the proof given in §13.3 we used (13.9) and Proposition 12.59. Here, under condition (13.30), we shall need to use the refinement in Propositions 12.58 and 12.59 and this will force us to consider a bouquet

$$\mathscr{B} = (S_1, \dots, S_m) \tag{13.31}$$

of currents as defined in §13.4.1.

Then by §§12.16, 12.17 we can find scalars $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ such that we can solve with bounds the b-equation in $\hat{S} \in \mathscr{D}'_{(r)}(G)$,

$$b\hat{S} = \sum \lambda_j S_j = S, \quad \sup|\lambda_j| = 1, \quad M_0(\hat{S}) \leqslant C \sum M_p(S_j) \leqslant C(\log R)^C, \tag{13.32}$$

where the constants C and p are independent of R. Furthermore, the transversality condition holds:

$$\begin{aligned} &\operatorname{supp} S \bigcap [|Y| \leqslant c] \subset \bigcup [E_P;\ P \in \partial_0 \mathscr{B}], \quad E_P = \{P\} \times V_c, \\ &[|Y| \leqslant c] = \pi^{-1}(V_c) \ \text{with } V_c \text{ as in (13.13)}, \end{aligned} \tag{13.33}$$

and where we use again the exponential coordinates $g = (X, Y) \in N \leftthreetimes A' = G$ of §13.3.1 for the group and $\pi\colon G \to A'$ is again the canonical projection. In (13.33) the left-hand side could be a proper subset, for example some of the λ could be zero.

The definition of the form ω of §13.3.3 is identical and satisfies properties (i), (ii) and (iii). Here, in condition (iii),

$$\varphi(P, Y) = \varepsilon_P \varphi_0(Y); \quad P \in \partial_0 \mathscr{B},\ \varepsilon_P = 0, \pm 1, \tag{13.34}$$

P runs through *all* the vertices and as before all but one of the ε_P is 0. As for φ_0, it is chosen as in (13.17). The only additional provision now is that in the choice of Y_0 and c_0 we have to make sure that we stay on the set on which *all* the mappings $f_j\colon \square_j \to G_r$ that define all the currents S_j of $\mathscr{B}$ are smooth (in fact affine).

The important new condition here is that if $P_0 \in \partial_0 \mathscr{B}$ is the vertex for which $\varepsilon_{P_0} = \pm 1$ (and $\varepsilon_P = 0$ for all the others) then

(i) $P_0 \neq e$ (= the common vertex);
(ii) $P_0 \in \partial_0 \square_{j_0}$ is a vertex of the cube $\square_{j_0}$ and for that index $|\lambda_{j_0}| = 1$. This is possible by our hypothesis on the λ in (13.32).

By these conditions it follows as before that for the currents $\hat{S}$ and S of (13.32) we have again as in (13.19), (13.20),

$$S[\omega] = \langle S, \omega \rangle \geqslant 1.$$
$$1 \leqslant \langle S, \omega \rangle = \langle \mathrm{d}\omega, \hat{S} \rangle \leqslant \|\mathrm{d}\omega\|_0 M_0(\hat{S}) = \mathrm{O}(R^{-c}). \tag{13.35}$$

Hence the contradiction.

Exercise 13.3 Fill in the details to make the same proof work in the Heisenberg alternative.

13.5 The Case $G = N \leftthreetimes (A' \oplus A)$ with $A = \mathbb{R}$ and $s = 1$

Unless otherwise stated, we shall fix ideas and assume that we are in the Abelian alternative of §13.1.1. This special case already contains the main new geometric idea that is needed. Furthermore, we can treat this case in just two steps and we can avoid the elaborate inductive process needed for general groups and therefore this new idea is highlighted much better. The hypothesis that will be made throughout in this section is that

$$H_{r-1}(\Lambda_P) = H_r(\Lambda_P) = H_{r+1}(\Lambda_P) = 0. \tag{13.36}$$

This hypothesis, together with the C-condition, will lead to a contradiction as in §13.3.3 and will thus complete the proof of Theorem 12.17 in this special case.

Here, remark (iv) in §13.3.4 very much applies again and we recommend that in a first reading you should assume that the complex Λ_P is acyclic. This is of course a stronger assumption but with it we avoid having to chase the dimensions around when using Proposition 12.59.

13.5.1 Notation

We shall introduce an important notational convention that will be adopted for the rest of the chapter and which will help us keep track of the dimensions in the various formulas. Several, in fact most, of the currents in this chapter will be denoted by capital letters: $S, \hat{S}, T, \ldots$. For these currents (but only for the ones that are denoted by *capital letters*) we shall often insert an exponent S^p, T^p, etc. to indicate that the dimension of the current is $(r-1+p)$, that is, $S^p, T^p, \ldots \in \mathscr{D}'_{(r-1+p)}$.

We shall also recall some of the notation that was introduced in Chapter 12 and which will be used here: $\delta_g \in \mathscr{E}'_{(0)}(G)$, with $g \in G$, is the Dirac mass at

g and, for $a \in A$ (the subgroup of G), $\delta_a \in \mathscr{E}'_{(0)}(A) \subset \mathscr{E}'_{(0)}(G)$. These are zero-dimensional currents. Similarly, $I = [0,1] \subset \mathbb{R} = A$ with the positive orientation is identified with a current $I \in \mathscr{E}'_{(1)}(A) \subset \mathscr{E}'_{(1)}(G)$ of dimension 1 (see (12.85)). We then have

$$bI = \delta_1 - \delta_0. \tag{13.37}$$

We shall also use the notation S ($= S^0$ with our notational convention) for the current constructed by the first basic construction in §13.2. This current will depend on the free parameter $R \gg 1$ and the letter R will be reserved just for that parameter.

13.5.2 The choice of the index p_1

In the general case $s \geqslant 1$, a whole sequence of indices has to be chosen backwards as explained in §12.15.3.2. We shall return to this in a systematic way in §13.6.1. Here, for $s = 1$ things simplify and we just have to choose one index $p_1 \geqslant 10$ and this is done by applying Proposition 12.57, or rather, the refinement of §12.17, and hypothesis (13.36). We define p_1 by the following condition.

For all closed currents U (i.e. $bU = 0$), as in Propositions 12.57 or 12.59 with finite M_{p_1}-norm, we can solve with bounds the following equation in T:

$$bT = U, \quad M_0(T) \leqslant CM_{p_1}(U), \tag{13.38}$$

with a constant C that is independent of U. Here we use hypothesis (13.36).

Having fixed p_1 from the same Proposition 12.59, it is then also true that for all closed currents $S \in \mathscr{D}'_{(r-1)}(G)$ we can solve the following equation in $\hat{S}$ with bounds

$$b\hat{S} = S, \qquad M_{p_1}(\hat{S}) \leqslant CM_p(S), \tag{13.39}$$

for some C and p that are independent of S. We have not indicated the fact that $M_p(S)$ is finite here because if not, the validity of (13.39) is formally correct (the same convention applies to (13.81), (13.88), (13.90) and other similar estimates below).

In (13.38) and (13.39) the dimensions are as follows: $\dim U = r$ (i.e. $U = U^1$ with our convention) and $T = T^2$. Also, with this notation we should set $S = S^0$, $\hat{S} = \hat{S}^1$, hence the notation for p_1. Similarly, for p in (13.39) we should perhaps write $p = p_0$. More precisely, in what follows the $S \in \mathscr{E}'_{(0)}(G_r)$ will be fixed and it will be the current that we constructed from the first basic construction in §13.2 for which we have

$$M_p(S) = \mathrm{O}((\log R)^{c_p}); \quad p > 0 \tag{13.40}$$

for constants c_p independent of R. Here and throughout we use the notation of §13.1.1 and identify the subgroup $N' \leftthreetimes A'$ with G_r. The current $\hat{S}$ will then be constructed as in (13.39) and it will satisfy

$$M_{p_1}\left(\hat{S}\right) = \mathrm{O}\left((\log R)^c\right) \tag{13.41}$$

for some constant independent of R.

13.5.3 The currents in the second basic construction

We shall maintain the notation δ_a, I and S, $\hat{S}$, p_1 from §§13.2 and 13.5.2 and use the product of §12.8.4 to define

$$S \cdot I = I \cdot S \in \mathscr{E}'_{(r)}(G). \tag{13.42}$$

For typographical reasons, when confusion does not arise, we shall drop the 'dot' in the product, that is, $S \cdot I$ above would become SI.

Exercise 13.4 Use the fact that A is central in G' and $I \in \mathscr{E}'(A)$, $S \in \mathscr{E}'(G')$ to verify this. This commutation relation is not essential for the construction. For more details see (13.99) below. On the other hand, the fact that A and G_r are direct factors in G makes (13.42) certainly not surprising (think about it). This will be used in an essential way in (13.48) below.

We shall also denote

$$I_D = \sum_{n=-D}^{D-1} \delta_n I = [-D, D] \in \mathscr{E}'_{(1)}(A) \tag{13.43}$$

and here the correct way to define the current I_D is by the chain $[-D, D]$ oriented by the same orientation as $I = [0,1]$ (see §12.8.6 and de Rham, 1960, §14). We shall now define the key new currents as follows:

$$T = T^1 = IS + \hat{S} - \delta_1 \hat{S}$$

$$T^1(I_D) = \sum_{n=-D}^{D-1} \delta_n T^1 = I_D S + \delta_{-D}\hat{S} - \delta_D \hat{S}; \quad D = 1, 2, \ldots. \tag{13.44}$$

We use the convention of §13.5.1; the dimension of these currents is $\dim T^1 = r$. By formula (13.37) we have (see (12.84))

$$b(IS) = (\delta_1 - \delta_0)\, S; \quad b(T) = 0 \tag{13.45}$$

because $bS = 0$ is closed. Therefore $bT^1(I_D) = 0$.

Given that G_r and A commute (i.e. every element of G_r commutes with every element of A), the reader should notice that, by (13.42), in the above definition we certainly have $IS = SI$ but a priori $\delta_a \hat{S} \neq \hat{S}\delta_a$, with $a \in A$, and the order by which we multiply in (13.44) cannot be interchanged (see (13.99)).

Now we come to another crucial feature of the definition (13.44). This is the choice of D, which will also depend on the basic parameter R. We shall have $D \sim (\log R)^{\hat{c}}$ for some $\hat{c} > 0$. We shall set

$$D = \left[(\log R)^{\hat{c}}\right] = \text{integer part of } (\log R)^{\hat{c}}. \tag{13.46}$$

The important point here is that the choice of the constant $\hat{c}$ is independent of R and that this choice will be made at the very end of the proof. This constant $\hat{c}$ will depend on the other previous constants of the construction of S and the other geometric constants of G. Putting it differently, the constants c_j and C from (13.8) and (13.39) for which

$$M_j(S) = \mathrm{O}\big((\log R)^{c_j}\big), \qquad M_{p_1}\big(\hat{S}\big) = \mathrm{O}\big((\log R)^{C}\big) \tag{13.47}$$

are independent of $\hat{c}$. So therefore are the constants in the following estimates:

$$M_{p_1}(IS), \quad M_{p_1}(\hat{S}), \quad M_{p_1}(T) = \mathrm{O}\big((\log R)^{c}\big). \tag{13.48}$$

The commutation between A and G_r is used for the first estimate (see the exercise below). From these and §12.8.2 we deduce that we have

$$M_{p_1}\left(T^1(I_D)\right) = \mathrm{O}\,(\log R)^{\hat{C}} \tag{13.49}$$

for some new $\hat{C}$ that now depends on the $\hat{c}$ of (13.46).

Exercise 13.5 In a product situation $M_0\big(\mathscr{A}(T \times S)\big) = M_0(T)M_0(S)$; cf. Exercise 12.41. Use the homomorphism $H \times K \to HK \subset G$ for two (pointwise) commuting subgroups $H, K \subset G$ to conclude that $M_0(TS) \leqslant M_0(T)M_0(S)$ for $T \in \mathscr{D}'(H), S \in \mathscr{D}'(K)$.

13.5.4 The restriction of these currents to the slice

As in §§13.1, 13.3 we shall write

$$\begin{gathered}\pi\colon G = N \leftthreetimes V \longrightarrow V, \quad V = A' \oplus A, \quad V_c = [v \in V;\ |v| \leqslant c], \\ N_c = \pi^{-1}(V_c) = [g = (X, Y);\ |Y| \leqslant c] \subset G\end{gathered} \tag{13.50}$$

for the canonical projection π, the c-ball V_c and the slice N_c. Here the same exponential coordinates as in §13.3 are used: $X = (x_1, \ldots, x_n)$ for N and $Y = (y_1, \ldots, y_r) \in \mathbb{R}^r$ for the Euclidean space V.

We shall denote by χ the characteristic function of the slice and multiply the corresponding integration currents (i.e. measures as coefficients and therefore this multiplication is possible):

$$\chi T^1 (I_D) = \chi (I_D S) + \text{error term}, \tag{13.51}$$

where

$$\text{error term} = \chi \left(\delta_{-D}\hat{S}\right) - \chi \left(\delta_D \hat{S}\right). \tag{13.52}$$

No confusion should arise in this notation, where $I_D S$ refers to the multiplication of §12.8.4 that is induced by the group multiplication, and $\chi(\cdots)$ is the multiplication of an integration current by the characteristic function of a closed set.

Furthermore, by the fact that

$$\operatorname{supp} I_D \subset A; \quad \operatorname{supp} S \subset N' \leftthreetimes A' \tag{13.53}$$

we deduce that

$$\text{principal term} = \chi (I_D S) = I_c (\chi S) \tag{13.54}$$

for the chain and current $I_c = [-c, c]$ of A with the same orientation as I_D. For the identity (13.54) to hold, the norm $|v|$ on the Euclidean space V will throughout be taken to be $v = \sup_i |v_i|$ for the Euclidean coordinates $v = (v', v_r) \subset V = A' \oplus A$, $v' \in A'$, $v_r \in A$. This new notation for the coordinates will be adopted; so for the exponential coordinates of the group $G = N \leftthreetimes V$, $g = (X, Y)$ that we used in (13.50) we now set $Y = (v_1, \dots, v_r) = (y_1, \dots, y_r)$.

With these coordinates the principal term in (13.51) can now be expressed quite explicitly provided of course that c is small enough. This will be done in the §13.5.4.1.

13.5.4.1 The principal term and the transversality condition From the transversality condition on S it follows from (13.14) that

$$\begin{aligned} &\operatorname{supp}(\chi S) = \bigcup_P E'_P, \quad E'_P = \left[\{P\} \times V'_c\right], \\ &V'_c = \left[v \in A';\ |v'| < c\right]; \quad P \in \partial_0 \square^r_R \subset N'. \end{aligned} \tag{13.55}$$

Here P runs through the vertices of the cube that was used in the first basic construction for S and to fix ideas, as said, we shall assume that we are in the Abelian alternative.

From this and (13.54) it follows that

$$\operatorname{supp}(\chi I_D S) = \bigcup_P E_P; \quad E_P = [\{P\} \times V_c],\ V_c = [v \in V;\ |v| \leqslant c], \tag{13.56}$$

for the same set of vertices P as in (13.55). In fact, considerably more than the above description of the supports can be asserted from the transversality properties of S.

We shall now express the principal term as a sum of explicit currents. Towards this let us denote by $[V_c']$ the current in A' defined by the chain

$$\left[v' \in A';\ |v'| < c\right] \in \mathscr{E}'(A'), \tag{13.57}$$

where the orientation of A' has been fixed by the order of the coordinates $v_1, v_2, \ldots$. Similarly, the chain

$$[V_c] = [v \in V;\ |v| < c] \in \mathscr{E}'(V) \tag{13.58}$$

can be defined for the same orientation of V. In all the above definitions the constant c is assumed sufficiently small. Both the above currents can be interpreted as the product currents $I^{r-1} \in \mathscr{E}'(A')$, $I^r \in \mathscr{E}'(V)$ that were examined in §12.8.6. These currents can be translated on the left in G by the vertices P of (13.55) and we obtain new currents $\delta_P[V_c']$, $\delta_P[V_c] \in \mathscr{E}'(G)$; cf. §12.8.4.

Using these currents and the full content of the transversality conditions as explained in §§9.3.3, 9.4.6, 9.5.4 we can then assert that there exist $\eta_P, \eta_P' = \pm 1$ such that

$$\chi S = \sum_P \eta_P' \delta_P \left[V_c'\right], \tag{13.59}$$

$$\chi I_D S = \sum_P \eta_P \delta_P \left[V_c\right], \tag{13.60}$$

where the ± 1 in the η are needed in both (13.59) and (13.60) to take care of the orientations.

These orientations for the chain S near the vertices were already an issue in Chapter 10. There was no problem to define these once we assumed that $S \subset G_r$, an embedded sphere. This is what was done in §10.2.4. Serious problems, however, arise when we try to define the orientation of a whole neighbourhood of a vertex $P \in S$ when S was not smoothed out and was only piecewise affine. Some explanation of how to do this was given in Exercise 10.9.

This additional difficulty was evaded in Chapter 10 and in §13.4.3 by the device of localising further. The same thing can be done here in §13.5.4.2 below, which, however, the reader is advised to skip, at least in a first reading, as additional awkward notation is needed to describe the corresponding currents.

13.5.4.2 The additional localisation near the vertices We shall be careful to explain again, in precise terms, this additional localisation.

We recall that the first basic construction as done in Chapter 9 was piecewise affine. This is reflected in the precise definition of the transversality in §9.4.6.

Observe also that this is essentially automatic when $\dim S = 1$, for example in the case of the Heisenberg alternative. Be that as it may, from this piecewise affine structure of S near the vertices we can proceed as follows.

First of all let us change the notation slightly, so that the formulas below are more readable, and write $A' = V' \subset V$. Also, for some point $v_0' \in V'$ sufficiently close to 0 and for $c_0 > 0$ sufficiently small we shall write

$$\begin{aligned} \widetilde{V}_{c_0} &= \left[v' \in V';\ |v' - v_0'| \leqslant c_0\right] \subset V_c', \\ \tilde{\chi} &= \text{characteristic function of } \pi^{-1}\left(\widetilde{V}_{c_0}\right) \subset N \leftthreetimes V' \\ &\quad \text{for } \pi\colon N \leftthreetimes V' \longrightarrow V'. \end{aligned} \tag{13.61}$$

By the piecewise affine nature of the construction we further see that if v_0' and c_0 in (13.61) are appropriately chosen then the mapping $f\colon \partial\,\square^r \to S \subset N \leftthreetimes V'$ is affine 'above all the points of $\widetilde{V}_{c_0}$', that is, when the point ξ is such that $\pi \circ f(\xi) \subset \widetilde{V}_{c_0}$. It follows that an orientation can be defined on each of these patches:

$$\pi^{-1}\left(\widetilde{V}_{c_0}\right) \cap (\text{small neighbourhood of } P \text{ in } S); \quad P \in \partial_0\,\square_R^r. \tag{13.62}$$

From that orientation the corresponding signs η_P' and $\eta_P = \pm 1$ can be defined on the corresponding patch. More explicitly, the analogue of (13.59) and (13.60) holds once localised by $\tilde{\chi}$. We have

$$\tilde{\chi} S = \sum \eta_P'\, \delta_P \left[\widetilde{V}_{c_0}\right], \tag{13.63}$$

where $\left[\widetilde{V}_{c_0}\right] \subset \mathscr{E}'(V') \subset \mathscr{E}'(G)$ is the chain and the current defined by $\widetilde{V}_{c_0}$ of (13.61) and some fixed orientation of $V' = \mathbb{R}^{r-1}$.

Once we have this localisation for S we can localise $I_D S$ by $(\tilde{\chi} S) I_{c_0}$ for some, say the same as in (13.61), constant c_0 and obtain

$$(\tilde{\chi} S) I_{c_0} = \sum \eta_P' \delta_P \left[\widetilde{V}_{c_0}\right] I_{c_0}. \tag{13.64}$$

Here we use the product notation of §12.8.4 for the two currents I_{c_0} and $\left[\widetilde{V}_{c_0}\right]$. A more natural notation would perhaps have been $\widetilde{V}_c \times [I_{c_0}]$ for the chain $\widetilde{V}_{c_0} \times I_{c_0} = \widetilde{V}_{c_0} \times [-c_0, c_0]$ and orientations that have been fixed (this could, however, have created confusion with the notation of §12.8.5). Notice that, in the product, I_{c_0} commutes with S and δ_P.

The Heisenberg alternative All the above constructions adapt with no problem in the case of the Heisenberg alternative where we start with the S that we obtain from the first basic construction for the Heisenberg case. In fact, things are easier because here $r = 2$ and S is an embedded circle for free.

13.5.5 The error term and the estimate of $T^1(I_D)$

We start with a number of observations. First of all, for the left-invariant Riemannian distance on G we have for the slice N_c of (13.50),

$$\text{distance } (N_c, a) \geqslant c_1 (|a| - c); \quad a \in A \tag{13.65}$$

for some appropriate $c_1 > 0$. Here we identify of course $A \subset G$ and one can see (13.65) by using the projection $\pi\colon G = N \leftthreetimes (A' \oplus A) \to A$ which contracts distances.

Also, quite generally, by the definition for any integration current U on G we have (see (12.68))

$$M_0 (U \mathbb{I} (|g| > r)) \leqslant CM_p(U) r^{-p}; \quad r > 0,\ p > 0, \tag{13.66}$$

where $\mathbb{I}(\cdot)$ is the indicator function of the set outside the r-ball of G centred at e and C is independent of r (but depends of course on p).

This estimate for the total mass outside a ball of radius r can be applied to $\hat{S}$ and we can translate and get $\delta_m \hat{S}$ for $m = -D, D-1$ in the error term of (13.52). This will give the estimate of the total mass of $\delta_m \hat{S}$ outside the r-ball centred at $m \in A$ (see §12.8.2).

It is here that the constant $\hat{c}$ involved in the choice of D in (13.46) comes into play for the first time. Using (13.65) we obtain from the above that

$$M_0 \left(\chi \left(\delta_m \hat{S}\right)\right) \leqslant CM_{p_1} \left(\hat{S}\right) (\log R)^{-\hat{c} p_1}; \quad |m| \sim (\log R)^{\hat{c}}. \tag{13.67}$$

We recall also (see (13.41)) that

$$M_{p_1}(\hat{S}) = \mathrm{O}(\log R)^c \tag{13.68}$$

for some c that is independent of the choice of $\hat{c}$. Note that p_1 is also independent of $\hat{c}$. The conclusion is that we can choose the $\hat{c}$ so that the error term in (13.67) satisfies

$$M_0(\text{error term}) = \mathrm{O}(\log R)^{-10}. \tag{13.69}$$

We shall recapitulate what was done in the last two subsections and we shall, abusing notation somewhat, write

$$\chi T^1 (I_D) = \sum_P \eta_P \delta_P [V_c] + \mathrm{O}\left[(\log R)^{-10}\right], \tag{13.70}$$

where the O$-$ is understood in the sense of (13.69).

It is this O-error term that is the new idea. This has to be incorporated in the second basic construction. This, and the way $\hat{c}$ is chosen in (13.46), is what one has to understand. The way $\hat{c}$ was chosen was of course already an issue in §9.5 where we had to deal with the $\mathscr{F}$-property, but there the difference was that, provided we went far enough, in other words by choosing $\hat{c}$ large enough, the error term disappeared altogether.

We have written (13.70) in terms of formula (13.60). The same thing can of course be written in terms of the finer localisation (13.64) except that the notation becomes more involved. The reader could take on the exercise of writing out complicated notation to spell that formula out. Similarly, an entirely analogous formula can be written in the Heisenberg alternative.

Exercise 13.6 Write these formulas out explicitly.

13.5.6 The construction of S^2

Having constructed T^1 in (13.44) with $bT^1 = 0$ we shall now use hypothesis (13.34) and the property of the index p_1 in (13.38) to apply Proposition 12.59. The conclusion is (see (13.45)) that we can construct $S^2 \in \mathscr{D}'_{(r+1)}(G)$ such that

$$bS^2 = T^1, \quad M_0\left(S^2\right) = \mathrm{O}\left((\log R)^C\right), \tag{13.71}$$

for some C independent of R.

With the same notation as before and D as in (13.46) we can go a step further and define

$$S^2\left(I_D\right) = \sum_{a=-D}^{D-1} \delta_a S^2, \quad bS^2\left(I_D\right) = T^1\left(I_D\right), \quad M_0\left(S^2\left(I_D\right)\right) = \mathrm{O}\left((\log R)^{\hat{C}}\right), \tag{13.72}$$

where $\hat{C}$ is independent of R (but does depend on $\hat{c}$ now!). This estimate follows from (13.71) by left translation as in §12.8.2 and (13.49).

13.5.7 The contradiction and the endgame in the proof

This is achieved by the same differential form ω on G as in §13.3.3:

$$\begin{aligned} \omega = \varphi(g)\,\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_r, \quad g = (X,Y); \\ X = \left(x_1,\ldots,x_n\right), \quad Y = \left(y_1,\ldots,y_r\right), \end{aligned} \tag{13.73}$$

for the same exponential coordinates as in (13.50) on $N \leftthreetimes V = G$. The only difference is now that $\dim V = \dim(A' \oplus A) = r$, hence the number of y-coordinates is r.

The differential form satisfies properties (i), (ii) and (iii) verbatim for an appropriate choice of $\varepsilon_P = 0, \pm 1$ (and all but one are 0) in (iii). The fact that there is one more coordinate y_r gives only a change of notation and this makes no difference.

There is, however, one difference and one more property that we must impose on ω.

(iv) There exists C such that

$$\|\omega\|_0 = \sup_g |\omega(g)| \leqslant C \tag{13.74}$$

for the Riemannian norm $|\omega|$. This is easy to achieve because $|\omega(g)| = C|\varphi(g)|$. To see this note that $\omega = \varphi(g)\omega_0$ with $\omega_0 = \mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_r$. The form ω_0 is the pullback of a constant form on V (see (13.50)) and $|\omega_0| = C$.

We are finally in a position to give the contradiction and complete the proof of the theorem as before.

First of all, by (13.60), (13.63), (13.64) and an appropriate choice of the ε_P of the differential form, we can guarantee for the principal term in (13.70),

$$\langle I_c S, \omega \rangle = 1 \tag{13.75}$$

and therefore by (13.70) and property (iv) of ω,

$$\left\langle S^2(I_D), \mathrm{d}\omega \right\rangle = \left\langle T^1(I_D), \omega \right\rangle = 1 + \mathrm{O}\left((\log R)^{-10}\right). \tag{13.76}$$

On the other hand by (13.72) and property (ii) of ω we have

$$\left|\left\langle S^2(I_D), \mathrm{d}\omega \right\rangle\right| \leqslant M_0\left(S^2(I_D)\right) \|\mathrm{d}\omega\|_0 = \mathrm{O}\left(R^{-c}\right) \tag{13.77}$$

for some $c > 0$ independent of R. Observe that the dependence of $\hat{C}$ on $\hat{c}$ in (13.72) here makes no difference.

Our required contradiction lies between (13.76) and (13.77) when $R \to \infty$.

13.5.8 A retrospective examination of the second basic construction

In this final subsection we shall be very informal. We shall try to explain why the way we went about the second basic construction was in some sense the only possible one. When we say this we refer to the qualitative geometric construction as presented in §9.5. Now, in the rest of this chapter the algebra will take over and the geometric ideas behind it will be hidden and lost. This, therefore, is an appropriate moment to make this retrospective.

The use of $S = S^0 \subset G_r$ as an 'opening move' is inevitable because this is the only thing that we have in hand. If we wish to make an endgame in the spirit of §13.3 we must somehow 'stretch' S out in the two directions of A. This has to be done in such a way that by the projection $\pi = G = N \leftthreetimes V \to V$, we cover some neighbourhood of $0 \in V$ and at the same time the error terms, that is, the terms that give the lateral boundary in the sense of Chapter 9, are far out.

Pretty much the only thing that we can try is to fill in S by $\hat{S}$ and then translate either from the left or the right. This gives

$$[-m,m]\cdot\hat{S} \quad \text{or} \quad \hat{S}\cdot[-m,m] \tag{13.78}$$

for m some large $m \in A$ and use left or right multiplication in G.

There are difficulties in both choices but the second choice is definitely excluded because, as pointed out in §12.8.2, the relevant norms cannot be controlled for the right multiplication $\hat{S}\cdot\delta_a$ and large $a \in A$. So it is the first choice in (13.78) that has to be made, despite the fact that we multiply on the right by the non-compactly-supported current $\hat{S}$, something that a priori causes problems (see §12.8.2). Here the twist in the proof that makes things work is this:

(i) S and A commute; therefore on the boundary $b\hat{S} = S$ at least it makes no difference whether we take left or right multiplication.
(ii) We actually do not consider (13.78) as a whole, but in §9.5 we make a long tube by glueing together slices of size 1 side by side, to look like empty food cans, and then fill each of these cans separately. This is how the idea was visualised in §9.5. Similarly, (13.78) in (13.72) looks like a long sausage cut into portions.

All in all, in the construction that we have made we have simply formalised these two ideas in a natural way.

13.6 The Proof of Theorem 12.17 under the Acyclicity Condition in the General Case

In this section we shall consider the general case of (13.1) when $G = N \leftthreetimes (A' \oplus A)$ where $A = \mathbb{R}^s$, with $s \geqslant 0$ and once more, to fix ideas, we shall assume that we are in the Abelian case unless otherwise stated. The hypothesis that we shall make throughout is the acyclicity

$$H_{r-1}(\Lambda_P) = \cdots = H_{r+s}(\Lambda_P) = 0 \tag{13.79}$$

and we shall show that this leads to a contradiction which in turn will prove the theorem.

The convention of §13.5.1 concerning the currents $S, T, U, \ldots \in \mathscr{D}'_{(r-1+p)}(G)$ of dimension $(r-1+p)$ that are denoted with an exponent p, that is, S^p, T^p, $U^p, \ldots$ will be used systematically here.

The proof in this general case is but a natural generalisation of the proof that we gave in the previous subsection. The only difficulty lies in choosing good notation without which the formulas in this section very rapidly become unreadable. Like everything else that has anything to do with homological algebra, the choice of notation is important. Without being unduly pedantic, entire sections of this chapter will be devoted to this problem of organising notation.

13.6.1 The choice of the indices

This subsection is the extension of §13.5.2 to the general case $s \geqslant 0$. The hypothesis is (13.79) and we shall repeatedly use Propositions 12.57 and 12.59. The former can only be applied under the more restrictive hypothesis of the acyclicity of Λ_P and remark (iv) in §13.3.4 here should be taken into account. With these we shall choose inductively indices

$$p_{s+1}, p_s, \ldots, p_0: \quad \text{all of these will be assumed to be} \geqslant 10^{10}(s+1), \quad (13.80)$$

and we shall start with $p_{s+1} = 10^{10}(s+1)$. Proposition 12.57 is used for these choices and we have already given a description of these indices in §12.15.3.2. Here we shall be more explicit.

Assume that p_{j+1} has been chosen for some $0 \leqslant j \leqslant s$. The choice of p_j is then made by the following condition:

> We choose p_j such that, for all closed currents $T = T^j$ (of dimension $r-1+j$ and $bT = 0$), we can solve in $S = S^{j+1}$ the following b-equation with bounds:
>
> $$bS = T, \quad M_{p_{j+1}}\left(S^{j+1}\right) \leqslant C M_{p_j}\left(T^j\right), \quad (13.81)$$
>
> where C is a constant that is independent of T. The additional condition $p_j \geqslant 10^{10}(s+1)$ can of course be imposed a posteriori.

We can work our way down on the dimensions starting in dimension $r+s$, where p_{s+1} is essentially arbitrary but large, to dimension $r-1$ where we end up with a possibly much larger p_0.

Concerning this last index p_0, let us point out that, in the last step for $j=0$, equation (13.81) will be used with the current obtained from the first basic construction in §13.2, $T = S^0$. For this compactly supported current we have

$$M_p\left(S^0\right) = \mathrm{O}\left((\log R)^C\right); \quad p \geqslant 0, \quad (13.82)$$

for some constant C that depends on p. From this it will follow that we do not need to worry about p_0.

We shall anticipate the way these indices and equations (13.81) will be used by starting from (13.82) and working our way up through the dimensions.

From this we shall conclude that the solution S^1 of (13.81) with $j = 0$ and $T = S^0$ as in (13.82) satisfies

$$M_{p_1}\left(S^1\right) = \mathrm{O}\left((\log R)^C\right) \tag{13.83}$$

for some C. And so on, step by step: if we have some closed current T^j that satisfies

$$M_{p_j}\left(T^j\right) = \mathrm{O}\left((\log R)^C\right) \tag{13.84}$$

for some $C > 0$, then in (13.81) we can solve in S^{j+1} which satisfies

$$M_{p_{j+1}}\left(S^{j+1}\right) = \mathrm{O}\left((\log R)^C\right) \tag{13.85}$$

for some other constant. And so on until we reach S^{s+1}.

This idea was described in §12.15.3.2 as 'closing the loop'.

13.6.1.1 The variant for the complex $\overline{\Lambda}_P$ We used hypothesis (13.79) and Proposition 12.59 for the choice of the indices $p_{s+1}, p_s, \dots$. We could have used a different hypothesis

$$H_{r-1}\left(\overline{\Lambda}_P\right) = \cdots = H_{r-1+s}\left(\overline{\Lambda}_P\right) = 0 \tag{13.86}$$

for the complex $\overline{\Lambda}_P$ of Remark 12.28. If we make that hypothesis then we can use Proposition 12.55 instead of Proposition 12.57 and make an analogous choice of indices $p_{s+1}, p_s, \dots$ verbatim by the same inductive condition. A point to note here is that in this variant, Proposition 12.55 is easier to prove and the intricate TVS considerations that were used in §12.15 can be bypassed. Note also that we have not exactly used Proposition 12.55 but rather the partial acyclicity version in the spirit of §12.17, so that (13.86) can be used. The reader can fill in the details, I am sure. The use of $\overline{\Lambda}_P$ rather than Λ_P^*, as in (13.22), was done here just for 'variety' (see Remark 12.56).

13.6.1.2 The finite homology Here, instead of (13.79), we shall make the finite homology assumption

$$h = \dim H_{r-1}\left(\Lambda_P\right) + \cdots + \dim H_{r+s}\left(\Lambda_P\right) < m, \tag{13.87}$$

where the integer m is assumed large enough, and explain how the choice of the indices is made under this more general assumption. The suggestion that we made to the reader in remark (iv) of §13.3.4 is very much valid again.

Since this new variant for the choice of the indices $p_{s+1}, \dots$ will not be needed immediately, the reader is advised to skip this material and come back to it when it is needed later in §13.6.4.2 and, more so, in §13.7.

Let us examine first the special case $s = 1$ where only one index $p_1 \geqslant 10$ has to be chosen. In that case we use the refinement of Proposition 12.57 as refined in §§12.16–12.17 and we can define p_1 by the following condition.

For any choice of closed currents $T_1^1, \dots, T_m^1$ (i.e. currents of dimension r) there exist real scalars $\lambda_1, \dots, \lambda_m$ such that $\sup_j |\lambda_j| = 1$ and for which we can solve with bounds the following b-equation in S^2:

$$bS^2 = \sum_j \lambda_j T_j^1; \quad M_0\left(S^2\right) \leqslant C \sum_j M_{p_1}\left(T_j^1\right), \tag{13.88}$$

for a constant C independent of the currents T_j^1.

As before in (13.39) and §13.5.2, once p_1 has been chosen the following also holds. For any choice $S_1^0, \dots, S_m^0$ of closed currents of dimension $(r-1)$ we can find scalars $\mu_1, \dots, \mu_m$ such that $\sup_i |\mu_i| = 1$ and for which we can solve the following b-equation with bounds in S^1:

$$bS^1 = \sum_i \mu_i S_i^0; \quad M_{p_1}\left(S^1\right) \leqslant C \sum_i M_p\left(S_i^0\right), \tag{13.89}$$

with p and the constant C independent of the currents S_i^0. This will suffice for the case $s = 1$ as we shall see in §13.6.4.2 below.

For the general case $s \geqslant 1$ in §13.7, however, the choice of the indices $p_{s+1}, p_s, \dots$ is slightly more elaborate. Once more we set $p_{s+1} = 10^{10}(s+1)$ and again assume that p_{j+1} has been chosen for $0 \leqslant j \leqslant s$. The analogue of the inductive step that is used to choose p_j is now given by the following property that p_j has to satisfy.

Let $(T_{\alpha,1}^j, \dots, T_{\alpha,m}^j;\ 1 \leqslant \alpha \leqslant \alpha_0)$ be $m\alpha_0$ closed currents of dimension $r - 1 + j$. Then there exist scalars $\lambda_1, \dots, \lambda_m$ such that $\sup_i |\lambda_i| = 1$ and for which we can solve the following b-equations with bounds in S_α^{j+1}:

$$bS_\alpha^{j+1} = \sum_i \lambda_i T_{\alpha,i}^j, \quad M_{p_{j+1}}\left(S_\alpha^{j+1}\right) \leqslant C \sum_i M_{p_j}\left(T_{\alpha,i}^j\right); \quad \alpha = 1, 2, \dots, \tag{13.90}$$

where the constant C is independent of the currents $T_{\alpha,i}^j$. The scalars λ_i are the same for all the $1 \leqslant \alpha \leqslant \alpha_0$ and several equations have to be solved now. The number α_0 of these equations will depend on the geometry of the situation in §13.7 and it will be fixed. The existence of such a $p_j < +\infty$ and of the coefficients λ_i follows from hypothesis (13.87) provided that m is large enough.

Exercise 13.7 The proof of this is just a refinement of the argument of §12.16. Indeed from the exercise in that section we see that if the right-hand side of (13.90) is finite for some appropriately large enough p_j, then we can define linear mappings

$$\mathbb{R}^m \ni (\lambda_1, \dots, \lambda_m) \underset{L_\alpha}{\longrightarrow} \sum_i \lambda_i T_{\alpha,i}^j \in (\operatorname{Im} \delta_r)^\perp,$$

where δ_r here is as in §12.16 (note, however, that in §12.16 the use of the letters T and S is different). Let us write $K_\alpha = L_\alpha^{-1}(\operatorname{Ker} d)^\perp \subset \mathbb{R}^m$, where again $\operatorname{Ker} d$ is as in §12.16. Hypothesis (13.87) implies that codimension $K_\alpha \leqslant h$. It follows that if m is large enough, depending on α_0 and h, these subspaces have a non-zero intersection and $0 \neq (\lambda_1, \dots, \lambda_m) \in \bigcap_\alpha K_\alpha$ will give the required scalars.

13.6.2 The Γ-free complex

With the same notation $G = N \leftthreetimes (A' \oplus A)$, $A' = \mathbb{R}^{r-1}$, $A = \mathbb{R}^s$, $s \geqslant 0$, we shall denote by $e_j = (0, \dots, 1, 0, \dots, 0)$, $1 \leqslant j \leqslant s$ the basis elements of A and by

$$\Gamma = \left(\sum_j n_j e_j;\ n_j \in \mathbb{Z} \right) \subset A; \quad \Gamma \cong \mathbb{Z}^s \tag{13.91}$$

the lattice they generate in A.

We shall use this lattice to define a complex of real vector spaces (see §12.4)

$$\begin{aligned} \mathscr{C}:\ \cdots \longleftarrow \mathscr{C}^0 \longleftarrow \mathscr{C}^1 \longleftarrow \cdots \longleftarrow \mathscr{C}^p \longleftarrow \cdots; \\ \mathscr{C}^n = 0 \text{ if } n < 0 \text{ or } n > s. \end{aligned} \tag{13.92}$$

The definition of this complex that we shall give below is not the most natural one but it is the one that best suits our purposes.

The coordinates of A in terms of the basis e_j are $(x_1, \dots, x_s) \in \mathbb{R}^s$. For every increasing multi-index $I = (1 \leqslant i_1 < i_2 < \cdots i_p \leqslant s)$ of length p, with $p = 0, \dots, s$ (for $p = 0$ we set $I = \emptyset$), we shall define the following current $e_I \in \mathscr{E}'(A) \subset \mathscr{E}'(G)$. We write $e_\emptyset = \delta_0$ for the current of dimension 0 given by the δ-mass at 0. Let $\mathrm{d}x_I = \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_p}$ be the basis of $\bigwedge T^*A$. We then define the current e_I by $e_I[a\,\mathrm{d}x_J] = 0$ if $J \neq I$ and $a \in C^\infty(\mathbb{R}^s)$. And if for such a function we write a_I for the restriction of a to the subspace spanned by the $(e_i, i \in I)$, we define

$$e_I\,[a\,\mathrm{d}x_I] = \int_{0<x_i<1} a_I\,\mathrm{d}x_{i_1} \dots \mathrm{d}x_{i_p}. \tag{13.93}$$

This current is none other than the one that we defined in §12.8.6.

These currents will now be translated by the elements of the group Γ and we shall obtain the currents $\tau_\gamma e_I = \delta_\gamma e_I$ where τ_γ is the translation $x \mapsto x + \gamma$ in A, $\gamma \in \Gamma$. These will be used as a basis over $\mathbb{R}$ to define the vector spaces

$$\mathscr{C}^p = \operatorname{Vec}\left[\delta_\gamma e_I;\ \gamma \in \Gamma,\ |I| = p\right]. \tag{13.94}$$

The above, with the usual boundary operator b for currents, define the complex $\mathscr{C}$. The e_I will be called the Γ-basis elements of $\mathscr{C}$.

It is clear that the basis of $\mathscr{C}^0$ can be identified with the group Γ which acts on the complex by translation. This complex qualifies for what, in homological algebra, one calls a Γ-free complex and a free resolution of Γ and this can be used to define the homology of the discrete group Γ (cf. Appendix G). This homology is the same as the homology of $\mathbb{R}^s/\Gamma = \mathbb{T}^s$. This aspect of things will not be relevant here.

Concerning this complex we shall fix once and for all some notation. In the same spirit as our previous convention in §13.5.1 we shall use *small* Greek letters, for example $\zeta^p, \xi^p \in \mathscr{C}^p$, to denote elements of dimension p of $\mathscr{C}$ so that ζ^p can be identified with a current in $\mathscr{D}'_{(p)}(G)$. This is not to be confused with $S^p \in \mathscr{D}'_{(r+p-1)}$ in §13.5.1. We shall also denote by $\gamma\xi = \tau_\gamma \xi$ the action of the group Γ on $\mathscr{C}$.

With this convention we shall fix some closed current $S = S^0 \in \mathscr{D}'_{(r-1)}(G)$ and using the definition given in §12.8.4 for the product in G of two currents we shall define

$$W(\xi) = \xi S; \quad \xi \in \mathscr{C}. \tag{13.95}$$

In this and in what follows, as with (13.42), in the definition of the product $T_1 \cdot T_2$ of (12.79) the dot will be suppressed when there is no confusion. No confusion shall arise either with the notation for the produit tensoriel of de Rham (1960, p. 36), that we briefly recalled in §12.8.5. The definition will depend on S, so a more accurate notation would have been W_S. In conformity with §13.5.1 we shall also denote by W^p the restriction of W to $\mathscr{C}^p$:

$$W^p \colon \mathscr{C}^p \longrightarrow \mathscr{D}'_{(r+p-1)}(G), \quad \text{that is, } W^p(\xi) \in \mathscr{D}'_{(r+p-1)}; \quad \xi \in \mathscr{C}^p. \tag{13.96}$$

When S is the current of §13.2 the relevant estimate for W is

$$M_j(W(\xi)) \leqslant \mathrm{O}((\log R)^{c_j}); \quad j \geqslant 0 \tag{13.97}$$

for constants c_j independent of R, provided that the support of $\xi \in \mathscr{C}$ stays in some fixed ball $[|x| \leqslant c]$ of $A = \mathbb{R}^s$. This is, of course, exactly as in Exercise 13.5, an immediate consequence of the commutativity between $N' \leftthreetimes A'$, which is identified with G_r (see §13.1.1) and A.

With this notation it is clear that

$$W(\gamma\xi) = \tau_\gamma W(\xi), \quad bW(\xi) = W(b\xi); \quad \gamma \in \Gamma; \tag{13.98}$$

the second relation holds because of (12.84) and because S is closed: $bS = 0$.

In our application we shall always set for S in the definition a current defined by the first basic construction as in §13.2. In that case, for $\xi = \xi^p \in \mathscr{E}'(A)$, by the commutation between G_r and A we deduce that in (13.95) we have

$$\xi S = S\xi. \tag{13.99}$$

Exercise No use of (13.99) will be made. However, it is good to keep it in mind and the proof is formal (see §12.8). Let $T_i \in \mathscr{D}'(G)$, $\operatorname{supp} T_i = E_i$, for $i = 1, 2$. Assume that E_1 and E_2 commute pointwise and at least one of the two is compact. Then, for the product of these two currents in (12.79), we have $T_1 T_2 = T_2 T_1$. For the proof let $H = G \times G$ and $p\colon H \to G$, where $p(x,y) = xy$ is the group product. Let $\theta(x,y) = (y,x)$ be the involution $\theta\colon H \to H$. We have $p \circ \theta = p$ on $E_i \times E_j$, for $i \neq j$. Now with $\mathscr{A}$ as in de Rham (1960, §13) or §12.8 we have $T_i T_j = p\mathscr{A}(T_i \times T_j)$ and also $\theta\mathscr{A}(T_i \times T_j) = \mathscr{A}(T_j \times T_i)$. Put these together and we are done.

To elaborate further go back to §12.8.5.2. The notation there was different because we wanted to stick with de Rham's notation and the reader should be careful to avoid confusion. At any rate, it follows from the discussion given there that the two products $T_1 \cdot T_2$ and $T_2 \cdot T_1$ in the group are the images under p and $p \circ \theta$ respectively of one and the same current R on $G \times G$ with support in $E_1 \times E_2$. The confusion that has to be avoided is that in de Rham, $T_1 T_2$ denotes 'le produit tensoriel' which is a double current; cf. §12.8.3.

13.6.3 The acyclicity of Λ_P and the mappings S^1, S^2

In this section we shall use once more the notation S^q. But this time it is to indicate an important mapping on the complex $\mathscr{C}$. To define these mappings the acyclicity conditions of (13.79) and (13.36) on the vanishing of some homology groups $H_{r-1}(\Lambda_P) = 0, \ldots$ will be needed and we shall also need and use freely the indices $p_{s+1}, p_s, \ldots$ of §13.6.1. When they can be defined, these mappings are

$$S^p\colon \mathscr{C}^{p-1} \longrightarrow \mathscr{D}'_{(r+p-1)}, \quad S^p\left(\xi^{p-1}\right) \in \mathscr{D}'_{(r+p-1)}; \qquad p = 1, \ldots, s+1,\ \xi^{p-1} \in \mathscr{C}^{p-1}, \tag{13.100}$$

and we shall assume that $s \geqslant 1$. We shall always denote $S^0 = S \in \mathscr{E}'(N' \leftthreetimes A')$ the current defined by the first basic construction in §13.2. As far as the dimensions are concerned, the above notation is consistent with our convention in §13.5.1 but there should be no confusion between these mappings and the currents of dimension $r + p - 1$.

The first two mappings S^1 and S^2 are easy to define. Let $s \geqslant 1$ and let $\hat{S} \in \mathscr{D}'_{(r)}(G)$ be as in §13.3.2 where we had

$$b\hat{S} = S, \quad M_0(\hat{S}) \leqslant CM_p(S), \tag{13.101}$$

where C and p are independent of the parameter R. This can be defined under the acyclicity condition (13.79) and, to be consistent with the notation of

§13.5.1, it is preferable to write $S = S^0$, $\hat{S} = \hat{S}^1$ and to rewrite (13.101) as

$$M_{p_1}(\hat{S}^1) \leqslant C M_{p_0}(S^0), \tag{13.102}$$

where now $\ldots, p_1, p_0$ are as in §13.6.1, so that this is essentially a rewriting of (13.39). Be that as it may, the first mapping (13.100) is then defined

$$S^1(\delta) = \delta\hat{S}; \quad \delta = \delta^0 \in \mathscr{C}^0. \tag{13.103}$$

This is the product of currents and is well defined by translating $\hat{S}$ on the *left* by elements of Γ.

With this definition and W as in the previous subsection, we have (see (12.84))

$$b\left(\xi\hat{S}\right) = S^1(b\xi) - W^1(\xi) = T^1(\xi); \quad \xi \in \mathscr{C}^1. \tag{13.104}$$

Notice that we have already encountered this notation in (13.44) but there it was given with the opposite sign. Not to worry, as no confusion will arise. Here we represent the first steps of a sequence of 'creatures' $T^1, T^2, \ldots$ that we shall elaborate on in the rest of this chapter. But before that, we shall immediately give the next step of the construction. For this we consider the e_j of the basis of Γ or equivalently of the Γ-free basis of $\mathscr{C}^1$ in (13.94), and under hypothesis (13.79) and the definition of $p_2 \geqslant 10^{10}(s+1)$ of §13.6.1 we can construct $S^2(e_j)$ that satisfies

$$bS^2(\xi) = S^1(b\xi) - W^1(\xi); \quad \xi = e_j, \tag{13.105}$$

$$M_{p_2}(S^2(e_j)) \leqslant C M_{p_1}(T^1(e_j)); \quad 1 \leqslant j \leqslant s, \tag{13.106}$$

where C is independent of $R > 0$ and of the particular choice of $S = S^0$. This can be done because by (13.104) it follows that $bT^1 = 0$.

Now we shall use the Γ-free nature of the complex $\mathscr{C}$ and of the Γ-basis (13.94) to extend S^2 to a unique Γ-mapping

$$S^2 \colon \mathscr{C}^1 \longrightarrow \mathscr{D}'_{(r+1)}; \quad S^2(\xi^1) \in \mathscr{D}'_{(r+1)}. \tag{13.107}$$

The mapping $S^2(\xi)$ satisfies the estimate $\mathrm{O}(\log R)^C$ of (13.83) and the dependence of the estimate on ξ will be one of the issues of the proof (see (13.114)). By Γ-mapping we mean here and throughout that the mapping is linear and that it commutes with the Γ-action on the two spaces. In algebraic terminology we have a Γ-module mapping. It also satisfies (13.105) for all $\xi \in \mathscr{C}^1$. To summarise,

$$bS^2(\xi) = S^1(b\xi) - W^1(\xi), \quad S^2(\gamma\xi) = \tau_\gamma S^2(\xi) = \delta_\gamma S^2(\xi); \\ \gamma \in \Gamma,\ \xi \in \mathscr{C}^1. \tag{13.108}$$

This mapping depends on the original current $S = S^0$ and on the parameter R. We should perhaps write in (13.108) more accurately

$$bS^2_S(\xi) = S^1(b\xi) - W^1_S(\xi); \quad \xi \in \mathscr{C}^1. \tag{13.109}$$

But of course these mappings also depend on the particular choice of $\hat{S}$ in (13.101) that we used to define S^1.

13.6.4 The first application of the mapping S^2

13.6.4.1 The acyclicity theorem (Theorem 12.17) when $s = 1$ As a first illustration of the mapping S^2 that we defined in the previous subsection we shall assume that $G = N \leftthreetimes (A' \oplus A)$ satisfies

$$A = \mathbb{R}, \quad \text{i.e. } s = 1, \quad H_{r-1}(\Lambda_P) = H_r(\Lambda_P) = H_{r+1}(\Lambda_P) = 0, \tag{13.110}$$

that is, that we are in the set-up of §13.5. Here we shall simply rewrite some of the proof that we gave in §13.5 using these new mappings and notation. This will give a good illustration of their use in proofs.

The current $S = S^0$ is the one that comes from the first basic construction as explained in the previous subsection. We then choose $\hat{S} = \hat{S}^1$ as in (13.101) and the two mappings S^1, S^2 of the previous subsection are then defined. These depend on the basic parameter R and the choice of S and $\hat{S}$.

Let $D = [(\log R)^{\hat{c}}]$ be as in (13.46) where $\hat{c} > 0$ is to be chosen later (it will be chosen last as explained in (13.46)). We then define

$$\xi_D = \sum_{n=-D}^{D-1} \delta_n e_1 \in \mathscr{C}^1 \tag{13.111}$$

for e_1, the unique Γ-basis element of $\mathscr{C}^1$ in (13.94). With the notation and the orientation specified in §13.5.3 we have

$$\xi_D = [-D, D], \quad b\xi_D = \delta_D - \delta_{-D} \tag{13.112}$$

and now our defining relation for S^2 gives the key identity (13.72) that is used in the proof in §13.5.5. This now reads (cf. (13.105) and (13.106)

$$bS^2(\xi_D) = S^1(b\xi_D) - W^1(\xi_D) = (\delta_D - \delta_{-D})\hat{S} - \xi_D S. \tag{13.113}$$

The estimate in §13.5.6 here reads

$$\begin{aligned} M_0\left(S^2(\xi_D)\right) &\leqslant CM_{p_1}\left(T^1(\xi_D)\right) \\ &\leqslant C\left(M_{p_1}\left(\delta_D\hat{S}\right) + M_{p_1}\left(\delta_{-D}\hat{S}\right) + \sum_{n=-D}^{D-1} M_{p_1}\left(\delta_n e_1 S\right)\right) \\ &= \mathrm{O}(\log R)^{\hat{C}}, \end{aligned} \tag{13.114}$$

where we use the notation (13.104) and for the last O-estimate we use again §12.8.2 as in (13.72), (13.48), (13.97). These are the ingredients that were used in §13.5.7 to finish the proof for the special case $s = 1$.

13.6.4.2 The case of finite homology when $s = 1$ This is a special case of §13.7 and the reader could defer reading this until later. Nonetheless this case illustrates well several of the ideas that we have already used. As before we shall assume that $A = \mathbb{R}$ but now we shall make the more general assumption that

$$\dim H_{r-1}(\Lambda_P) + \dim H_r(\Lambda_P) + \dim H_{r+1}(\Lambda_P) < m < +\infty, \tag{13.115}$$

for some integer m

Under these conditions we shall produce a contradiction and thus give the proof of the finite homology part of Theorem 12.17 for this special case $s = 1$. This will illustrate the use of the bouquets of currents of §13.4 as well as the new twist of double bouquets that has to be used in this proof.

First of all, unless we introduce a bouquet of currents

$$\mathscr{B} = (S_1, \ldots, S_m) \tag{13.116}$$

as in (13.31), we cannot take off because, since we do not have the acyclicity (13.79), we cannot solve $b\hat{S} = S$ and define the first mapping S^1 of (13.103). On the other hand, the index p_1 and the corresponding scalars λ_j of §13.6.1.2 and Proposition 12.59 can be used. What we obtain is what follows (cf. 13.6.1.2).

Once the bouquet $\mathscr{B}$ has been introduced we can find scalars $\lambda_1, \ldots, \lambda_m$ such that $\sup_j |\lambda_j| = 1$ for which we can solve with bounds the following b-equation in $\hat{S}$,

$$b\hat{S} = \sum \lambda_j S_j, \quad M_{p_1}\left(\hat{S}\right) \leqslant C \sum_j M_{p_0}(S_j), \tag{13.117}$$

for some C and p_0 independent of R.

This $\hat{S}$ can now be used to define $S^1(\delta) = \delta\hat{S}$ as in (13.103) and we set as before

$$T^1(\xi) = S^1(b\xi) - W(\xi) = b\left(\xi\hat{S}\right); \quad \xi \in \mathscr{C}^1. \tag{13.118}$$

These are closed currents but we are faced with the same problem as before because (13.17) is not assumed to hold and we cannot a priori solve the b-equation $bS^2(e_1) = T^1(e_1)$ as we did in (13.105) *with bounds*. Here we emphasise ‘with bounds’ because otherwise this equation can of course be solved by some current.

It is here that the notion of a double bouquet comes to our rescue. Instead of one bouquet $\mathscr{B}$ in (13.116) we introduce a collection of m-bouquets

$$\begin{gathered}\mathscr{B}^2 = (\mathscr{B}_1, \dots, \mathscr{B}_m); \quad \mathscr{B}_j = (S_{j,1}, \dots, S_{j,m}), \\ [\mathscr{B}^2] = (S_{j,i};\ i,j = 1,\dots,m \text{ is a bouquet in the sense of §13.4}).\end{gathered} \tag{13.119}$$

Each of these bouquets $\mathscr{B}_j$, with $1 \leqslant j \leqslant m$, gives rise to coefficients and currents

$$\lambda_{j,1}, \dots, \lambda_{j,m} \in \mathbb{R}, \quad \sup_i |\lambda_{j,i}| = 1, \quad S_j = \sum_i \lambda_{j,i} S_{j,i}; \quad j \geqslant 1, \tag{13.120}$$

for which we can solve with bounds the corresponding b-equations

$$b\hat{S}^1_j = S_j; \quad M_{p_1}\left(\hat{S}^1_j\right) \leqslant C \sum_i M_{p_0}(S_{ji}),\ j \geqslant 1, \tag{13.121}$$

for C and p_0 that are independent of the parameter R, and we recall that $\dim S_j = r-1$. When confusion does not arise, for typographical reasons with these multiple indices, the commas will be suppressed; for example, S_{ji} rather than $S_{j,i}$

These $\hat{S}^1_j$ can in turn be used to define the corresponding mappings $S^1_j(\delta) = \delta\hat{S}^1_j$ of (13.103). So now, instead of having a closed current $T^1(\xi)$ as in (13.104), we have m such currents $T^1_1(\xi), \dots, T^1_m(\xi)$.

Having these currents we shall use the definition of p_2 in §13.6.1.2. We conclude that we can find scalars $\mu_1, \dots, \mu_m$ such that $\sup_j |\mu_j| = 1$ for which we can solve with bounds the following b-equation when $\xi = e_1$ is the element in the Γ-basis of $\mathscr{C}^1$ (13.94):

$$\begin{aligned} bS^2(\xi) &= \sum_j \mu_j T^1_j(\xi) = \sum_j \mu_j \left(S^1_j(b\xi) - W^1_{S_j}(\xi)\right) \\ &= \sum_j \mu_j S^1_j(b\xi) - \sum_{i,j} \mu_j \lambda_{ji} \xi S_{ji}; \quad \xi = e_1, \end{aligned} \tag{13.122}$$

because W_X is linear in X. Since $\mathscr{C}^1$ is a free Γ-module we can extend S^2 to a Γ-linear mapping that satisfies (13.122) for all the $\xi \in \mathscr{C}^1$.

Furthermore, the corresponding bound

$$M_0\left(S^2(e_1)\right) \leqslant CM_{p_2}\left(S^2(e_1)\right) \leqslant C \sum_j M_{p_1}\left(T^1_j(e_1)\right) \tag{13.123}$$

holds for constants that are independent of R and from this, by the definition of T^1, it follows that

$$M_0\left(S^2(e_1)\right) = \mathrm{O}\left((\log R)^C\right) \tag{13.124}$$

for a similar constant C. We now apply (13.122) to ξ_D of (13.111) with D as in (13.46) and we have

$$bS^2(\xi_D) = \sum_j \mu_j S_j^1(b\xi_D) - \sum_{i,j} \mu_j \lambda_{ji} \xi_D S_{ji}. \tag{13.125}$$

That this can now be used to give a contradiction in the case when $s = 1$, exactly as was done in §13.5, is but an exercise. We shall not give the details here because the general case $s \geqslant 1$ will be treated in §13.7.

Exercise 13.8 The reader is urged to complete the details for themselves here. The pivot is that the coefficients $\mu_j \lambda_{ji}$ of S_{ij} in (13.125) are 'triangular' in the sense that we can first choose j such that $|\mu_j| = 1$ and then i so that $|\lambda_{ji}| = 1$. For these specific values of the indices, the contribution to the principal term is then given by $\mu_j \lambda_{ji} \xi_D S_{ji}$ and the error term can be estimated by (13.69) because $b\xi_D$ is far out as in §13.5.5.

13.6.5 The mappings S^q; $q \geqslant 1$

Here we shall assume that $s \geqslant 2$ and go back to the original standing hypothesis of acyclicity, (13.79). The mappings S^1 and S^2 were defined in §13.6.3. The general mapping S^q, with $1 \leqslant q \leqslant s+1$, will now be constructed under hypothesis (13.79). This will be done inductively for $q = 1, 2, \ldots$.

13.6.5.1 The construction of S^3 For the convenience of the reader we shall start with the next case $q = 3$ and we shall assume that the current $S = S^0$ of the first basic construction has been fixed and it depends of course on the parameter $R > 0$. We shall use the mappings S^1, S^2 and the mapping $W = W_S$ that we defined in §13.6.3. From the definition of S^2 it follows that

$$\begin{aligned} b\left(S^2(b\xi) + W\xi\right) &= bS^2(\zeta) + b(W\xi) \\ &= \left(S^1(b\zeta) - W^1(\zeta)\right) + W^1 b\xi = 0; \\ &\quad \zeta = b\xi,\ \xi \in \mathscr{C}^2, \end{aligned} \tag{13.126}$$

and therefore the current $T^2(\xi) = S^2(b\xi) + W(\xi)$ is closed. As a consequence, by the definition of the indices $p_{s+1}, p_s, \ldots$ in §13.6.1 and the acyclicity hypothesis (13.79), we can solve the following b-equations with bounds.

For every $e_{i,j}$, with $i < j, i, j = 1, \ldots, s$, the basis element of $\mathscr{C}^2$ as in (13.94), we can solve the equations

$$\begin{aligned} &bS^3(\xi) = S^2(b\xi) + W^2(\xi) = T^2(\xi); \quad \xi = e_{ij}, \\ &M_{p_3}\left(S^3(e_{ij})\right) \leqslant CM_{p_2}\left(T^2(e_{ij})\right); \quad i < j, \end{aligned} \tag{13.127}$$

for some constant C independent of R. We can now use the Γ-basis of $\mathscr{C}^2$ in (13.94) to extend the definition of S^3 for all $\xi \in \mathscr{C}^2$ so that it is a Γ-mapping. This completes the definition of S^3.

13.6.5.2 The definition of S^q in the general case Here $s \geqslant 1$ and, as before, both $S = S^0$ from the first basic construction and the corresponding mapping W (13.95) will be fixed. The indices $p_{s+1}, p_s, \ldots$ from §13.6.1 will also be assumed fixed. Let $2 \leqslant q \leqslant s$ be given. We shall assume that the Γ-mappings $S^q, S^{q-1}, \ldots$ have been constructed and that they satisfy

$$bS^q(\xi) = S^{q-1}(b\xi) + (-1)^{q-1}W\xi; \quad \xi \in \mathscr{C}^{q-1},$$
$$M_{p_q}\left(S^q(e_I)\right) \leqslant (\log R)^C; \quad R \geqslant C, \tag{13.128}$$

where C is independent of R and where e_I are the Γ-basis elements of $\mathscr{C}^{q-1}$ as I runs through the corresponding multi-indices of length $(q-1)$. From this we have

$$bS^q(b\xi) = (-1)^{q-1}Wb\xi; \quad \xi \in \mathscr{C}^q. \tag{13.129}$$

Therefore we have (see (13.98))

$$bT^q(\xi) = 0; \quad T^q(\xi) = S^q(b\xi) + (-1)^q W^q \xi,\ \xi \in \mathscr{C}^q. \tag{13.130}$$

By the definition of the index p_{q+1} and our acyclicity hypothesis (13.79) we can therefore solve in S^{q+1} the following b-equations with bounds:

$$bS^{q+1}(\xi) = S^q(b\xi) + (-1)^q W\xi = T^q(\xi); \quad \xi = e_I \in \mathscr{C}^q \tag{13.131}$$
$$M_{p_{q+1}}\left(S^{q+1}(e_I)\right) \leqslant CM_{p_q}\left(T^q(e_I)\right) \leqslant C(\log R)^C; \quad |I| = q,\ R \geqslant C. \tag{13.132}$$

This is done for all the basis elements e_I, (13.94), of $\mathscr{C}^q$ with a multi-index I of length $|I| = q$ and with constants that are independent of R. The final estimate in (13.132) is a consequence of (13.128) and of (13.97). From this we can use the Γ-free nature of the complex $\mathscr{C}$ and we can extend S^{q+1} to a Γ-linear mapping defined for all $\xi \in \mathscr{C}^q$ that satisfy (13.131).

This completes the inductive step and we can define thus all the mappings $S^1, S^2, \ldots, S^{s+1}$. This was referred to in §12.15 as ‘closing the loop’.

An explanation of the construction. This construction is quite involved. In particular, we have to use all the dimensions $r-1, r, \ldots$ of (3.79). The following offers some explanation of why this is so.

The reason that we have to proceed by induction on the dimensions is that once some current that has already been constructed spills out of $N' \leftthreetimes A'$ we cannot multiply it by some current of A and estimate the product in a useful way (cf. §8.1). The only hope is to translate that current on the left and use §12.8.2. This is what was done in (13.67), (13.72) and will be done again in the next subsection and in §13.7.

13.6.6 The end of the proof in the general case

Once the mapping S^{s+1} has been defined under the acyclicity condition (13.94), a contradiction can be obtained exactly as we did for the special case $s = 1$ in §13.5.7.

To see this we define

$$\begin{aligned}\xi_D &= \xi_D^s \\ &= \sum_\gamma [\gamma e;\ \gamma = (n_1, \dots, n_s) \in \Gamma,\ -D \leqslant n_j \leqslant D-1,\ 1 \leqslant j \leqslant s]. \end{aligned} \quad (13.133)$$

This is $[-D,D]^s$, the chain with the same orientation on $\mathbb{R}^s$ that gives $e \in \mathscr{C}^s$ the unique basis element $e = e_{[1,\dots,s]}$ of (13.46). Here, as before, $D = [(\log R)^{\hat{c}}]$ exactly as in (13.46) where the new constant $\hat{c} > 0$ will be chosen at the end.

We shall then apply (13.131) to the current ξ_D and we obtain

$$bS^{s+1}(\xi_D) = (-1)^s W^s \xi_D + S^s(b\xi_D) = T^s(\xi_D) \quad (13.134)$$

and now we have to examine separately,

$$\text{principal term} = (-1)^s W \xi_D, \quad \text{error term} = S^s(b\xi_D). \quad (13.135)$$

With these formulas one sees the *algebraic reasons* behind the definitions that we have made. For instance, we use the Γ-module property of S^s and we have

$$\sum_\gamma \gamma S^s(be) = S^s\left(\sum \gamma be\right) = \text{error term}.$$

A special case when $s = 1$ is the *telescopic sum* of §13.5.3.

The way to go from here is verbatim what we did in §13.5 for $s = 1$. It could have been left as an exercise for the reader to write this proof out. For convenience, however, we shall first recall the notation as adapted here in the general case $s \geqslant 1$.

In fact, the only difference is that in $G = N \leftthreetimes V$ with $V = A' \oplus A$, now A is $A \cong \mathbb{R}^s$, with $s \geqslant 1$. The notation of §13.5.4 will therefore be adapted. For $g = (X,Y) \in G$, we have $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_{r+s-1}) = (v_1, \dots, v_{r+s-1})$ for the exponential coordinates for N and the Euclidean coordinates for V, respectively, as in §13.5.4, with

$$v = \left(v', v_r, \dots, v_{r+s-1}\right) \in A' \oplus A, \quad (v_r, \dots, v_{r+s-1}) \in A \quad (13.136)$$

as in §13.5.4.1. The norms and the corresponding balls in V,

$$|v| = \sup_i |v_i|\,, \quad [|v| < c] = V_c \subset V, \tag{13.137}$$

are defined as in (13.50) for some appropriately small $c > 0$. Using the canonical projection $\pi\colon G \to V$ we define the slice $N_c = \pi^{-1}(V_c) \subset G$ as in (13.50) and denote by χ the characteristic function of that slice. This together with (13.130) gives the analogue of formula (13.70) which now becomes

$$\begin{aligned} \chi T^s(\xi_D) &= (\text{principal term}) + (\text{error}),\\ \text{principal term} &= (-1)^s \chi W \xi_D = (-1)^s \chi \xi_D S^0, \\ \text{error} &= \chi S^s(b\xi_D)\,. \end{aligned} \tag{13.138}$$

From here the proof unfolds exactly as in the special case $s = 1$ in §13.5, even with the same notation, provided that we bear in mind that in §13.5 the one-dimensional current ξ_D was denoted by I_D. What was done in §13.5, and what has to be done here again in the general case, is to take the following three steps.

(a) **Localisation of the principal term**

$$\chi(\xi_D S) = \sum_P \eta_P \delta_P [V_c] \tag{13.139}$$

with summation on the vertices and the same notation as (13.59), (13.60) (we recall that $P \in \partial_0 \square^r_R$, δ_P are Dirac masses, $\eta_P = \pm 1$). Here $[V_c] \in \mathscr{E}'(V)$ is an $(r+s-1)$-dimensional current. The same comments about the orientations and of course the finer localisation as in §13.5.4.1 apply here again.

(b) **Estimate of the error term and the formula** We have, again with a slight abuse of notation, the analogue of formula (13.70):

$$\chi T^s(\xi_D) = \sum_P \eta_P \delta_P [V_c] + \mathrm{O}\left((\log R)^{-10}\right), \tag{13.140}$$

where the O is interpreted in terms of the M_0-norm as in (13.69) and (13.70).

It is the choice of $\hat{c}$ in the definition of D and the estimate of the error term in (13.132) that allows us to write (13.140). Here in the error term $\chi S^s(b\xi_D)$ the current $b\xi_D$ is supported on ∂I^s_D the boundary of the cube $[-D,D]^s$ and is not just the two-point $\delta_{\pm D}$ of the case $s = 1$. This makes no difference because the only thing that counts is that [distance $(0, \partial I^s_D)$] in V is $\sim (\log R)^{\hat{c}}$. From this the same estimate on the error term follows.

Exercise 13.9 We shall outline below the proof of this and the reader should fill in the details. Notice that the same argument in a more intricate situation will be spelled out in detail in §13.7.4.3 below.

Write $b\xi_D = \sum_\alpha \zeta_\alpha$, for $\zeta_\alpha \in \mathscr{C}^{s-1}$ which are in fact of the form $\zeta_\alpha = \pm\delta_\gamma e_J$, with $\gamma \in \Gamma$ and e_J the basis elements in (13.94). This is the natural decomposition of the boundary of the cube ∂I_D^s into subcubes of size 1 and dimension $(s-1)$. The translations γ involved are all $|\gamma| = D \sim (\log R)^{\hat{c}}$. The total number of these subcubes is $\approx(s-1)$-dimensional measure of $\partial I_D^s \approx D^{s-1} \approx (\log R)^{\hat{c}(s-1)}$.

By the argument of §13.5.5 and §12.8 the contribution of each ζ_α in the slice N_c is

$$M_0\big(\chi S^s(\delta_\gamma e_J)\big) \leqslant C|\gamma|^{-p_s} M_{p_s}\big(S^s(e_J)\big) \leqslant (\log R)^{-p_s\hat{c}}(\log R)^C; \quad R \geqslant C, \tag{13.141}$$

where the important point is that the constants C are independent of the choice of the constant $\hat{c}$ because $\hat{c}$ is the last constant that was chosen. The index p_s is independent of $\hat{c}$ also. Furthermore, $p_s \gg s$ in §13.6.1. This is something we did not have to worry about when $s = 1$. Putting all these estimates together we obtain the required estimate for the error term

$$M_0\big(\chi S^s(\xi_D)\big) \leqslant C(\log R)^{(s-1-p_s)\hat{c}}(\log R)^C = \mathrm{O}\big((\log R)^{-10}\big), \tag{13.142}$$

where for the last O-estimate $\hat{c}$ has to be chosen sufficiently large. It is at this point that we finally make the choice of $\hat{c}$ in (13.46). This completes the proof of (13.140).

(c) **Final step in the proof** With (a) and the estimate (b), the contradiction that is needed for the proof of the theorem is produced by the same differential form as §13.5.7,

$$\omega = \varphi(g)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{r+s-1}; \quad \text{for } y_1, \ldots, y_{r+s-1} \text{ the coordinates of } V, \tag{13.143}$$

where ω satisfies the same properties (i), (ii) and (iii) of §13.3.3 and property (iv) of §13.5.7. Furthermore, the coefficients ε_P of property (iii) will again be chosen so that

$$\langle \xi_D S, \omega \rangle = (-1)^s \tag{13.144}$$

for the principal term in (13.140) and therefore

$$\big\langle S^{s+1}(\xi_D), \mathrm{d}\omega \big\rangle = \big\langle T^s(\xi_D), \omega \big\rangle = 1 + \mathrm{O}(\log R)^{-10}. \tag{13.145}$$

This follows from (13.140), (13.144) and property (iv) of ω. These are the analogues of (13.75), (13.76). The required contradiction is thus obtained from this and

$$\left|\left\langle S^{s+1}(\xi_D), \mathrm{d}\omega\right\rangle\right| \leqslant M_0\left(S^{s+1}(\xi_D)\right) \|\mathrm{d}\omega\|_0 = \mathrm{O}\left(R^{-c}\right) \tag{13.146}$$

which is the exact analogue of (13.77) and where §12.8.2 is used again to estimate the first factor on the right-hand side.

Adapting the above for the Heisenberg alternative will be left as an exercise.

13.7 The Use of Bouquets and the Proof of Theorem 12.17 under the Finite Homology Condition

In this section the group will be as in the previous section, but instead of (13.79) the standing hypothesis will be that

$$h = \dim H_{r-1}(\Lambda_P) + \cdots + \dim H_{r+s}(\Lambda_P) < +\infty \tag{13.147}$$

and as in the previous section we shall obtain a contradiction that will give the theorem.

This has already been examined for $s = 0, 1$ in §§13.4 and 13.5. In both cases $s = 0, 1$ this was done by generalising the mappings S^1, S^2 of the previous section. In this section we shall see how these mappings S^q are generalised for $1 \leqslant q \leqslant s+1$ and obtain the required contradiction from these.

Bouquets of currents were used for this already when $s = 1$ in §13.6.4.2. Nonetheless, as we shall see, for $s = 0, 1$ the construction was special in a very particular way. To explain this problem as clearly as possible we shall recapitulate in §13.7.1 the constructions of S^1 and S^2, but this time for a general $s \geqslant 1$. Once this is well understood, the way the general construction of S^q has to be made should be clear. But to write the details down for the general S^q a well-chosen set of notation is needed. This gives rise to a non-trivial exercise in organising notation that will be carried out in §13.7.3 below.

13.7.1 The construction of S^1, S^2 for $s \geqslant 1$

The notation $G = N \leftthreetimes (A' \oplus A)$, $A' = \mathbb{R}^{r-1}$, $A = \mathbb{R}^s$ and all the other notation on the groups is as in §13.1.1 and once more we fix ideas the reader should assume are in the Abelian alternative. Also, the notation on the complex $\mathscr{C}$ is as in §13.6.2. The standing hypothesis here will be (13.147) and we shall fix m, a very large integer depending on h and s, $m = m(h,s) \gg h$, that will be chosen later.

The definition of S^1, as we saw in §13.6.4.2, necessitates the use of the bouquets of currents $\mathscr{B} = (S_1, \ldots, S_m)$ of §13.4. Once the bouquet has been fixed we shall specify scalars

$$\lambda_1,\ldots,\lambda_m \in \mathbb{R}, \quad \sup_i |\lambda_i| = 1 \tag{13.148}$$

for which we can solve the following b-equation with bounds

$$b\hat{S} = \sum_i \lambda_i S_i = S; \quad M_{p_1}\left(\hat{S}\right) \leqslant C \sum M_{p_0}(S_i) = \mathrm{O}\left((\log R)^C\right), \tag{13.149}$$

where p_1, p_0 are the indices chosen in §13.6.1.2 and C is independent of R.

We shall use the notation $I = [0,1]$ and δ_a for the currents of §§12.8.2–12.8.6, and the notation for the mapping W and the complex $\mathscr{C}$ as in §13.6.2.

For $e_1 = I = [0,1]$, a basis element of $\mathscr{C}^1$, we then have (see (12.84))

$$b\left(I\hat{S}\right) = (\delta_1 - \delta_0)\hat{S} - IS = S^1(be_1) - W^1 e_1 = T^1(e_1), \tag{13.150}$$

where we set $S^1(\xi^0) = \xi^0 S$ with $\xi^0 \in \mathscr{C}^0$.

A similar definition is given for each Γ-basis element $e_i \in \mathscr{C}^1$ of (13.46) and we have

$$bT^1(e_i) = 0; \quad T^1(e_i) = S^1(be_i) - We_i = (be_i)\hat{S} - e_i S; \quad i = 1,\ldots,s. \tag{13.151}$$

From (13.151) it does not follow that we can solve the equation $bS^2(\xi) = T^1(\xi)$, where $\xi \in \mathscr{C}^1$, with bounds because we do not have a priori the acyclicity condition (13.79). For this reason we have to start not with just one bouquet of currents but with several bouquets:

$$\mathscr{B}^2 = (\mathscr{B}_1,\ldots,\mathscr{B}_m), \quad \mathscr{B}_j = (S_{j,1},\ldots,S_{j,m}) \tag{13.152}$$

with the same large m. We shall choose the above bouquets so that the union $[\mathscr{B}^2] = (S_{j,i};\ 1 \leqslant i,j \leqslant m)$ is also a bouquet in the sense of §13.4. We call $\mathscr{B}^2$ a double bouquet. Using this we constructed S^2 in §13.6.4.2 when $s=1$, and for this, together with coefficients λ as in (13.148) that correspond to each bouquet $\mathscr{B}_j$, a new choice of coefficients $\mu_1,\ldots,$ has to be made. In generalising this construction when $s \geqslant 1$ this has to be made separately for each basis element $e_1,\ldots$ of $\mathscr{C}^1$. Care is therefore needed in applying this argument.

Let us denote by $T_j^1(e_i)$ the $T^1(e_i)$ defined from the jth bouquet $\mathscr{B}_j$ by some set of coefficients. More explicitly, let the coefficients

$$\lambda_{j,1},\ldots,\lambda_{j,m}, \quad \sup_i \left|\lambda_{ji}\right| = 1; \quad 1 \leqslant j \leqslant m \tag{13.153}$$

be as in (13.148). We solve first as in (13.149) to obtain $\hat{S}_j$ and then define

$$\begin{aligned} b\hat{S}_j = \sum_i \lambda_{ji} S_{ji} = S_j, \quad S_j^1(\delta) = \delta \hat{S}_j; \quad \delta \in \mathscr{C}^0, \\ W_j(\xi) = W_{S_j}(\xi) = \xi S_j; \quad \xi \in \mathscr{C}^1, \\ T_j^1(e_i) = S_j^1(be_i) - W_j(e_i). \end{aligned} \tag{13.154}$$

We obtain thus s sets of closed currents

$$\left(T_1^1(e_i), \dots, T_m^1(e_i)\right); \quad 1 \leqslant i \leqslant s. \tag{13.155}$$

By Exercise 13.7, *common* coefficients $\mu_1, \dots, \mu_m$ can now be found for all these different sets provided that m is large enough. More precisely, there exist scalars

$$\mu_1, \dots, \mu_m, \quad \sup_j |\mu_j| = 1, \tag{13.156}$$

such that we can solve with bounds the following s equations in S^2:

$$\begin{aligned} bS^2(\xi) = \sum_j \mu_j T_j^1(\xi); \quad \xi = e_i, \\ M_{p_2}\left(S^2(e_i)\right) \leqslant C \sum_j M_{p_1}\left(T_j^1(e_i)\right); \quad 1 \leqslant i \leqslant s, \end{aligned} \tag{13.157}$$

where p_2, p_1 are as in §13.6.1.2 and C is independent of R. We can then use Γ-linearity to extend (13.157) to all $\xi \in \mathscr{C}^1$ and thus complete the definition of a Γ-linear mapping S^2.

This definition depends of course on the double bouquet $\mathscr{B}^2$ of (13.152) and on the choice of coefficients λ_{ji}, μ_j and also on the particular solutions chosen for the b-equations of (13.154) and (13.157). In the notation it will be convenient to suppress this additional information, which will be implicit, and simply denote by $S^2_{\mathscr{B}^2}(\xi)$, where $\xi \in \mathscr{C}^1$, the Γ-linear mapping so constructed. We shall also denote

$$W^1_{\mathscr{B}^2}(\xi) = \sum_{i,j} \mu_j \lambda_{ji} \xi S_{ji} = \sum \mu_j W_j(\xi); \quad \xi \in \mathscr{C}^1 \tag{13.158}$$

and then the defining equation for S^2 is

$$bS^2_{\mathscr{B}^2}(\xi) = \sum \mu_j S_j^1(b\xi) - W^1_{\mathscr{B}^2}(\xi); \quad \xi \in \mathscr{C}^1. \tag{13.159}$$

From this definition it follows that

$$bT^2(\xi) = 0; \quad T^2_{\mathscr{B}^2} = T^2(\xi) = S^2_{\mathscr{B}^2}(b\xi) + W^2_{\mathscr{B}^2}(\xi),\ \xi \in \mathscr{C}^2. \tag{13.160}$$

To see this write

$$bS^2_{\mathscr{B}^2}(\zeta) = \sum \mu_j S^1_j(b\zeta) - W^1_{\mathscr{B}^2}(\zeta) = -W^1_{\mathscr{B}^2}(b\xi); \quad \zeta = b\xi,\ \xi \in \mathscr{C}^2 \tag{13.161}$$

and thus (13.160) follows from $bW(\xi) = W(b\xi)$ of §13.6.2.

Remark 13.10 It is a fact that the coefficients μ_j are common for all the basis elements e_i which allows us to factor the coefficients μ_j out and write (13.158) for *all* $\xi \in \mathscr{C}^1$.

Using Exercise 13.7 we can in fact choose common coefficients (i.e. independent of the first index) $\lambda_{j,i} = \lambda_i$. But for this to work out in (13.152), the number of bouquets $\mathscr{B}_1, \ldots, \mathscr{B}_{m_1}$ has to be chosen first, and afterwards, the number of currents, say m_0, in each bouquet is chosen in terms of m_1. We pick up some obvious notational advantages in this approach but the global understanding of the proof is then loaded with this extra 'twist' and this does not help matters. As a consequence, we shall avoid this.

13.7.2 The construction of S^3 and the proof of the theorem when $s = 2$

Before we go on to give the general definition of S^q under the finite homology condition (13.147) we shall define S^3 in the special case $s = 2$. This will allow us to obtain the required contradiction and prove the theorem in that special case. It will also illustrate the idea of the proof well.

To define this S^3 we only need to solve one b-equation $bS^3(e) = T^2(e)$ with bounds because for $s = 2$ there is only one Γ-basis element $e = e_{12}$ for $\mathscr{C}^2$. Even that of course a priori cannot be done since we do not have the acyclicity (13.79). To cope with this difficulty we shall need to consider several double bouquets $\mathscr{B}^2_k$ as in (13.152):

$$\begin{aligned} \mathscr{B}^3 &= \left(\mathscr{B}^2_1, \ldots, \mathscr{B}^2_m\right) = \text{multiple (triple) bouquet,} \\ \mathscr{B}^2_k &= \left(\mathscr{B}_{k1}, \ldots, \mathscr{B}_{km}\right) = \text{double bouquet for each } k \text{ fixed,} \\ \mathscr{B}_{kj} &= \left(S_{kj1}, \ldots, S_{kjm}\right) = \text{bouquet for each } k,\, j \text{ fixed,} \end{aligned} \tag{13.162}$$

where as before the commas have been suppressed between the indices. These are chosen so that

$$[\mathscr{B}^3] = (S_{kji};\ 1 \leqslant i, j, k \leqslant m) \tag{13.163}$$

is a bouquet in the sense of §13.4 and as long as m has been chosen large enough the constructions below can be carried out.

For each of these double bouquets $\mathscr{B}^2_k$ coefficients λ_{kji}, and μ_{kj}, where $i, j =$

$1, \ldots, m$, can be chosen for which we can solve the corresponding b-equations (13.154), (13.159) with bounds, and define $T_k^2(e) = T^2_{\mathscr{B}_k^2}(e)$ as in (13.160).

This is the end of the story as far as S^3 is concerned when $s = 2$. Indeed, we can then choose scalars

$$\nu_1, \ldots, \nu_m \in \mathbb{R}, \quad \sup_k |\nu_k| = 1 \tag{13.164}$$

and for which we can solve with bounds the following equation in S^3:

$$\begin{gathered} bS^3(\xi) = \sum_k \nu_k T_k^2(\xi) = \sum_k \nu_k \big(S^2_{\mathscr{B}_k^2}(b\xi) + W^2_{\mathscr{B}_k^2}(\xi)\big); \quad \xi = e, \\ M_{p_3}\big(S^3(e)\big) \leqslant C \sum_k M_{p_2}\big(T_k^2(e)\big), \end{gathered} \tag{13.165}$$

with p_3, p_2 as in §13.6.1.2 and C independent of R.

As before, the definition of S^3 can be extended to a Γ-linear mapping for every $\xi \in \mathscr{C}^2$ and the defining equation that it satisfies is

$$\begin{gathered} bS^3(\xi) = \sum_k \nu_k S^2_{\mathscr{B}_k^2}(b\xi) + W^2_{\mathscr{B}^3}(\xi); \\ W_{\mathscr{B}^3}(\xi) = \sum_{k,j,i} \nu_k \mu_{kj} \lambda_{kji} S_{kji}; \quad \xi \in \mathscr{C}^2. \end{gathered} \tag{13.166}$$

From this, the required contradiction that gives the theorem follows for the case $s = 2$. This was outlined in §13.5 for $s = 1$ and it will be treated in the general case $s \geqslant 1$ in §13.7.4 below. The reader is urged to work this out for themselves using the 'triangular' nature of the coefficients ν_k, μ_{kj}, λ_{kji} (see Exercise 13.8). This together with (13.148), (13.153) and (13.164) allows us to find indices k_0, j_0, i_0 for which $|\nu_{k_0} \mu_{k_0 j_0} \lambda_{k_0 j_0 i_0}| = 1$; cf. also Remark 13.10 where one sees that, by adopting a slightly different approach, the notation in this side of things simplifies.

Apart from this, the other overall information that is used for the contradiction is contained in (13.149), (13.157), (13.165), where we have obtained the following estimates:

$$M_{p_1}(\hat{S}),\ M_{p_1}(T^1),\ M_{p_2}(S^2),\ M_{p_2}(T^2),\ M_{p_3}(S^3) = \mathrm{O}\big((\log R)^C\big) \tag{13.167}$$

for some constant that is independent of R and where $T^q = T^q(e)$, $S^q = S^q(e)$ for the basis elements e of $\mathscr{C}$.

13.7.3 The general definition of S^q, $1 \leqslant q \leqslant s+1$

The general definition that we shall give here is not likely to be very clear unless the reader has understood well how the two special cases S^1, S^2 of the previous two sections work out. In contrast, if one has acquired a good understanding of these special cases, then one probably does not need to read any further because it should be pretty clear how to go about things. For completeness, however, in this subsection we shall spell out the details of this general construction. The other thing the reader has to make sure that they understand properly is the construction of the indices $p_{s+1}, p_s, \dots$ of §13.6.1.2 because in what follows, these indices will be used freely without further explanation.

13.7.3.1 The lexicographical indices for multiple bouquets We have already encountered double bouquets in (13.152) and (13.162). This construction will now be iterated s times with s as in §13.1.1 for our SSAC group G. A bouquet is as defined in §13.4 and will be denoted by $\mathscr{B}^1$ here, to indicate that it is a simple bouquet. By taking a collection, $\mathscr{B}_i^1, \dots$, of simple bouquets we obtain double bouquets $\mathscr{B}^2$. And so on. This definition is summarised as follows:

$$\begin{aligned} \mathscr{B}^1 &= \mathscr{B} = (S_1, \dots, S_m) = \text{bouquet}, \\ \mathscr{B}^2 &= (\mathscr{B}_1, \dots, \mathscr{B}_m), \quad \mathscr{B}^3 = (\mathscr{B}_1^2, \dots, \mathscr{B}_m^2), \dots, \\ \mathscr{B}^{j+1} &= (\mathscr{B}_1^j, \dots, \mathscr{B}_m^j), \dots; \quad 1 \leqslant j \leqslant s. \end{aligned} \tag{13.168}$$

The additional requirement is that if we put together all the currents S_α that occur in (13.168) and denote this set of currents by $[\mathscr{B}^{s+1}] = (S_\alpha; \alpha)$, then we have a bouquet in its own right that consists of m^{s+1} currents S_α. We shall call $\mathscr{B}^j$ a (multiple) bouquet of order j. The length m of these bouquets will be large and will be chosen later.

The double bouquet $\mathscr{B}_p^2$ can be represented by currents $(S_{p,1}, \dots, S_{p,m})$ and the final bouquet $\mathscr{B}^{s+1}$ will be indexed lexicographically as follows:

$$\left[\mathscr{B}^{s+1}\right] = \left(S_{i_1,\dots,i_{s+1}};\ 1 \leqslant i_j \leqslant m,\ 1 \leqslant j \leqslant s+1\right), \tag{13.169}$$

where $[\cdots]$ in (13.169) indicates that we put together all the currents in the set $\mathscr{B}^{s+1}$; that is, (13.169) gives the index set for α in a lexicographic order as follows.

If we fix the first t indices $i_1, \dots, i_t$, where $t = 0, 1, \dots, s$ (the case $t = 0$ means that we fix no index at all), then $i_{t+1} = 1, \dots, m$ is the index of the bouquet of order $\mathscr{B}^{s-t}$ in the bouquet of order $\mathscr{B}^{s-t+1}$ to which the current S_α belongs in (13.168). With this indexing we can rewrite (13.168) backwards as follows:

$$\begin{aligned} &\mathscr{B}^{s+1} = (\mathscr{B}_{i_1}^s;\ 1 \leqslant i_1 \leqslant m), \quad \mathscr{B}_{i_1}^s = (\mathscr{B}_{i_1 i_2}^{s-1};\ 1 \leqslant i_2 \leqslant m); \quad 1 \leqslant i_1 \leqslant m, \\ &\vdots \\ &\mathscr{B}_{i_1,\dots,i_t}^{s+1-t} = (\mathscr{B}_{i_1,\dots,i_{t+1}}^{s-t};\ 1 \leqslant i_{t+1} \leqslant m); \quad 1 \leqslant i_1, \dots, i_t \leqslant m,\ t = 0, \dots, s, \end{aligned} \tag{13.170}$$

where for the last equation with $t = s$ to make sense we shall agree that a bouquet of order zero, $\mathscr{B}^0 = S$, is one single current. The indexing by $i_1, \dots, i_{s+1}$ in (13.169) is then what we get in (13.170) when $t = 0$. This indexing is thus lexicographic and every fixed index $i_1, \dots, i_{s+1}$ contains all the information that indicates which previous bouquets of successive orders this current S_α belongs to (this information on the 'history' of each current will, however, not be needed).

13.7.3.2 The coefficients and the construction of $\hat{S}$, S^1 Here the multiple bouquets and all the notation will be as in the previous subsection. The homologies in (13.147) do not vanish but they are of dimension $\leqslant h$. We conclude that for the indices p_1, p_0 of §13.6.1.2 and for every fixed $i_1, \dots, i_s = 1, \dots, m$ we can introduce an additional index i_{s+1} and find coefficients as follows:

$$\lambda_{i_1,\dots,i_s,i_{s+1}}, \qquad \sup_{1 \leqslant i_{s+1} \leqslant m} |\lambda_{i_1,\dots,i_s,i_{s+1}}| = 1; \quad 1 \leqslant i_i, \dots, i_s \leqslant m, \tag{13.171}$$

for which the following holds.

We can solve with bounds the following b-equation in $\hat{S} = \hat{S}^1$:

$$\begin{aligned} b\hat{S}_{i_1,\dots,i_s} &= \sum_{i_{s+1}} \lambda_{i_1,\dots,i_s,i_{s+1}} S_{i_1,\dots,i_s,i_{s+1}}, \\ M_{p_1}(\hat{S}_{i_1,\dots,i_s}) &\leqslant C \sum_{i_{s+1}} M_{p_0}\left(S_{i_1,\dots,i_s,i_{s+1}}\right) = \mathrm{O}\left(\log R\right)^C, \end{aligned} \tag{13.172}$$

for constants C that are independent of R. We can use these to define

$$\begin{aligned} S^1_{i_1,\dots,i_s}\left(\xi^0\right) &= \xi^0 \hat{S}_{i_1,\dots,i_s}, \\ W_{i_1,\dots,i_s}\left(\xi^1\right) &= \sum_{i_{s+1}} \lambda_{i_1,\dots,i_s,i_{s+1}} \xi^1 S_{i_1,\dots,i_s,i_{s+1}}, \\ T^1_{i_1,\dots,i_s}\left(\xi^1\right) &= S^1_{i_1,\dots,i_s}\left(b\xi^1\right) - W^1_{i_1,\dots,i_s}\left(\xi^1\right); \\ &\quad \xi^0 \in \mathscr{C}^0,\ \xi^1 \in \mathscr{C}^1,\ 1 \leqslant i_1, \dots, i_s \leqslant m. \end{aligned} \tag{13.173}$$

Here we have used our notation of §13.5.1 and written $W(\xi^1) = W^1(\xi^1)$. Also as before, S^1, W^1 and T^1 are Γ-linear mappings on the complex $\mathscr{C}$ of §13.6.2 with values in $\mathscr{D}'(G)$ and the convention for the exponents $\xi^p \in \mathscr{C}^p$ and S^p, T^p and $W^p \in \mathscr{D}'_{(r+p-1)}$ (cf. §§13.5.1 and 13.6.2) will be used in this section.

By this definition we have (see (12.84) and (13.104))

$$T^1_{i_1,\dots,i_s} = b\left(\xi^1 \hat{S}_{i_1,\dots,i_s}\right), \quad bT^1_{i_1,\dots,i_s} = 0; \quad \xi^1 \in \mathscr{C}^1,\ 1 \leqslant i_1, \dots, i_s \leqslant m. \tag{13.174}$$

This completes the first step of the construction.

13.7.3.3 The general case S^q The notation of §§13.7.3.1 and 13.7.3.2 will be preserved. In order to work out the inductive step that gives the Γ-mappings $S^1, S^2, \ldots, S^{s+1}$ we proceed as follows. We assume that Γ-mappings

$$\begin{aligned} S^q_{i_1,\ldots,i_{s-q+1}}(\xi) \in \mathscr{D}'_{(r+q-1)}(G); \quad \xi \in \mathscr{C}^{q-1}, \\ W_{i_i\ldots,i_{s-q+1}} : \mathscr{C} \longrightarrow \mathscr{D}'(G); \quad i_1,\ldots,i_{s-q+1} = 1,\ldots,m \end{aligned} \tag{13.175}$$

have been defined for a certain $1 \leqslant q \leqslant s$ and that these have a number of properties that we shall enumerate below. From this we shall work out an inductive step and construct the next two Γ-mappings down the road:

$$\begin{aligned} S^{q+1}_{i_1,\ldots,i_{s-q}}(\xi) \in \mathscr{D}'_{(r+q)}(G), \quad W_{i_1\ldots,i_{s-q}} : \mathscr{C} \longrightarrow \mathscr{D}'(G); \\ i_1,\ldots,i_{s-q} = 1,\ldots,m, \ \xi \in \mathscr{C}^q \end{aligned} \tag{13.176}$$

that enjoy the same properties. Notice that in this next step of the induction, the length of the index set decreases. To be consistent, in the last step when $q = s$ the multi-index of (13.169) is $\emptyset$ and the mapping $S^{s+1} = S^{s+1}_\emptyset$ is also defined, but for $W_\emptyset$ we do not need to bother. As for the first step $q = 1$ the mappings S^1 and W^1 are the ones that we constructed in §13.7.3.2 with new notation.

Now we come to the properties that the mappings S^q, W of (13.175) will have to satisfy. First for the W. These will be defined explicitly in terms of the currents that occur in the bouquets of §13.7.3.1 and depend on a choice of coefficients. This definition will be given in due course but for the moment we shall simply require that these mappings have the following property:

$$\begin{aligned} W : \mathscr{C}^p \longrightarrow \mathscr{E}'_{(r+p-1)}(G), \quad bW(\xi) = W(b\xi); \quad \xi \in \mathscr{C}, \\ \text{i.e. } b \circ W = W \circ b, \quad M_j(W(e)) = \mathrm{O}(\log R)^{c_j}; \\ j = 0,1,\ldots, \ e = e_I, \text{ the elements of the Γ-basis of } \mathscr{C}, \end{aligned} \tag{13.177}$$

for constants $c_0, c_1, \ldots$. Here, for $\xi^p \in \mathscr{C}^p$, we shall use the notation of §13.5.1 and write $W(\xi^p) = W^p(\xi^p)$. This condition is certainly satisfied by W^1, W^2 in §§13.7.2, 13.7.3.2.

The conditions for S^q are more involved. To write them down we put

$$\begin{aligned} T^q_{i_1,\ldots,i_{s-q+1}}(\xi^q) = S^q_{i_1,\ldots,i_{s-q+1}}(b\xi^q) + (-1)^q W_{i_1,\ldots,i_{s-q+1}}(\xi^q); \\ \xi^q \in \mathscr{C}^q, \ i_1,\ldots,i_{s-q+1} = 1,\ldots,m. \end{aligned} \tag{13.178}$$

What is required is that

$$bT^q = 0, \quad M_{p_q}(S^q(e)) = \mathrm{O}(\log R)^c; \quad e \text{ basis elements of } \mathscr{C}^{q-1} \tag{13.179}$$

for some $c > 0$ independent of R. From definition (13.178), and (13.177), it also follows that

$$M_{p_q}(T^q(e)) = \mathrm{O}(\log R)^c \tag{13.180}$$

for the same basis elements e of $\mathscr{C}^q$.

13.7.3.4 The choice of the coefficients and the inductive step: the overall strategy To make the inductive step, as in §13.6, we must solve with bounds the b-equation $bS^{q+1} = T^q$. But since the relevant homologies are only assumed finite-dimensional and not necessarily zero we must use Proposition 12.58. Hence the need to consider the bouquets as we did in §13.7.1. Furthermore, since several b-equations have to be solved, one for each Γ-basis element of $\mathscr{C}^q$, here we are obliged to use the further refinement of Proposition 12.58 that was explained in Exercise 13.7.

13.7.3.5 The explicit details Let e_I, with $|I| = q$, be the Γ-basis of $\mathscr{C}^q$. We know by the inductive step that $bT^q(e_I) = 0$ and we shall assume that $m \geqslant m_0(h,s)$ is large enough depending on s and h of (13.147). Then from §13.6.1.2 we conclude that we can find appropriate scalars λ with the properties

$$\begin{aligned} &\lambda_{i_1,\ldots,i_{s-q},i_{s-q+1}}; \quad 1 \leqslant i_t \leqslant m,\ t = 1,\ldots,s-q+1, \\ &\sup_{1\leqslant j\leqslant m} \left|\lambda_{i_1,\ldots,i_{s-q},j}\right| = 1; \quad 1 \leqslant i_t \leqslant m,\ t = 1,\ldots,s-q, \end{aligned} \tag{13.181}$$

and these scalars are such that we can solve with bounds the following b-equations in S^{q+1}:

$$\begin{aligned} bS^{q+1}_{i_1,\ldots,i_{s-q}}(\xi) &= \sum_{1\leqslant j\leqslant m} \lambda_{i_1,\ldots,i_{s-q},j} T^q_{i_1,\ldots,i_{s-q},j}(\xi), \\ M_{p_{q+1}}\left(S^{q+1}_{i_1,\ldots,i_{s-q}}(e_I)\right) &= \mathrm{O}(\log R)^C; \quad i_1,\ldots,i_{s-q} = 1,\ldots,m,\ \xi = e_I, \end{aligned} \tag{13.182}$$

for some C that is independent of R. Using the Γ-free property of $\mathscr{C}$ we also see that S^{q+1} in (13.182) can be extended to all the $\xi \in \mathscr{C}^q$ and we have thus Γ-linear mappings. Of course, the point here is that the particular choice of ξ in (13.182) among the finitely many e_I does not affect the choice of the λ. Putting things the other way round, we use §13.6.1.2 to choose *common* coefficients for all the basis elements e_I. To be more explicit, we *fix* the indices $i_1,\ldots,i_{s-q}$ and then the number of equations that have to be solved is equal to $\dim(\mathscr{C}^q)$; therefore if m is large enough, Exercise 13.7 can be used. This is done for every choice of these indices. To avoid confusion we stress that the λ depend on $i_1,\ldots$ that have been fixed but are the same for all basis elements.

For the definition of the S^{q+1} apart from the choice of the constants λ we also have made a choice of how we solve the b-equations (13.182). For the inductive definition of the W, no such additional choice is used. We simply set

$$W_{i_1,\ldots,i_{s-q}} = \sum_{1\leqslant j\leqslant m} \lambda_{i_1,\ldots,i_{s-q},j} W_{i_1,\ldots,i_{s-q},j}; \quad i_1,\ldots,i_{s-q} = 1,\ldots,m. \quad (13.183)$$

The verification that $b\circ W = W\circ b$ and that $M_{p_{q+1}}(W(e)) = \mathrm{O}(\log R)^C$ for the basis elements $e\in\mathscr{C}$ is thus immediate. To complete the proof that this is a legitimate inductive step (i.e. that (13.179) is satisfied after we have taken this step), it remains to verify that if we write

$$\begin{gathered} T^{q+1}_{i_1,\ldots,i_{s-q}}(\xi) = S^{q+1}_{i_1,\ldots,i_{s-q}}(b\xi) + (-1)^{q+1} W^{q+1}_{i_1,\ldots,i_{s-q}}(\xi), \\ i_1,\ldots,i_{s-q} = 1,\ldots,m,\ \xi\in\mathscr{C}^{q+1}, \end{gathered} \quad (13.184)$$

then $bT^{q+1}_{i_1,\ldots,i_{s-q}} = 0$. This follows from (13.182), (13.178) because these give

$$\begin{aligned} bS^{q+1}_{i_1,\ldots,i_{s-q}}\left(b\xi^{q+1}\right) &= \sum_{1\leqslant j\leqslant m} \lambda_{i_1,\ldots,i_{s-q},j} T^{q}_{i_1,\ldots,i_{s-q},j}(b\xi) \\ &= \sum_{1\leqslant j\leqslant m} (-1)^q \lambda_{i_1,\ldots,i_{s-q},j} W_{i_1,\ldots,i_{s-q},j}(b\xi); \\ &\quad \xi = \xi^{q+1}\in\mathscr{C}^{q+1}, \end{aligned} \quad (13.185)$$

and we need only to use $b\circ W = W\circ b$.

13.7.3.6 The final mapping S^{s+1} We shall end this subsection by collecting together our previous information for the final dimension $q = s+1$ and the mapping $S^{s+1}_{\emptyset} = S^{s+1}(\xi)$, where $\xi\in\mathscr{C}^s$. We shall write

$$\begin{aligned} &bS^{s+1}(\xi) = (-1)^s W(\xi) + \text{error term}, \\ &\quad W(\xi) = \xi\left(\sum \lambda_{i_1}\lambda_{i_1,i_2}\cdots\lambda_{i_1,\ldots,i_{s+1}} S_{i_1,\ldots,i_{s+1}}\right); \quad \xi\in\mathscr{C}^s, \end{aligned} \quad (13.186)$$

where the $S_{i_1,\ldots}$ are the currents that occur in the multiple bouquets of §13.7.3.1, and the λ in the definition of the coefficient

$$\mu_{i_1,\ldots,i_{s+1}} = \lambda_{i_1}\lambda_{i_1 i_2}\ldots\lambda_{i_1,\ldots,i_{s+1}}, \quad \text{the product of the } \lambda, \quad (13.187)$$

are the λ that we have picked up in the inductive construction. The only thing that counts and will be used below is the following property of these coefficients:

$$\sup_{1\leqslant i_{t+1}\leqslant m} \left|\lambda_{i_1,\ldots,i_t,i_{t+1}}\right| = 1; \quad 1\leqslant i_1,\ldots,i_t\leqslant m,\ 1\leqslant t\leqslant s. \quad (13.188)$$

From this it follows in particular that we can make one choice $i^0_1, i^0_2,\ldots,i^0_{s+1}$ for which the coefficient in (13.187) is

$$\left|\mu_{i^0_1,\ldots,i^0_{s+1}}\right| = 1. \quad (13.189)$$

So much for the principal term. (You can also use Remark 13.10 and 'simplify'.)

Now, how about the error term? This of course is much more involved to write down explicitly and it also depends on the particular solutions of the b-equations that we have chosen in the successive inductive steps. However, as we shall see, for this error term it is quite easy to give appropriate estimates.

13.7.4 The estimate of S^{s+1}; the principal term; the error term

13.7.4.1 The test current As we did before we shall now apply the mapping S^{s+1} to the current

$$\xi_D = \sum_{\gamma} \delta_\gamma e = [-D, D]^s \in \mathscr{C}^s. \tag{13.190}$$

Here $e = e_{(1,2,\dots,s)}$ is the unique Γ-basis element of $\mathscr{C}^s$ and the summation is taken on

$$\gamma = \left[(n_1, \dots, n_s) \in \Gamma = \mathbb{Z}^s;\ -D \leqslant n_j \leqslant D - 1,\ 1 \leqslant j \leqslant s\right]. \tag{13.191}$$

This is the chain and the current $[-D, D]^s \in \mathscr{E}'(A) \subset \mathscr{E}'(G)$ for the orientation $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_s$ which is the Lebesgue measure (cf. §12.8).

Now we come to

$$D = \left[(\log R)^{\hat{c}}\right] = \text{integer part of } (\log R)^{\hat{c}}. \tag{13.192}$$

The important point here is that the constant $\hat{c} > 0$, which is independent of R, will be chosen at the very end and it will depend on the geometric constants of the group G and on the various other constants that had to be introduced in the process of the construction of S^{s+1}. Putting this the other way round, all the constants that crop up in the construction of S^{s+1} are independent of this $\hat{c} > 0$.

Formula (13.182) will now be applied to this current and we have

$$\begin{aligned} bS^{s+1}(\xi_D) &= (-1)^s W(\xi_D) + \mathscr{P}_D, \\ \text{error term} = \mathscr{P}_D &= \sum_{1 \leqslant i \leqslant m} \lambda_i S_i^s (b\xi_D). \end{aligned} \tag{13.193}$$

The other piece of information that will be used is a global one on $S^{s+1}(\xi_D)$ and not on its boundary. We have the estimate

$$M_0\left(S^{s+1}(\xi_D)\right) \leqslant M_{p_{s+1}}\left(S^{s+1}(\xi_D)\right) \leqslant \sum_{\gamma} M_{p_{s+1}}\left(\delta_\gamma S^s(e)\right) = \mathrm{O}(\log R)^{\hat{C}}. \tag{13.194}$$

For this we use (13.182), §12.8.2 and the size of D. Here $\hat{C}$ also depends on $\hat{c}$ and not only on the constants that crop up in the construction of S^{s+1}.

13.7.4.2 The localisation of the principal term to the central slice In $G = N \leftthreetimes (A' \oplus A) = N \leftthreetimes V$ we consider as before the characteristic function χ of the central slice

$$N_c = \pi^{-1}(V_c) \subset G; \quad \pi\colon G \to V = \text{canonical projection}, \\ V_c = [v \in V;\ |v| \leqslant c], \tag{13.195}$$

for some appropriately small $c > 0$. We shall localise the current W in the principal term of (13.186) and examine

$$\chi W(\xi_D) = \sum_{i_1,\dots,i_{s+1}} \mu_{i_1,\dots,i_{s+1}} \chi\left(\xi_D S_{i_1,\dots,i_{s+1}}\right). \tag{13.196}$$

The localisation of each individual term (see (13.139))

$$\chi\left(\xi_D S_{i_1,\dots,i_{s+1}}\right) = \sum \eta_P \delta_P [V_c] \tag{13.197}$$

has already been analysed twice before (in §13.5.4 in the special case $s = 1$ and §13.6.6 in the general case). This will therefore not be repeated. Summing over all the indices we put all these together and write

$$\chi W(\xi_D) = \sum_{i_1,\dots,i_{s+1};P} \mu_{i_1,\dots,i_{s+1}} \eta_P \delta_P [V_c], \tag{13.198}$$

where in that summation, for every fixed choice of the indices $i_1, \dots, i_{s+1}$, the P runs through the vertices of the cube $\square_{cR}$ that is used to define the current $S_{i_1,\dots,i_{s+1}}$ in the definition of the bouquet (13.169); see §13.7.3.1.

The comments that we made in §13.5.4 about the orientations apply again. We recall that to be able to define the $\eta_P = \pm 1$ we need to specify the orientation that is used when we define the corresponding $S = S_{i_1,\dots,i_{s+1}}$ near the vertex P. This was explained in Chapter 10 and again in §§13.5 and 13.6. This, as already pointed out, is easy to see when this current has been smoothed. However, if we have not done that smoothing, this orientation can again be defined, but this amounts to a non-trivial exercise (see Exercise 10.9). Alternatively, as we have already explained (e.g. §13.5.4.2), we can localise further and consider instead $\tilde{\chi} W(\xi_D)$ with $\tilde{\chi}$ the characteristic function of an appropriate smaller slice, as in §13.5.4.2.

13.7.4.3 The localisation of the error term This is easier but here we have to keep track of the constants in the corresponding powers of $\log R$.

First of all observe that, exactly as in Exercise 13.9, we have

$$b\xi_D = \sum_{\gamma,J} \pm \delta_\gamma e_J; \quad |J| = s - 1 \text{ (see (13.191))}, \ |\gamma| = \sup_j |n_j| = D. \tag{13.199}$$

This is just the building up of the boundary $\partial[-D,D]^s$ by subcubes of dimension $(s-1)$ and size 1, and they are all at a distance D from 0 in A.

The total number of terms in the summation is

$$\# \text{ of terms} \sim D^{s-1}. \tag{13.200}$$

We recall on the other hand that quite generally

$$M_0\left(T\mathbb{I}(|g|>r)\right) \leqslant r^p M_p(T) \tag{13.201}$$

for every integration current T, any index p, any $r>0$ and the indicator function $\mathbb{I}$ of the set outside the ball of radius r (cf. (12.68)). We observe also that in the summation (13.199) the Riemannian distance in G of the support of each current $\delta_\gamma e_J$ from the slice N_c of (13.195) is given by

$$\text{distance} \gtrsim D \sim (\log R)^{\hat{c}}. \tag{13.202}$$

This follows from the fact that the canonical projections $G \xrightarrow{\pi} V \to A$ contract the distances.

The localisation of the error term can now be completed as follows. First we have

$$\mathscr{P}_D = \sum_{1\leqslant i\leqslant m} \lambda_i S_i^s\left(b\xi_D\right) = \sum_{\gamma,iJ} \lambda_i \delta_\gamma S_i^s\left(e_J\right), \tag{13.203}$$

where we sum as in (13.199). The total number of terms in the second summation by (13.200) is

$$\# \text{ of terms} \leqslant CmD^{s-1} = C_{h,s}D^{s-1} \leqslant C_{h,s}(\log R)^{s\hat{c}}, \tag{13.204}$$

where, by the definition of m, the term $C_{h,s}$ depends on s and the dimension of the homologies h in (13.147) but is in particular independent of R and $\hat{c}$.

By (13.201) and (13.182), for the localisation of each term in (13.199) we have

$$M_0\left(\chi\left(\delta_\gamma S_i^s(e_J)\right)\right) \leqslant (\log R)^{-\hat{c}p_s}(\log R)^C; \quad R \geqslant C, \tag{13.205}$$

where C is independent of R and also of the choice of $\hat{c}$.

Putting these together we finally obtain

$$M_0\left(\chi\mathscr{P}_D\right) \leqslant (\log R)^C(\log R)^{(s-p_s)\hat{c}}; \quad R \geqslant C. \tag{13.206}$$

The exponent of $\log R$ in (13.206) is

$$\text{exponent} = C + (s-p_s)\,\hat{c}; \tag{13.207}$$

here p_s is also independent of $\hat{c}$. Furthermore, p_s, which was chosen in §13.6.1, is larger than $10^{10}+s$. It is therefore down to the choice of $\hat{c}$ which will be made at this point to make sure that the exponent (13.207) is $\leqslant -10$.

We shall finally put together in abbreviated terms the information we have obtained on the principal term and the error term localised on the central slice N_c. We shall write

$$\chi\left(bS^{s+1}(\xi_D)\right) = \sum \mu_{i_1,\dots,i_{s+1}} \eta_P \delta_P[V_c] + \mathrm{O}(\log R)^{-10}, \tag{13.208}$$

where the summation is taken over the $i_1,\dots,i_{s+1} = 1,\dots,m$ and the corresponding vertices as in (13.198) and where the O-error is interpreted in terms of the M_0-norm as in (13.182), (13.205).

13.7.5 The endgame and the contradiction

The same differential form on G as in (13.143) is used:

$$\omega(g) = \varphi(g)\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{r+s-1} \tag{13.209}$$

where y_i are the Euclidean coordinates of V.

This form satisfies properties (i), (ii), (iii) and (iv), which were given in §§13.3.3, 13.5.7 and used again in §13.6.6(c). As before, the value of φ is preassigned near *all* the vertices that are used in the summation (13.208). In other words, we must have, exactly as in §13.4.3, (13.34),

$$\varphi(g) = \varepsilon_P \varphi_0(v), \quad \varepsilon_P = 0, \pm 1; \quad g = (P, v) \in G,\ v \in V_c, \tag{13.210}$$

where P is viewed as a point of N (since $N' \subset N$) and P runs through all these vertices.

Now, as before, all these ε_P but one will be chosen to be 0. The one vertex P_0 for which ε_{P_0} is not zero is chosen on the support of the current $S_{i_1^0,\dots,i_{s+1}^0}$ for which $\mu_{i_1^0,\dots,i_{s+1}^0} = \pm 1$ (cf. (13.189)) and this P_0 is not the identity of G. The conclusion as in §13.6.6 is that a choice can be made so that

$$\langle W(\xi_D), \omega \rangle = 1. \tag{13.211}$$

But then by (13.208) the other two estimates (13.145), (13.146) of §13.6.6 hold verbatim and the contradiction follows.

Adapting the above to the Heisenberg alternative will be left as an exercise to the reader.

13A Appendix

13A.1 The use of an infinite bouquet

For simplicity let us stick with the group $G_2 = \mathbb{R}^2 \rtimes \mathbb{R}$ (see §§13.1, 7.6.3, 9.2.2) and construct a sequence of cubes $\square_j \subset \mathbb{R}^2$, as in §13.4, of size R_j but now for an infinite sequence of parameters that grow very rapidly: $1 \ll R_1 \ll \cdots$. By the first basic construction, we can construct currents $S_j \in \mathscr{E}'_{(1)}$ for each cube as in §13.4; we shall define the new current $S = \sum \alpha_j S_j \in \mathscr{D}'_{(1)}$, where the positive coefficients α_j are chosen such that $\sum \alpha_j (\log R_j)^c < +\infty$ for all $c > 0$. This current has the following properties.

First, $S \in \mathscr{C}^*(\mathrm{pol})$ (see (12.52), (13.8)) and, denoting by $\tilde{S}_j \in \mathscr{E}'_{(2)}$ an arbitrary compactly supported current such that $b\tilde{S}_j = S_j$, then we have $S = \lim bT_n$ for the topology of $\mathscr{C}^*$ induced by the seminorms (12.49), with $T_n = \sum_{j \leqslant n} \alpha_j \tilde{S}_j$. All this is clear enough; but what is also true is that the coefficients α_j and the sizes R_j can be chosen in such a way that it is impossible to solve the equation $bT = S$, where $M_0(T) < +\infty$.

To see this we closely follow §§13.3, 13.4 and construct a sequence $\omega_n = \varphi_n \, \mathrm{d}y$ of smooth differential forms that live in $[|y| < c]$ (with the same notation for the coordinates (x_1, x_2, y) as in §13.3) that have the following properties. We have $\|\mathrm{d}\omega_n\|_0 \leqslant C R_n^{-c}$ for some $c > 0$, and the values of φ_n near the vertices are pre-assigned as in (13.34). More explicitly, near the vertices of $\square_n$, other than the identity e we set, say, $\varphi_n = \pm \alpha_n^{-1}$ or 0 and φ vanishes near all the other vertices. It follows that $\langle S, \omega_n \rangle \approx 1$ (for a proper choice of the $\pm$) and, additionally, if $bT = S$ solves the equation with a bound as above then $\langle S, \omega_n \rangle = \langle T, \mathrm{d}\omega_n \rangle = O(R_n^{-c})$ for some $c > 0$. This contradiction proves our assertion (see (13.35)).

From this and §12A.4 we deduce that $\mathrm{Im}\,\mathrm{d}$ is *not closed* for the group G_2. More precisely, for the complex $\Lambda_P(G_2)$, assigned with the inductive limit topology of §12.13, the subspace

$$\mathrm{Im}\left[(\text{forms of degree } 1) \to (\text{forms of degree } 2)\right]$$

is not closed because otherwise we would contradict §12A.4.

It is clear that the above argument can be generalised to some of the other SSAC groups that are considered in this chapter (no doubt all!). From this, using the material of the next chapter, it is safe to assume (insofar as such a thing can be said about a mathematical statement) that for every B-group one can prove that $\mathrm{Im}\,\mathrm{d}$ is never closed in Λ_P. This matter, however, will not be pursued further.

13A.2 The topological homotopy

Position of the problem The complexes $\cdots \to \Lambda_n \xrightarrow{\mathrm{d}} \Lambda_{n+1} \to \cdots$ that we have considered in this book consist of topological vector spaces; therefore the closure for that topology, $\overline{\mathrm{Im}\,\mathrm{d}}$, can be considered as in §13A.1 and a new homology group $\overline{H} = \mathrm{Ker}\,\mathrm{d}/\overline{\mathrm{Im}\,\mathrm{d}}$ (actually a vector space) can be defined.

This new way of defining polynomial homology for a Riemannian manifold M is easily seen (verify this) to be invariant under polynomial homotopy and, together with polynomial mappings, this gives a new contravariant functor. Because of (12.128), that is, Poincaré duality, one can define a canonical mapping

$$\overline{H}(M) = \overline{H}(\Lambda_P(M)) \to H(M, \text{de Rham}). \tag{13A.1}$$

If the manifold is an NB-connected Lie group, then (13A.1) is an isomorphism because of the polynomial homotopy equivalence of §12.9.8.

In the rest of this appendix we shall stick to the same group $\mathbb{R}^2 \rtimes \mathbb{R}$ of §13A.1 and prove that (13A.1) is an isomorphism. This will be done in a series of steps (i.e. exercises). In what follows G will always denote this group. The reason we do this is that it shows that (13A.1) is an isomorphism for a group that is not NB.

The distance from the origin. Notation and facts On the group $G = G_2 = \mathbb{R}^2 \rtimes \mathbb{R}$, the same coordinates (x_1, x_2, y) as in §13.3.2 will be used to identify G with $\mathbb{R}^3$ (see §§7.6.3, 9.2.2). Throughout, we shall use the notation $G \ni g \leftrightarrow (x_1, x_2, y) = X \in \mathbb{R}^3$ and also denote $\|X\| = \log(|x_1| + 10) + \log(|x_2| + 10) + |y|$. Then there exist constants such that

$$C_2\|X\| \leqslant |g| \leqslant C_1\|X\|; \quad g \in G, \ |g| = \text{dist}(e, g) \geqslant C. \tag{13A.2}$$

For the proof we just have to combine the following two observations. Let $g = xy$ be the group product with $x \in \mathbb{R}^2$ and $y \in \mathbb{R}$. Then

(a) $|y| \leqslant |g|$ (to see this project onto $\mathbb{R}$) and $|y| = |g|$ when $x = 0$;
(b) when $y = 0$ and $g = (x_1, x_2) \in \mathbb{R}^2$ then $|g| \approx \log(|x_1| + 10) + \log(|x_2| + 10) \geqslant C$ – this follows easily from the considerations of §8.2.2 (if you get stuck, look at the explicit general proof of this in Varopoulos, 2000a).

We shall now streamline the notation for the polynomial complexes of the group G and denote

$$(\Lambda_P, \mathrm{d})\colon 0 \to E_0 \to E_1 \to E_2 \to E_3 \to 0,$$

where E_p denotes the space of forms of degree p. The topologies are, as usual, the inductive limit. We can then assert the following:

(i) $\text{Im}(E_0 \to E_1) = \text{Ker}(E_1 \to E_2)$. This holds for any Riemannian manifold because we can integrate any closed form ω along an appropriate path from e to g and then $V(g) = \int \omega$ solves the Poincaré equation $\mathrm{d}V = \omega$.

(ii) $\mathrm{Im}(E_2 \to E_3) = E_3$. Let $\omega = f\,\mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}y \in E_3$; then

$$\theta = \left(\pm \int_{y_0}^{y} f(x_1, x_2, y)\,\mathrm{d}y \right) \mathrm{d}x_1 \wedge \mathrm{d}x_2 \in E_2; \quad y_0 = \mp\infty \text{ or } 0$$

solves $\mathrm{d}\theta = \omega$. To see that $\theta \in E_2$ use the orthonormal basis of Exercise 7.17 for the cotangent space

$$\omega_1 = \mathrm{e}^{-\alpha y}\,\mathrm{d}x_1, \quad \omega_2 = \mathrm{e}^{\beta y}\,\mathrm{d}x_2, \quad \omega_3 = \mathrm{d}y, \tag{13A.3}$$

where $\alpha, -\beta$ are the two roots. The details are left to the reader and they are, if anything, easier than what is done below.

The only interesting new assertion is the following.

(iii) $\overline{\mathrm{Im}(E_1 \to E_2)} = \mathrm{Ker}(E_2 \to E_3)$. The proof of this uses duality. In concrete terms, this is the first step (a) in the proof below.

Proof of (iii) (a) We shall assume by contradiction that (iii) fails. In the notation of §12.7 this means that there exist some $p > 0$ and some closed $\omega \in \mathscr{C}_p$ such that $\omega \notin \overline{\mathrm{Im}\,\mathrm{d}}$. It follows that for all $q > p$ we can find some $T \in (\mathscr{C}_q)^*$ (see (12.52)) such that $bT = 0$ and $T[\omega] = 1$. To see this we go back to (12.63), (12.64) for the explicit description of the dual space Λ_P^* and use Hahn–Banach. This gives $S = T + bT_1 \in \Lambda_P^*$ with T as above and $T_1 \in (C_r^0)^*$ for some $r \gg p$. We shall show that if q is chosen large enough, a contradiction will be obtained.

(b) *Smoothing by a mollifier*. As we pointed out, the coordinates (x_1, x_2, y) are used to identify G and $\mathbb{R}^3$ and the spaces Λ_P, Λ_P^* with the spaces of currents on $\mathbb{R}^3$. On $\mathbb{R}^3$ we can use (Euclidean) convolution by some $f \in C_0^\infty$ and smooth out in these spaces because by (13A.2), (13A.3), we readily see that $\Lambda_P * f \subset \Lambda_P$ and $\Lambda_P^* * f \subset \Lambda_P^*$. As a consequence, the current T that was constructed in (a) can be assumed to be smooth: $T = a_1\,\mathrm{d}x_1 + a_2\,\mathrm{d}x_2 + b\,\mathrm{d}y$ (note that the convolution commutes with the d and b operators on the complexes).

Furthermore, from (13A.3) it follows that the form $\mathrm{e}^{\gamma y}\,\mathrm{d}x_1 \wedge \mathrm{d}x_2 = \omega_1 \wedge \omega_2$ is bounded when $\gamma = \beta - \alpha$ and therefore $\|X\|^q \mathrm{e}^{\gamma y}\,\mathrm{d}x_1 \wedge \mathrm{d}x_2 \in \Lambda_q$. From this we deduce that

$$\int |b(X)|\; \|X\|^q \mathrm{e}^{\gamma y}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y < +\infty. \tag{13A.4}$$

If need be we can change the direction of the y-axis and assume that $\gamma \geqslant 0$. We can also convolve with an additional mollifier so that for the Euclidean gradients $\nabla = \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial y}$, and all $k = 0, 1, \ldots$ and $q, \gamma \geqslant 0$ as above, we have

$$\int_{-\infty}^{\infty} |\nabla^k b(x_1, x_2, y)|\; \|X\|^q \mathrm{e}^{\gamma y}\,\mathrm{d}y < +\infty; \quad x_1, x_2 \in \mathbb{R}. \tag{13A.5}$$

(c) *The Poincaré equation and the contradiction*. Because of (13A.5) we can define $V = -\int_y^\infty b\,\mathrm{d}y$ and verify that $\mathrm{d}V = T$ (the form $\mathrm{d}V - T = \mathrm{e}\,\mathrm{d}x_2$ with $\mathrm{e} = -\int_{-\infty}^{+\infty} \frac{\partial b}{\partial x_2}\mathrm{d}y$ – see (13A.5) – is closed, therefore $\mathrm{e} = 0$ because of (13A.4), (13A.5)). We shall presently show that

$$\int |V|\ \|X\|^r \mathrm{e}^{\gamma y}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y; \quad r = q - 10. \tag{13A.6}$$

From this, the contradiction easily follows: let $\chi \in C_0^\infty(\mathbb{R}^3)$ be some appropriate cut-off function that is $\equiv 1$ in larger and larger subsets (see Exercise 12.26). Then $\mathrm{d}(\chi V) = \chi T + V\mathrm{d}\chi \to T$ in $(\mathscr{C}_r^0)^*$ and if $r > p$ we have the contradiction $0 = \langle \mathrm{d}\omega, \chi V\rangle = \langle \omega, b(\chi V)\rangle \to T[\omega] = 1$.

(d) *The proof of* (13A.6). For this we shall use a different expression for $\|X\|$ in the integrals and instead we set $\|X\| = (1+|y|)\big(\log(|x_1|+10) + \log(|x_2|+10)\big)$. By changing the values of q and r, we see that this makes no difference. This allows us to split the integrals into y and (x_1, x_2) integrals. In fact, if $\gamma > 0$, then (13A.6) simply follows from the observation that

$$\int_{-\infty}^{+\infty} \mathrm{e}^{\gamma y}(1+|y|)^r \left(\int_y^\infty |b|\,\mathrm{d}y\right)\mathrm{d}y \leqslant C\int_{-\infty}^{+\infty} \mathrm{e}^{\gamma y}(1+|y|)^{r+1}|b|\,\mathrm{d}y. \tag{13A.7}$$

This is just integration by parts because the integrated term so obtained is $\lesssim \mathrm{e}^{\gamma y}(1+|y|)^{r+1}\int_y^\infty |b|\,\mathrm{d}y$ and vanishes at $y = \pm\infty$. To see this note that $\int_y^\infty |b|\,\mathrm{d}y \lesssim \mathrm{e}^{-\gamma y}(1+|y|)^{-q}$ by (13A.4) for $y \in \mathbb{R}$ (check separately for $y > 0$ and $y < 0$).

If $\gamma = 0$, observe that because of (13A.5) and the fact that $\mathrm{d}T = 0$, we have

$$a_1 = -\int_y^\infty \frac{\partial a_1}{\partial y}\mathrm{d}y = -\int_y^\infty \frac{\partial b}{\partial x_1}\mathrm{d}y,$$

$$a_2 = \int_{-\infty}^y \frac{\partial a_2}{\partial y}\mathrm{d}y = \int_{-\infty}^y \frac{\partial b}{\partial x_2}\mathrm{d}y,$$

$$-\int_y^\infty \frac{\partial^2 b}{\partial x_1 \partial x_2}\mathrm{d}y = \frac{\partial a_1}{\partial x_2} = \frac{\partial a_2}{\partial x_1} = \int_{-\infty}^y \frac{\partial^2 b}{\partial x_1 \partial x_2}\mathrm{d}y,$$

where all the integrals are absolutely convergent. This means that $B = \int_{-\infty}^\infty b\,\mathrm{d}y$ satisfies $\frac{\partial^2 B}{\partial x_1 \partial x_2} = 0$, that is, $B = f_1(x_1) + f_2(x_2)$ and therefore by (13A.4) again $B = 0$.

Having this, integration by parts is used again on $\int_{-\infty}^\infty (1+|y|)^r \left|\int_y^\infty b\,\mathrm{d}y\right|\mathrm{d}y$, and here, to show that the integrated term vanishes for $y = -\infty$, we use (13A.4) and $B = 0$. We then finish as in (13A.7). (Added in proof): A general theorem in that direction for ‘rank 1’ groups is stated without proof in the epilogue. Results of that nature are elementary, but tricky to prove. What one should conjecture for higher rank is anything but clear.

13A.3 Further comments

The argument that we gave is ad hoc and gives no indication of how to go about the homomorphism (13A.1) in general. By a similar ad hoc argument we can show that for any Riemannian manifold M (assume for simplicity that M is diffeomorphic to $\mathbb{R}^n$, with $n \geqslant 2$) and any $T \in \Lambda_P^*(M)$, a closed current of *dimension* 1, we have $T \in \overline{\operatorname{Im} b}$ for the projective limit topology; cf. §12.13. To see this, it suffices to find a sequence of *compactly supported closed* currents T_n that converges to T in Λ_P^* because all these currents lie in $\operatorname{Im} b$ (see Bott and Tu, 1982, §4.7.1; de Rham, 1960, §19). The interested reader can find some details of how this is done in the exercise below.

Exercise Verify the following facts and for simplicity assume $T \in \mathscr{C}^*(M, \text{pol})$ – see (12.52) – and $bT = 0$.

(i) Use the cut-off function as in Exercise 12.24 and verify that
  (1) $S_R = \mathrm{d}(\chi T) = \mathrm{d}\chi \wedge T$ is supported in the annulus $P_R = [m;\ R \leqslant |m| \leqslant R+1]$ and $M_p(S_R) = \mathrm{O}(R^{-a})$ for all $p, a > 0$ where R is a 'free parameter' that will be made to tend to infinity;
  (2) $\langle S_R, 1\rangle = \int \mathrm{d}(\chi T) = 0$ and therefore the zero-dimensional current S_R (i.e. a Radon measure) can be written as a vector integral (see Bourbaki, 1963, Chapter 6) $S_R = \int_{x,y \in P_R} (\delta_x - \delta_y)\, \mathrm{d}\mu(x,y)$ where μ is a Radon measure of total mass bounded by $M_0(S_R)$.

(ii) We have $\delta_x - \delta_y = b\ell_{x,y}$ where $\ell_{x,y}$ is a one-dimensional current given by a path (as in §12.3.4) that joins x to y of length $\lesssim R$.

(iii) Put these together and set $T_R = \chi T - \text{error}$, with

$$\text{error} = \int_{x,y \in P_R} \ell_{x,y}\, \mathrm{d}\mu(x,y).$$

Then $\mathrm{d}T_R = 0$, and error $\to 0$ as $R \to \infty$, that is, $T_R \to T$ in Λ_P^*. Done!

Note finally that in the case at hand the same fact can be seen by abstract methods simply by dualising (iii) of §13A.2. For this we use the functional analysis of the remark in §12.11 combined with the regularisation of §12.10, which allows us to reduce the problem to the spaces (12.137). The details are left to the reader, who may be interested in putting this appendix in the context of the TVS and the dualities that we presented in the second half of Chapter 12. We shall not, however, pursue this matter.

14
Cohomology on Lie Groups

14.1 Introduction: Scope and Methods of the Chapter

This chapter differs in many essential ways from the rest of the book and it is desirable to make it as self-contained as possible. We shall start therefore by summarising what has been done in Chapters 12 and 13 and highlighting the problem that remains to be addressed.

14.1.1 The de Rham complex revisited

On M, some C^∞ manifold, we shall denote by $\Omega^*(M) = \sum \Omega^p(M)$ the complex of C^∞ differential forms on M, where $p = 0, 1, \ldots, n = \dim M$ denotes the degree of the form, and for $\omega \in \Omega^*$ we have the corresponding decomposition $\omega = \sum \omega^p$. For typographical reasons we sometimes omit the $*$.

Notation In previous chapters we have used the letter Λ to indicate the various spaces of differential forms and, more often than not, the forms that we considered in those chapters were not necessarily smooth but had continuous or even L^∞_{loc} coefficients. We change the notation here to stress the point that in this chapter we shall work exclusively in the C^∞ category (this Ω notation is one that many topologists use, e.g. Bott and Tu, 1982; on the other hand de Rham, 1960 uses the letter $\mathscr{E}$).

In our case, more often than not the manifold M will be some connected Lie group G and then (see §12.5) we can fix $\omega_1, \ldots, \omega_n \in \Omega^1(G)$ some global basis of left-invariant 1-forms (i.e. we choose some basis of $T^*_e(G)$ the cotangent space at the identity and move it about by left translation). An arbitrary $\omega \in \Omega^p$, can then be written

$$\omega = \sum_I a_I(g)\omega_I; \quad \omega_I = \omega_{i_1} \wedge \cdots \wedge \omega_{i_p},\ g \in G, \tag{14.1}$$

where $I = (i_1, \ldots, i_p)$, $1 \leqslant i_j \leqslant n$ runs over all the increasing multi-indices of length $|I| = p$ (i.e. $i_1 < i_2 < \cdots$). It follows that for any $\omega \in \Omega(G)$ we can define

$$|\omega(g)| = \sum_p \sum_{|I|=p} |a_I(g)|; \quad g \in G, \tag{14.2}$$

and that this, up to equivalence (i.e. within multiplicative constants), is unambiguous and independent of the basis $\omega_1, \ldots, \omega_p$. With a little additional care, this norm can be defined in a Riemannian manifold M (see §12.5 or, say, Warner, 1971). Uniform norms of differential forms can be defined using (14.2).

We shall say that a differential form ω is of polynomial growth, denoted by $\omega \in \Omega(G; \text{Pol})$, if there exist constants C such that

$$|\omega(g)| \leqslant C(1+|g|)^C; \quad g \in G, \tag{14.3}$$

where $|g|$ is the distance of g from the identity for the left-invariant distance on G. The notion also extends to any Riemannian manifold (see §12.5).

We shall use these forms to construct cohomology: we define the polynomial complex

$$\Omega_P(G) = [\omega \in \Omega^*(G);\ \omega, \mathrm{d}\omega \in \Omega^*(G, \text{Pol})]. \tag{14.4}$$

We can also make the same definition Ω_P for any Riemannian manifold M. The homology of this complex will be denoted $H(G;\text{Pol})$ (or $H_{\text{DR}}(\ldots)$ to emphasise that it is the de Rham cohomology we are using). When the polynomial growth is clear from the context we shall even drop the 'Pol' and simply write $H_P(G)$ or even just $H(G)$.

This homology is naturally graded and we can define the Betti numbers $b_p = \dim H^p(G)$, where $H^p = \text{Ker}\left(\Omega_P^p \xrightarrow{\mathrm{d}} \Omega_P^{p+1}\right) / \text{Im}\left(\Omega_P^{p-1} \to \Omega_P^p\right)$, and where Ω_P^p is the natural grading of (14.4) and as usual we set $\Omega_P^p = 0$ for $p < 0$. When all the Betti numbers are finite (i.e. $b_p < +\infty$) we shall write $H_{\text{DR}}(G;\text{Pol}) < +\infty$ and say that G (or M) has finite (de Rham) polynomial homology.

14.1.2 What has been done and what remains to be done

It is useful at this point to resurrect terminology that we used in §1.5. We said there that a Lie group G is a *model* if G is diffeomorphic with some Euclidean space $\mathbb{R}^n$. Such a group is of course nothing more than a simply connected soluble group such as the ones that we considered in Chapters 12 and 13.

Exercise To see this use the Levi decomposition (Varadarajan, 1974, §3.15); when G is a model then $G = Q \leftthreetimes S$ and the semisimple factor S is $\{0\}$ because as a differential manifold it is contractible. Here we ask the reader to look back at Exercise 11.18.

We can now summarise Chapters 12 and 13 in the following result.

Theorem 14.1 *Let G be some model. Then the following are equivalent:*

(i) *G is an NB-group*;
(ii) $H_{\mathrm{DR}}(G;\mathrm{Pol}) < +\infty$.

This was proved in Chapters 12 and 13 but for a different complex of forms that were not necessarily smooth. We denoted this complex by $\Lambda_P(G)$. In Proposition 12.18 and §12.10 we showed how a standard regularisation procedure can be used to pass from this complex to the smooth complex $\Omega_P(G)$. It is the implication

$$H(G,\mathrm{Pol}){<}{+}\infty \implies G \text{ is NB} \tag{14.5}$$

that is the issue in those two chapters. The implication the other way round was done in §12.9 and in Appendix F with the use of the polynomial retract and this was done for general Lie groups without the assumption that the group is a model. Note that the homotopy we constructed in Appendix F can be made to be smooth (see also §12.2.3). The implication (14.5) will be the missing link needed for a complete proof of Theorem 12.9 (see §12.9.8 and §§1.6–1.8 for a refresher of the 'overall' picture).

What remains to be done – and this is the aim of this chapter – is to prove implication (14.5) for a general Lie group without the assumption that the group is a model. From this, the theorem of §1.6.3 follows at once.

Exercise Show that the general case of (14.5) can be used to complete the proof of Theorem 12.9. When G is polynomially contractible to a compact set as in §§12.2.2–12.2.4, we can follow the contracting homotopy with the obvious local homotopy coming from (12.1) and deduce that there exists a mapping $\Lambda_P^* \ni T \to fT \in \mathscr{D}'(K)$ where $K \subset G$ is some compact subgroup, and such that when T is closed then fT is closed and polynomially homologous to T. This implies that $H(\Lambda_P^*(G))$ is finite-dimensional. This in turn implies that $H(G,\mathrm{Pol}) < +\infty$ by §12.13.3. We then apply (14.5).

14.1.3 Simply connected groups and the general strategy in the proofs

It is very easy to see that with what we have we can deal at once with simply connected groups. Indeed, we saw in §11.1.3 that such a group G is (smoothly) quasi-isometric with a group of the form $Q \times K = G_1$, where Q is a model and K is compact. On the other hand it is easy to see that $H(G_1, \mathrm{Pol}) < +\infty$ if and only if $H(Q; \mathrm{Pol}) < +\infty$.

Remark Only the implication $H(G_1) < +\infty \Longrightarrow H(Q) < +\infty$ is needed and this is automatic from $Q \to G_1 \to Q$ and the induced mapping on $H(Q) \to H(G_1)$ which is one-to-one (cf. Exercise 12.20). The other way round, which is not needed, comes from the Künneth formula for polynomial cohomology. We shall come back to this in §14.6, below.

Be this as it may, what follows is that $H(G) < +\infty \Longrightarrow H(Q) < +\infty$. Since in the construction from §11.1.3 it is also true that Q is an NB-group if and only if G is, we see that the required implication (14.5) does hold for simply connected groups.

The above argument is typical of what we shall be doing in this chapter. To wit, we start from a group G and by a suitable procedure we construct some new group G_1, and in that construction, G_1 will be NB if and only if G is NB, and also the following implication will hold: $H(G) < +\infty \Longrightarrow H(G_1) < +\infty$. In our previous example of a simply connected group this new group could have been taken to be the model Q.

Once this is done it follows that if the implication (14.5) holds for G_1 it also holds for G. In other words, we have a reduction.

We shall see that it is possible to make a series of reductions like this, which will ultimately prove the implication (14.5) in full generality.

14.1.4 The pivotal reduction

This is contained in the following result.

Proposition 14.2 *Let G be some connected Lie group and let K be some connected compact normal subgroup. Then the following are equivalent*:

(i) $H(G; \mathrm{Pol}) < +\infty$;
(ii) $H(G/K; \mathrm{Pol}) < +\infty$.

This will be proved in §14.6 below and will be shown to hold in even greater generality when K is not necessarily normal and $G|K = [gK;\ g \in G]$ is the right homogeneous space (as we shall see, the definition of $H(G|K; \mathrm{Pol})$ easily extends; see also Example 11.7). The proposition also holds more generally for the case of a *bundle* with compact fibre: $K \subsetneq E \to B$ provided that natural conditions are imposed (see the next subsection and §14.2 below). A lot of background material has to be developed before we come to the proof. We shall do that in §§14.2, 14.4, 14.5 and in between in §14.3 we shall go back to structure theorems to show that once we have this type of reduction the required implication (14.5) follows.

In fact only the following special cases of the above proposition will actually be used:

(i) G is connected and soluble, and K is some normal – and therefore also central – torus (see §§2.6 and 11.3.1).

(ii) $K = F$ is some finite subgroup. Once more it is only the case when F is central that will be needed (see §11.1.5): the point is that the proposition still holds here despite the fact that F is not connected.

(iii) **The ultimate reduction**. This consists in showing, for the polynomial homologies, that

$$H(G) < +\infty \iff H(G/\Gamma) < +\infty$$

for any connected Lie group G and $\Gamma \lhd G$ some discrete central subgroup. We know a posteriori that this holds but we have not been able to give a direct proof. Had we been able to do so no other reduction would have been needed because this would reduce everything to the simply connected case.

(iv) **The 0-distorted case**. The implication that we suggested in (iii) will be proved in the special case when Γ is 0-distorted in G. We recall (see §§2.14.1, 5.7.1) that this means that the distance induced on Γ as a subset of G is equivalent to the intrinsic word distance of Γ. This will be postponed to Appendix G but it will not be an essential step to our final goal. Nonetheless, this is one of the most interesting aspects of this circle of ideas because it connects naturally with the cohomology of the discrete group Γ.

Remark In proving such reductions for a discrete central group Γ it is clear, by passing to intermediate subgroups $\Gamma \supset \Gamma' \supset \{0\}$, that we can assume that either Γ is finite (this is dealt with in §14.3) or $\Gamma = \mathbb{Z}$, which will be the subject matter of Appendix G. For this reduction recall that Γ is finitely generated and Abelian.

14.1.5 The methods and the background for the proofs

The proof of the proposition in §14.1.4 makes essential use and needs the full thrust of some basic concepts and ideas from algebraic topology.

To explain these we recall what a fibre bundle, $F \subsetneq E \xrightarrow{\pi} B$, is (see Steenrod, 1951; Hilton and Wylie, 1960; Bott and Tu, 1982). Here, π is a continuous surjective mapping of topological spaces such that each fibre $\pi^{-1}(x) \cong F$ for each $x \in B$, that is, they are homeomorphic. Furthermore, there exists $\mathfrak{U} = (U_\alpha;\ \alpha \in I)$, an open cover such that we have homeomorphisms $\pi^{-1}U_\alpha \xrightarrow[\cong]{\varphi_\alpha} U_\alpha \times F$ for which the natural fibres $\pi^{-1}(x)$ and $\{x\} \times F$ correspond. Here we shall be working throughout in the C^∞ category, that is, all the spaces are smooth. The open sets U_α are called the charts of the bundle, and the mappings φ_α are the corresponding trivialisations.

In fact, in this chapter we shall be exclusively in the special case of $G \to G/K$ (or $G|K$) for a connected Lie group and some closed subgroups K. When the subgroup K is not normal, $G|K$ denotes the homogeneous space (recall that $G|K$ is the space of cosets $(gK;\ g \in G)$ assigned with the canonical manifold structure; see Helgason, 1978, §II.4). These special cases fall under the more general class of principal bundles (see Greub et al., 1973, Chapter V). We recall that this means that F is some Lie group that acts on the 'total' space E and φ_α respects the natural group actions, that is, if $x \in \pi^{-1}y$ with $y \in U_\alpha$ and $f \in F$ then $f \cdot x \in E$ (the group action of f on E) is such that $f \cdot x \in \pi^{-1}y$ and $\varphi_\alpha(f \cdot x) = (y, f \cdot f_1)$ when $\varphi_\alpha(x) = (y, f_1)$, where $f_1 \in F$ and where $f \cdot f_1$ is the group multiplication.

The case that is important for the proposition is when, as above, $F = K$ is connected and compact.

At the other extreme $F = \Gamma$ is discrete and acts 'discretely' on E so that we have a covering space. This is the situation that we shall consider in Appendix G.

At this juncture let us completely forget about the Lie group structure, the possible Riemannian structure and the polynomial cohomology, but still work in the differential category of smooth manifolds. The fibre F will be assumed to be connected and we shall consider the classical de Rham cohomology of the spaces involved $H^*(X)$, $X = E, B, F$, that is, nothing polynomial or Riemannian here. The result from algebraic topology that will be pivotal for us is the following:

Assume that $H^*(F) < +\infty$; that is, finite Betti numbers. Then, under 'appropriate conditions', $H^*(E) < +\infty$ if and only if $H^*(B) < +\infty$.

For the 'appropriate conditions' on the fibre bundle we could demand, for instance, that B is simply connected.

We shall say more on this result in §14.6 below where our task will be to introduce the Riemannian structure on the fibre bundle and adapt the proofs to make them apply to our polynomial cohomologies $H^*(X; \mathrm{Pol})$.

14.1.6 About the style and the presentation of the chapter

The methods needed for the proof of the result on the homology of a fibre bundle are difficult but also central in algebraic topology. In fact, when they were discovered they marked a turning point in the subject.

A bit of history These are the methods that use the algebraic tool of spectral sequences. The discoverer of these methods was J. Leray and he did this when he was a prisoner of war in the period 1940–1945. It seems that the Germans respected the Geneva convention for French officers. I am sure, however, that had Leray been a Soviet officer, nobody would ever have heard either of him, or his spectral sequences.

At any rate, between 1945 and the late 1950s these methods were developed by some of the greatest mathematicians of the century: A. Borel, H. Cartan, J.-P. Serre and others, and the algebraic machinery that we now call spectral sequences was worked out by J. L. Koszul. The work of the great topologist H. Hopf was also decisive in these developments. Much of this was presented in the celebrated Séminaire Cartan, ENS, 1948–1959.

Be this as it may, we shall be needing all of this, and as a consequence, when writing this chapter I was faced with a dilemma:

- Should I give all the background material? Given the scope and the size of the book this would have been really out of the question.
- Or, should I address this chapter only to people with good knowledge of algebraic topology? Given who the average user of this book is likely to be, that would have cut the audience considerably.

The advantage of taking the second position would have been of course that the chapter would not have been a chapter but just a few pages.

I opted for the following hybrid policy. The presentation is on the one hand informal, and on the other more what one finds in a research paper rather than a book. This has been done in order to economise on space, and, since so much background is missing anyway, I felt this was a good compromise.

However, and partly to compensate, I have tried to place this chapter in the more general context of modern algebraic topology with the following elements: a number of digressions; a good guide to references in the subject; indications of alternative proofs depending on the aspects that one wishes to stress.

I also feel that this chapter, and especially Appendix G, contains the germ of further developments; and the presentation is designed to encourage readers to study and pursue these matters.

14.2 Notions from Algebraic Topology and Riemannian Geometry

Some basic definitions and facts from these subjects will be recalled in this section and these will be essential background material for the rest of the chapter.

14.2.1 Cohomology attached to a cover

14.2.1.1 Čech cohomology We denote $\mathfrak{U} = (U_\alpha, \alpha \in I)$ some open cover of the C^∞ manifold M where I is a finite or countable *ordered* index set. For multi-indices $\underline{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ that are increasing (i.e. $\alpha_0 < \alpha_1 < \cdots < \alpha_p$) and of length $|I| = p+1$, we shall use the notation $U_{\underline{\alpha}} = U_{\alpha_0} \cap \cdots \cap U_{\alpha_p}$ when that set is not empty. We shall associate with this $C^*(\mathfrak{U}) = C^*(\mathfrak{U}; M)$, the vector space $\prod_{\underline{\alpha}} \mathbb{R}_{\underline{\alpha}}$, where all the $\mathbb{R}_{\underline{\alpha}} = \mathbb{R}$ are the reals, that is, $C^* = (c_{\underline{\alpha}} \in \mathbb{R};\ \underline{\alpha})$. This is the space of Čech cochains of the cover.

We denote by C^p, with $p \geqslant 0$, the cochains of length $p+1$, that is, $c_{\underline{\beta}} = 0$ unless the length of $\underline{\beta} = p+1$; we have $C^* = \sum_{p \geqslant 0} C^p$. We can then define the following differential $\delta : C^p \to C^{p+1}$ (i.e. a linear map with $\delta^2 = 0$):

$$(\delta c)_{\alpha_0, \ldots, \alpha_{p+1}} = \sum_j (-1)^j c_{\alpha_0, \ldots, \hat{\alpha}_j, \ldots, \alpha_{p+1}}; \quad c = (c_{\underline{\alpha}}) \in C^*. \tag{14.6}$$

Here, $c, \delta c \in C^*$ and the indices, for example $(\delta c)_{\underline{\alpha}}$, indicate the $\underline{\alpha}$ coordinate of δc. As usual, the hat ˆ indicates that the term is omitted. That this is a differential can be easily verified and so we obtain a graded differential complex

$$0 \to C^0 \underset{\delta}{\to} C^1 \underset{\delta}{\to} C^2 \to \cdots \to C^p \to \cdots. \tag{14.7}$$

The homology of this complex $H^*(\mathfrak{U}) = H^*(\mathfrak{U}, M)$ is the Čech cohomology of the cover $\mathfrak{U}$ on M.

14.2.1.2 Presheaf cohomology Čech cohomology can be generalised into what is called presheaf cohomology $H^*(\mathfrak{F})$ for some presheaf

$$\mathfrak{F} = \big(\mathfrak{F}(U_{\underline{\alpha}});\ \alpha_0 < \alpha_1 < \cdots < \alpha_p,\ p \geqslant 0\big).$$

First of all $\mathfrak{F}(U_{\underline{\alpha}})$ denotes, for all $U = U_{\underline{\alpha}}$, some vector space (other objects, e.g. Abelian groups, or more general modules over some ring, could be considered) and the corresponding presheaf cochain $C^*(\mathfrak{F})$ is then the vector space $\prod_U c_U$, with $c_U \in \mathfrak{F}(U)$, and where here U runs through the set of all $U_{\underline{\alpha}}$, as above. The corresponding differential complex is then defined as (14.6), with C^p defined as before and where, in this definition of the differential, we replace the right-hand side with $\sum_j(-1)^j i_j(c_{\alpha_0,\dots,\hat{\alpha}_j,\dots,\alpha_{p+1}})$; here i_j is an appropriate linear mapping

$$i_U^V \colon \mathfrak{F}(U) \to \mathfrak{F}(V); \quad V = U_{\alpha_0,\dots,\alpha_{p+1}},\ U = U_{\alpha_0,\dots,\hat{\alpha}_j,\dots,\alpha_{p+1}}. \tag{14.8}$$

Now we explain how the mappings in (14.8) are defined and for this we shall start with the easier case when $\mathfrak{F}(U)$ is actually defined for all the open sets $U \subset M$. One then simply demands

$$i_U^U = \text{Identity}, \quad i_U^W = i_V^W \circ i_U^V; \quad W \subset V \subset U. \tag{14.9}$$

This condition qualifies $\mathfrak{F}$ to be a functor that is defined in the full category Open(M) and not only on the subcategory $(U_{\underline{\alpha}})$ as in (14.8).

Example 14.3 We could set $\mathfrak{F}(U) = \Omega^*(U)$ or $\mathfrak{F}_0(U) \simeq \mathbb{R}$, the constant functions on U which is $\cong \mathbb{R}$ (the reals). In these cases we set i_U^V the restriction mapping. For $\mathfrak{F}_0$ we of course obtain back the Čech cohomology of (14.7).

For the presheafs that we are considering in this chapter, an open cover $\mathfrak{U}$ will be considered and only the values $\mathfrak{F}(U)$ for $U = U_{\underline{\alpha}}$ will be involved. But this functor will be the restriction of a functor that is defined on the whole Open(M) with condition (14.9) on the i_U^V. We shall therefore stop here and not give the more sophisticated definition for functors only defined on the subcategory (14.8); see Bott and Tu (1982, §13).

Remark In fact, what will happen in §14.2.3 below is this. All the sets $U_{\underline{\alpha}} \cong \mathbb{R}^n$ (diffeomorphic) and the functor $\mathfrak{F}(U_{\underline{\alpha}})$ will be the restriction of a functor $\mathfrak{F}(V)$ defined on (V open in M, $V \cong \mathbb{R}^n$ and V contained in some U_α) with condition (14.9) satisfied for the mappings between these sets.

Once the δ has been defined on $C^*(\mathfrak{F})$ we easily verify that $\delta^2 = 0$ and we therefore have a complex. (To stress the presence of the cover we could denote this by $C^*(\mathfrak{U};\mathfrak{F})$). The homology of this complex is denoted $H^*(\mathfrak{U};\mathfrak{F})$ and is called the cohomology of the cover $\mathfrak{U}$ with values in the presheaf $\mathfrak{F}$.

14.2.1.3 Constant and locally constant presheafs When all the $\mathfrak{F}(U) = L$ are identical and all the i_U^V are the identity mappings we say that the sheaf is the trivial presheaf L (where L can be a vector space, an Abelian group, etc.). For instance, the original Čech cohomology is defined on the trivial presheaf $L = \mathbb{R}$.

When the presheaf $\mathfrak{F}$ is isomorphic to a trivial presheaf, we say that $\mathfrak{F}$ is constant. A homomorphism between two presheafs $(\mathfrak{F}; i) \to (\mathfrak{G}; j)$ is a family of linear mappings $\varphi_U : \mathfrak{F}(U) \to \mathfrak{G}(U)$ that intertwine the defining homomorphisms i and j (i.e. for $V \subset U$ as above, we have $\varphi_V \circ i_U^V = j_U^V \circ \varphi_U$).

As long as we are in the setting of the above remark, there is a very important halfway notion for a presheaf that demands that all the $\mathfrak{F}(U) \simeq L$ are isomorphic to some fixed vector space and all the i_U^V are isomorphisms. Such a presheaf on the cover $\mathfrak{U}$ is called *locally constant*.

Exercise Show that a presheaf could be locally constant without being constant. Simple examples can be found in Bott and Tu (1982, §10.7).

Exercise Show that for a presheaf to be constant we need to be able to specify isomorphisms $\theta_U : \mathfrak{F}(U) \underset{\cong}{\rightarrow} L$ such that $\lambda = \theta_V \circ i_U^V \circ \theta_U^{-1}$ is the identity. In the case of a locally constant presheaf we have analogous mappings but then the isomorphism λ is not necessarily the identity (cf. the proofs of Bott and Tu, 1982, Theorem 13.2 and Example 13.5).

Note The reader is urged to elaborate these definitions further and consult at least the rapid review of the subject that can be found in most books on algebraic topology, for example Bott and Tu (1982, §§10,13). A comprehensive but very readable account can be found in Godement (1958). For instance, the setting of the above remark is artificial and the correct notion of a locally constant presheaf on an open cover is elaborated on further in Bott and Tu (1982, §13, p. 142). To do this one has to use an abstract simplicial complex called the nerve of the cover.

14.2.2 Good covers, notation and fundamental facts

Both M and $\mathfrak{U}$ are as before. We say that $\mathfrak{U}$ is a *good* cover if all the sets $U_{\underline{\alpha}}$ (that by definition are not empty open sets) are diffeomorphic with $\mathbb{R}^n$, $n = \dim M$. Here are some basic facts.

First of all, every manifold admits some good cover. At the end of this section we shall see how this is done.

Monodromy If M is a simply connected manifold and if $\mathfrak{U}$ is a good cover of M then any locally constant presheaf attached to $\mathfrak{U}$ is constant.

A proof of this can be found in Bott and Tu (1982, §13). In the same book, several proofs, some direct, some less so, can be found for the following basic fact:

Let $M, \mathfrak{U}$ be as above with $\mathfrak{U}$ some good cover. Then we have an isomorphism between the de Rham cohomology and the Čech cohomology with respect to the cover $\mathfrak{U}$.

In symbols, we have $H^*_{\mathrm{DR}} \cong H^*(\mathfrak{U})$.

One of these proofs, not the most direct, but one that uses spectral sequences, will be reproduced in §14.5 below. The reader who is serious about the material of this chapter is, however, urged to study a more direct proof as given in Bott and Tu (1982, §8).

For later use the following notions and notation that are related to a cover $\mathfrak{U}$ on the manifold will be specified.

One is that of a partition of unity (ρ_α; $\alpha \in I$) that is subordinated to the cover. These are $\rho_\alpha \in C_0^\infty(U_\alpha)$ for each $\alpha \in I$ and such that $\sum_{\alpha\in I} \rho_\alpha = 1$ and $\operatorname{supp} \rho_\alpha \subset U_\alpha$.

The other is the retract $h_{\underline{\alpha}}(x,t) \in U_{\underline{\alpha}}$, with $x \in U_{\underline{\alpha}}$, $0 \leqslant t \leqslant 1$. These are smooth and defined for all $\underline{\alpha}$, not just the elements $\alpha \in I$ in the index set, and they retract the set $U_{\underline{\alpha}}$ to some fixed point $x_{\underline{\alpha}} = h_{\underline{\alpha}}(x,0)$, for $x \in U_{\underline{\alpha}}$, while for $t = 1$ we have $h_{\underline{\alpha}}(x,1) = x$ for all $x \in U_{\underline{\alpha}}$, and also $h_{\underline{\alpha}}(x_{\underline{\alpha}},t) = x_{\underline{\alpha}}$, with $0 \leqslant t \leqslant 1$. These clearly only exist if the cover $\mathfrak{U}$ is a good cover.

14.2.2.1 A homotopy operator A good reference for this material is Bott and Tu (1982, §8.2). Let $\mathfrak{U} = (U_\alpha)$ be some open cover that is not necessarily a good cover, and let $\mathfrak{F} = \Omega^*(U_{\underline{\alpha}})$ be the presheaf of Example 14.3. For the decomposition $C^*(\mathfrak{U};\mathfrak{F}) = \sum C^p$ let

$$0 \to \Omega^*(M) \underset{r}{\to} C^0 \underset{\delta}{\to} \cdots \underset{\delta}{\to} C^p \to \cdots$$

be the corresponding complex with δ as in (14.7) and r induced by the inclusions $U_\alpha \subset M$, that is, the restriction mappings. It is clear that $\operatorname{Ker} r = 0$ and $\operatorname{Im} r = \operatorname{Ker}[\delta\colon C^0 \to C^1]$.

We can also construct a 'homotopy operator' $K\colon C^p \to C^{p-1}$, for $p \geqslant 1$, by the formula

$$(K\omega)_{\alpha_0\cdots\alpha_{p-1}} = \sum_\alpha \rho_\alpha \omega_{\alpha\alpha_0\cdots\alpha_{p-1}} \tag{14.10}$$

for the partition of unity of §14.2.2. In this formula we use the convention that when two indices are interchanged the form becomes its negative:

$$\omega_{\cdots\alpha\cdots\beta\cdots} = -\omega_{\cdots\beta\cdots\alpha\cdots} \tag{14.11}$$

and therefore in the indices of $\omega_{\alpha_0\cdots\alpha_p}$ we no longer have to assume that $\alpha_0 < \cdots < \alpha_p$.

By a routine verification we now see that (see Bott and Tu, 1982, §8.5)

$$\delta K + K\delta = \text{Identity}. \tag{14.12}$$

From these facts it follows that the homology of the above complex vanishes.

Remark The 'convention' (14.11) comes straight out of Bott and Tu (1982, Exercise 8.4), where an informal presentation of the material in hand is given.

We follow that presentation closely and make it even more informal! For our purposes this is the best way of proceeding, no doubt about it. There are places, however (e.g. §§14.2.4.4 and G.1.2), where more formal definitions would clarify matters; see standard references on algebraic topology, for example Massey (1991) or Godement (1958), for these.

14.2.3 The cohomology presheaf of a bundle

We recall that we are working throughout in the C^∞ category and in what follows in this subsection we shall be using the de Rham cohomology. We shall also consider a fibre bundle $F \hookrightarrow E \underset{\pi}{\to} B$ and on every open set $U \subset B$ we shall define $\mathfrak{F}(U) = H^*(\pi^{-1}U)$ so that now $i_U^V \colon \mathfrak{F}(U) \to \mathfrak{F}(V)$ for $V \subset U$ is induced by the inclusion $\pi^{-1}V \subset \pi^{-1}U$.

We shall now fix $\mathfrak{U} = (U_\alpha)$ a good cover of B and restrict the definition of $\mathfrak{F}$ to the subcategory of open sets of the form $U_{\alpha_0\cdots\alpha_p}$ as in §14.2.1. We obtain therefore a presheaf attached to the cover $\mathfrak{U}$; this presheaf is locally constant since any two subsets $V \subset U$ as above are contractible and therefore for any $x \in V$ we have

$$H^*(\pi^{-1}U) \cong H^*(\pi^{-1}V) \cong H^*(\pi^{-1}x) \cong H^*(F). \tag{14.13}$$

To see this we use the standard fact that a bundle with a contractible base space is trivial (i.e. a product; see Steenrod, 1951, §11.6). In everything that follows, however, all the sets U will always be chosen so small that $\pi^{-1}U \cong U \times F$ is a chart of the bundle, so that we do not have to worry about this.

It follows in particular that this presheaf is constant when B is simply connected (cf. §14.2.2). On the other hand, one can easily verify (see Bott and Tu, 1982, Examples 10.1 and 13.1) that when the fibre bundle is a Möbius strip this presheaf is not constant.

Proposition *Let the fibre bundle be given by the homogeneous space $G \underset{\pi}{\to} G|K$ where G is a connected Lie group and K is a connected closed subgroup. Further, let $\mathfrak{U}$ be some good cover of $G|K$ as above; then the above presheaf is constant.*

To see this, let us fix both $x \in U_{\alpha_0\cdots\alpha_p} = U_{\underline{\alpha}}$, and $g \in \pi^{-1}(x)$. Then we have an isomorphism $H^*(K) \to H^*(\pi^{-1}U_{\underline{\alpha}})$ that is induced by $\dot{g} \colon K \to \pi^{-1}(x)$, $k \mapsto gk$, and the inclusion $\pi^{-1}(x) \to \pi^{-1}U_{\underline{\alpha}}$. For the proof it suffices to show that this isomorphism is independent of the two choices of x and g (see the following exercise). Furthermore, this argument extends to all principal bundles.

Exercise (See also Greub et al., 1973, Chapter V.) For some fixed $x \in G|K$

and two different choices $g_1, g_2 \in \pi^{-1}x$ the corresponding mappings $\dot{g}_i$ differ by group multiplication on K, $\dot{k}_0\colon k \to k_0 k$, $k \in K$, with $k_0 = g_1^{-1} g_2$. And if $k_0(t) \in K, 0 \leqslant t \leqslant 1$ is a path that joins e to k_0, the mappings $k \to k_0(t)k$ give a homotopy between $\dot{k}_0$ and the identity.

Similarly, for the two choices $x_1, x_2 \in U_{\underline{\alpha}}$, by considering some trivialisation we can find $g_i \in \pi^{-1}(x_i)$ such that the two mappings from $K \to \pi^{-1}U_{\underline{\alpha}}$, namely $\dot{g}_i\colon K \to \pi^{-1}x_i \subset \pi^{-1}U_{\underline{\alpha}}$, are homotopic.

14.2.4 How to construct a good cover

14.2.4.1 The set-up In this subsection we shall show how to construct a good cover on a connected Lie group G and also on a homogeneous space $G|K$, where $K \subset G$ is some compact subgroup.

Some of the special features of the constructions below are difficult to adapt to a general C^∞ manifold M but the essential properties of the construction go through for M and even for more general topological spaces (see Bott and Tu, 1982, §§5, 13, p. 147). At any rate, in this chapter the only manifolds that we shall be considering will be $M = G$ or the homogeneous space $G|K$, as in §14.1.4.

We shall fix once and for all some left-invariant Riemannian structure on G. Once this structure is chosen also to be K-right invariant (this is always possible when K is compact because it amounts to choosing the Riemannian scalar product $\langle \cdot, \cdot \rangle$ on $T_e(G)$ to be invariant by the compact $\operatorname{Ad} K$ action on the Lie algebra $\mathfrak{g}$; this can be done by taking the average $\int_K \langle (\operatorname{Ad} k)\xi, (\operatorname{Ad} k)\zeta \rangle dk)$, then the Riemannian structure on G induces canonically a G-invariant Riemannian structure on $G|K$ (i.e. invariant under $\dot{x} \to g\dot{x}$, $\dot{x} \in G|K$, $g \in G$, where, when $\dot{x} = xK$ is the coset, $g\dot{x}$ is the coset gxK). The way this is done is as follows: let $\mathfrak{k} \subset \mathfrak{g}$ be the Lie subalgebra of K and let $\mathfrak{k}^\perp$ be the orthogonal complement with respect to the Riemannian scalar product. This subspace can be identified with $T_{\dot{e}}(G|K)$, the tangent space at the identity class $\dot{e} = K \in G|K$. This defines a scalar product on that space and one can easily verify that the G-group action defines unambiguously a global Riemannian structure on $G|K$ (see Kobayashi and Nomizu, 1963, Example IV.1.3; Kobayashi and Nomizu, 1969, Chapter X; Helgason, 1984, §II.4, p. 284).

In fact, for most of our purposes and certainly for all the essential ones, K is a compact normal subgroup (even central!) and then G/K is a Lie group. In that case the above Riemannian structure is (up to equivalence) just the left-invariant structure of the group G/K.

14.2.4.2 The Whitehead lemma and geodesically convex sets in Riemannian manifolds Here we shall explain and comment on some well-known facts from local Riemannian geometry.

We recall first what it means to say that $U \subset M$, some open set in a (complete) Riemannian manifold, is geodesically convex. For any two points $x, y \in U$ there exists $\gamma(t) \subset M$, $t \geqslant 0$, some geodesic in M that joins x to y, with the following properties:

(i) *The geodesic $\gamma \subset U$ and no other geodesic that joins x to y lies entirely in U.*
(ii) *The geodesic γ gives distance; that is, the length of γ equals the Riemannian distance between x and y. Such a geodesic is called a minimising geodesic. Furthermore, γ is the only minimising geodesic between x and y in M.*

What is clear is that if $U, V \subset M$ are geodesically convex then so is $U \cap V$, and since we expect these sets to be diffeomorphic to $\mathbb{R}^n$, we see that these sets can be used to give good covers on a manifold (see Bott and Tu, 1982, §5).

The Whitehead lemma is as follows.

Lemma 14.4 (Whitehead lemma) *For every $x_0 \in M$ there exists ρ_0 such that for all $0 < \rho < \rho_0$ the Riemannian open ball $U = B_\rho(x_0)$ centred at x_0 with radius ρ is geodesically convex.*

It is also clear (at least if ρ_0 is small enough) that the above sets U are normal neighbourhoods of any of their points. More explicitly, we use here the fact that for all $x_0 \in M$ there exists $\rho_1 > 0$ such that for all $x \in M$ with $\mathrm{d}(x_0, x) < \rho_1$, the set $[\xi \in T_x(M);\ |\xi| < \rho_1]$, that is, the ball of radius ρ_1 on the tangent space, is mapped bijectively and smoothly by $\mathrm{Exp}_x \colon T_x(M) \to M$ onto what is called a normal neighbourhood of $x \in M$ and the coordinates that we obtain near x by $\mathrm{Exp}_x \colon \mathbb{R}^n \to M$ are called normal coordinates. Here, Exp_x denotes the mapping that sends every vector $\xi \in T_x(M)$ small enough, to $\mathrm{Exp}(\xi) \in M$ at distance $|\xi|$ from x and lying on the geodesic from x in the direction ξ. Furthermore, the geodesic from x to $\mathrm{Exp}(\xi)$ is minimising.

These well-known facts from Riemannian geometry can be found in standard books on the subject, for example Kobayashi and Nomizu (1963, §3.8) or Hicks (1971, §9.4).

It follows therefore, as asserted (at least if ρ_0 is small enough), that the geodesically convex sets that we constructed, U, are diffeomorphic with $\mathbb{R}^n$. Furthermore, they are ‘geodesically star shaped’ with respect to any of their points.

More generally, if $x_0, x_1, \ldots, x_k$ are finitely many points (and ρ_0 is small enough to work simultaneously for all the convex subsets $U_j = B_\rho(x_j)$), then the intersection $U = U_0 \cap \cdots \cap U_k$ is diffeomorphic to $\mathbb{R}^n$ when not empty.

Let us elaborate more on this finite intersection so that we do not have to come back later. Provided that ρ_0 as above is small enough, this is what can be said:

> The set U could certainly have the nasty shape of a long thin lens, the intersection of two discs that barely overlap. But whatever the shape might be, U is 'geodesically star shaped', with centre any of its points $\bar{x} \in U$.

More formally we can give normal coordinates for U with zero at $\bar{x}$ and these identify U with a star-shaped domain in $\mathbb{R}^n$ centred at the origin. This follows from what we have explained. These normal coordinates can then be used to define a smooth retract $h(x,t) \in U$ with $h(x,0) = \bar{x}$, $h(x,1) = x$ and $h(\bar{x},t) = \bar{x}$, for $x \in U$, $0 \leqslant t \leqslant 1$. Furthermore, the gradient (both in x and t) satisfies $|\nabla h| \leqslant C$, where C depends only on x_0, ρ_0 and the Riemannian structure of the manifold in some neighbourhood of x_0. A special case, where we do not have to worry about this Riemannian structure in the dependence of $|\nabla h| \leqslant C$, is when M is a Lie group with its left-invariant Riemannian structure, or more generally the homogeneous spaces $G|K$ that we considered earlier in this section. This holds because this structure is homogeneous and is the same as we move from one point to another. This will be used in the construction of the good cover that we shall use in the next subsection.

The essential exploitation that will be made of this retract h for U is the following. Let $\omega \in \Omega^*(U)$ be some closed form without constant term (i.e. $\mathrm{d}\omega = 0$, $\omega = \omega_1 + \cdots$); then the Poincaré equation $\mathrm{d}\theta = \omega$ can be solved with $\theta \in \Omega^*(U)$ with control on the uniform norm $\|\theta\|_\infty = \sup_{x \in U} |\theta(x)|$. That is, $\|\theta\|_\infty \leqslant C_1 \|\omega\|_\infty$, where C_1 depends only on C (and a priori also possibly on the Riemannian structure on U – but in our case, for Lie groups, it will essentially depend only on the group).

Exercise 14.5 Verify the above. This is sometimes referred to as the Poincaré lemma (see de Rham, 1960, lemma in §19). One elementary way of proceeding is to use the previous observation so that, for each U as above, there exists U^*, some star-shaped domain around the origin in the Euclidean space, and a diffeomorphism $F\colon U^* \to U$ that, together with its inverse F^{-1}, has bounded gradients (depending again on ρ_0 and the Riemannian geometry). This allows us to transport the problem to $\mathbb{R}^n$. Then we can use well-known formulas (see Warner, 1971, §4.18) that solve the Poincaré equation in $\mathbb{R}^n$. Alternatively, we can do what we did in §12.9 with polynomial homotopy. The homotopy that we use here is h and is in fact 'bounded'.

14.2.4.3 Construction of the good cover We shall deal only with the case for a group G. The construction for $G|K$ is identical and will be left to the reader.

This good cover on G depends on the following construction. We fix some $c_0 > 0$ and then choose a sequence of points $g_1, g_2, \ldots \in G$ that satisfy the condition $\mathrm{d}(g_i, g_j) > c_0$ for $i \neq j$ for the left-invariant distance, and are such that if we add one additional point this condition no longer holds, that is, the sequence is maximal under our condition. Zorn's lemma can be used for the construction. This means that the balls of radius $10c_0$, centred at the points g_i, give a cover of G. These balls are $U_\alpha = g_\alpha U$ where $U = (|g| < 10c_0)$.

By its very construction this cover has a number of special properties. In particular, a partition of unity (ρ_α), as in §14.2.2, subordinated to the cover, can be constructed so that it is of *bounded gradient*. More precisely, $|\nabla\rho_\alpha| \leqslant C$ uniformly on α (here C depends on G and c_0). Also, such a cover is *locally finite* and no more than C balls (depending on G and c_0) can have non-empty intersection. The above also imply that the $|\rho_\alpha|$ are also uniformly bounded.

Exercise 14.6 Verify this as follows. If C balls intersect, the C disjoint sub-balls $g_\alpha U'$, with $U' = [|g| < c_0]$, are close together. Then use the Haar measure to bound C.

Furthermore, if we use the Whitehead lemma (Lemma 14.4) and the properties enumerated in the previous subsection, we see that the above cover $\mathfrak{U} = (U_\alpha)$ is a good cover and also the uniformly bounded retracts $h_{\underline{\alpha}}$ of the $U_{\underline{\alpha}}$ (with $\underline{\alpha} = (\alpha_0, \ldots, \alpha_p)$, as in §14.2.1) can be constructed. For all of these it suffices to choose $c_0 > 0$ small enough so that in particular the Poincaré lemma, from Exercise 14.5, applies.

14.2.4.4 An illustration for the covering spaces We shall consider in Appendix G a covering map $\widetilde{G} \underset{\pi}{\to} G$ between two connected Lie groups. More precisely, $G = \widetilde{G}/\Gamma$ and $\Gamma = \ker\pi$ is a discrete central subgroup. It is then clear that if the good cover $\mathfrak{U} = (U_\alpha)$ that we constructed on G in the previous subsection is fine enough (i.e. c_0 is small enough) then we can lift it to a similar good cover $\widetilde{\mathfrak{U}} = (\pi^{-1}U_\alpha)$ where each $\pi^{-1}U_\alpha$ breaks up into disjoint balls and the deck transformation group Γ is in one-to-one correspondence with these balls, that is, $\pi^{-1}(U_\alpha) = (\gamma\widetilde{U}_\alpha;\ \gamma \in \Gamma)$, for some fixed $\widetilde{U}_\alpha$. Furthermore, as long as the diameter of the original ball U in our construction has been chosen sufficiently small, it is clear that $\mathrm{d}(\gamma_1\widetilde{U}_\alpha, \gamma_2\widetilde{U}_\alpha) \geqslant c$ for $\gamma_1 \neq \gamma_2$ and some $c > 0$ independent of γ_1, γ_2 and α. A consequence of this is that, when $U_{\alpha_1} \cap U_{\alpha_2} \neq \emptyset$, then for all $\gamma_1 \in \Gamma$ there exists one and only one γ_2 such that $\gamma_1\widetilde{U}_{\alpha_1} \cap \gamma_2\widetilde{U}_{\alpha_2} \neq \emptyset$.

The connection between $H^*(\mathfrak{U})$ and $H^*(\widetilde{\mathfrak{U}})$ is a very important issue (cf. McCleary, 2001, §8$^{\text{bis}}$.9) and it will also be the subject matter of Appendix G.

14.2.5 The polynomial Čech cohomology

In this final subsection the new notion of the polynomial Čech complex will be introduced. We go back to the manifold M and the open cover $\mathfrak{U} = (U_\alpha)$ of §14.2.1 and now a base point $O \in M$ will be fixed and a Riemannian structure will be introduced and we shall assume that $\operatorname{diam}(U_\alpha) \leqslant C$ for some C uniformly in α.

The polynomial Čech subcomplex of the corresponding Čech complex consists of the chains $c \in C^*(\mathfrak{U})$ of §14.2.1 for which there exist constants $C > 0$, with C depending on the particular chain, such that

$$|c_{\underline{\alpha}}| \leqslant C\big(1 + \operatorname{dist}(O, U_{\underline{\alpha}})\big)^C. \tag{14.14}$$

These are polynomially growing chains and they clearly form a subcomplex of $C^*(\mathfrak{U})$ under the differential δ provided that the cover is uniformly *locally finite*, as in Exercise 14.6. We shall denote this subcomplex by $C^*_{\mathrm{P}}(\mathfrak{U})$ and its cohomology will be the polynomial Čech cohomology of the cover $\mathfrak{U}$ and will be denoted by $H^*_{\mathrm{P}}(\mathfrak{U})$.

In §14.5 below we shall show that, provided the cover $\mathfrak{U}$ satisfies the conditions of §14.2.4.3, we have $H^*_{\mathrm{DR}}(M; \mathrm{Pol.}) \cong H^*_{\mathrm{P}}(\mathfrak{U})$. For this we shall use the spectral sequences of §14.4. But for the reader who followed our suggestion in §14.2.2 to study Bott and Tu (1982, §8), that is, the 'elementary' proof of $H^*_{\mathrm{DR}} \cong H^*(\mathfrak{U})$ without using spectral sequences, there will be no difficulty in being convinced that the same proof with easy modifications will also give this corresponding fact for polynomial cohomologies.

Note The reader must not get the impression that we have, all of a sudden, fallen in love with spectral sequences. Anything but! But they *have* to be used for the main result of this chapter and the use one makes of the isomorphism $H^*_{\mathrm{DR}}(M; \mathrm{Pol.}) \cong H^*_{\mathrm{P}}(\mathfrak{U})$ offers a very instructive illustration. Apart from this, the spectral-sequence-free approach that we have urged the reader to work out is preferable from every point of view. That proof is due to A. Weil and it played an important role in the development of the subject. See the footnote at the end of the preface in Godement (1958). This comes from someone who was a young man when it was all happening in École Normale Supérieure in the late 1940s; cf. §14.1.6.

14.3 Revisiting Structure Theorems

In view of the reductions that we explained in §14.1 we have to go back and recycle some of the structure theorems of Chapter 11.

14.3.1 Soluble groups

The critical reductions of §14.1.4 depend on what we shall recall here. Let G be some connected soluble group and let T denote its maximal compact central torus (see §§11.3.1 and F.3, especially (F.6)). Then we have the following commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow[\pi]{} & G_1 = M \leftthreetimes K \\ & \alpha \searrow \quad \swarrow \beta & \\ & G_1|K. & \end{array} \tag{14.15}$$

Here π is the canonical group projection on $G_1 = G/T$ and M is a model (i.e. a simply connected soluble group) and K is compact. Thus π is a group homomorphism and the other two mappings, α and β, give canonical mappings on the homogeneous space $G_1|K = G|KT$ and $\alpha = \beta \circ \pi$ (see §14.2). The two mappings α, β are used to give an independence proof and they may be ignored for the first approach that we give.

This first approach consists in proving that

$$H(G) < +\infty \implies H(G_1) < +\infty. \tag{14.16}$$

We are using here the terminology of §14.1 and talking of polynomial de Rham cohomology of course. Once we have proved (14.16) we are essentially done. The reason is that we have a quasi-isometry $G_1 \approx M \times K$ and this, as we pointed out in Exercise 12.20, implies that

$$H(G_1) < +\infty \implies H(M) < +\infty. \tag{14.17}$$

On the other hand, both the groups G_1 and M are NC if and only if G is. To see this, for G_1 use the fact that T is central and §2.2.1. As for M, use the fact that M is normal in G_1 and since the action of $\mathrm{Ad}(K)$ on the Lie algebra of M is compact, it has unimodular roots (this is obvious; but in case you have a problem, go back to the linear algebra of Part I, e.g. §§3.4.4, 3.9.1). Notice that in the construction (14.15), the nilradical of G_1 lies in M and therefore any of the algebraic characterisations of the NC-condition of Chapter 2 can be used. (Explicitly, we use the Lie algebra $\mathfrak{g}_1 = \mathfrak{m} \leftthreetimes \mathfrak{k}$ and Varadarajan, 1974, §3.7.3 to triangulate the action $\mathrm{ad}\,\mathfrak{g}_1$ on $\mathfrak{m} \otimes \mathbb{C}$, then the real parts of the roots vanish on $\mathrm{ad}\,\mathfrak{k}$ because of the above remark. Since $\mathfrak{k}$ is Abelian, we obtain thus a triangulation of the $\mathrm{ad}\,\mathfrak{g}_1$ action on $\mathfrak{g}_1 \otimes \mathbb{C}$ on which we apply §2.3.3.

We have therefore the reduction of our problem, as explained in §14.1.3, from an arbitrary connected soluble group G to a model M.

The alternative approach consists in using the two mappings α, β in (14.15) – observe that these both give fibrations with fibres homeomorphic to a torus because these fibres are connected compact soluble groups (see Hochschild, 1965, XIII, §1.3). We arrive then at the same conclusion but now by proving

$$H(G)<+\infty \implies H(G_1|K)<+\infty, \tag{14.18}$$

$$H(G_1|K)<+\infty \implies H(G_1)<+\infty \tag{14.19}$$

for the corresponding polynomial cohomologies. The polynomial cohomology on $G_1|K$ which is not necessarily a Lie group is given by the definition on a general Riemannian manifold (see §14.1.1 or Chapter 12).

This approach is at first sight more involved. However, here the advantage is that the base spaces of the fibrations are simply connected and therefore in the proofs that we give in the next section we could use the monodromy of §14.2.2. The proposition of §14.2.3 in is particular not needed.

The reductions (14.16), (14.18), (14.19) that we described above will be proved in the next three sections: see (14.64).

14.3.2 The endgame

Once we have the reductions of the previous subsection we can put things together and play the endgame in the proof of (14.5) as follows.

Let G be some arbitrary connected Lie group. We can find some new connected Lie group $\widetilde{G}$ with the following two properties: $G \cong \widetilde{G}/F$ (group isomorphism) for some finite central subgroup F, and $\widetilde{G} \simeq Q \times K$ (a quasi-isometry) where Q is soluble, connected and K compact. Furthermore, Q is a C-group if G is a B-group (and vice versa – but this is not needed). This construction was carried out in §11.1.5.

Now it is easy to see that if we assume that $H(G) < +\infty$ then it follows that $H(\widetilde{G}) < +\infty$ too (see the exercise below) and from this it also follows that $H(Q) < +\infty$ as we have already observed (see the remark in §14.1.3). Since Q is soluble the reductions of the previous subsection apply and it follows from Theorem 14.1 that Q has to be an NC-group. Hence G is an NB-group; in other words we have a proof of the implication (14.5) and our goal has been achieved.

Exercise Here G is an arbitrary connected Lie group and $F \subset G$ a finite subgroup (in our previous considerations, $F \subset \widetilde{G}$ was a central subgroup of $\widetilde{G}$). Go through the arguments below and verify them.

Let $\Omega_I(G)$ denote the differential forms on G that are invariant under right translation by elements of F (recall that in our case, F is central). Show that the cohomology of this complex is identical to the cohomology of $\Omega(G|F)$, and that the same holds for the complexes of forms of polynomial growth (for more on this see Greub et al., 1973, 1976).This is evident because the projection $\pi\colon G \to G|F$ induces an isomorphism between $\Omega(G|F)$ and $\Omega_I(G)$. Now the injection $\Omega_I(G) \to \Omega(G)$ clearly induces $\alpha\colon H^*(\Omega_I) \to H^*(\Omega)$. We shall show that this is an isomorphism.

To see that α is one-to-one we use the average operator $\overline{\omega} = \frac{1}{|F|}\sum_{x\in F} \tau_x^* \omega$, with $\omega \in \Omega(G)$, where $\tau_x\colon g \to gx$ is the *right* translation. From this we conclude, for some $\omega \in \Omega_I(G)$, that if we can solve $\omega = \mathrm{d}\theta$ for some $\theta \in \Omega(G)$ we can also choose this θ to belong to Ω_I.

To prove that α is onto, we start from some closed $\omega \in \Omega(G)$ and then show that for all $x \in F$, we have $\omega \sim \tau_x^*\omega$ are cohomologous (i.e. give the same cohomology class). From this we are done because this shows that $\omega \sim \overline{\omega} \in \Omega_I(G)$.

This last point is slightly less formal and to prove it we must use the connectedness of G. Indeed, this implies that all the mappings $\tau_y\colon g \to gy$, with $y \in G$ fixed, and in particular all the mappings τ_x, with $x \in F$, are polynomially homotopic to the identity. We have already observed this fact in the proof of the proposition in §14.2.3: the homology is supplied by $g \to gy(t)$ where $y(t) \in G$, $0 \leqslant t \leqslant 1$ is some path that joins y to the identity (compare with Exercise 12.30).

A digression This will not be needed in this chapter but is interesting because it brings out an aspect of the polynomial homology where it differs from the classical theory. This is the fact that if we define τ_y by *left* translations the same conclusion in general no longer holds. One can already see this on the group $\mathbb{R}^2 \leftthreetimes \mathbb{R}$ and the differential forms of Example 7.16 by arguing on the inner automorphisms $g \to y^{-1}gy$. But to write the whole thing out is quite long (though it is 'fun'!). The details will not be given below, but for the reader who wants to try their hand, here is a hint:

- Use the notation of Example 7.16 and set $F(x_1,x_2) = \sum j^{-2}\chi\left(\frac{x_1}{R_j}, \frac{x_2}{R_j}\right)$, where χ is the characteristic function of the unit square and $R_1 < R_2 < \cdots$ is rapidly increasing. Then $\mathbb{R}$ acts by inner automorphism on $\mathbb{R}^2$: denote that action by I_y, with $y \in \mathbb{R}$.
- Let $\omega = F(x_1,x_2)\mathrm{d}x_1 \wedge \mathrm{d}x_2$ and show that the R_j can be chosen so that $\varphi = (I_1^* - I_{-1}^*)\omega$ cannot be written as $\mathrm{d}\theta$ for some $\theta \in \Lambda_P$.
- To see this restrict φ on $\mathbb{R}^2$ and use Stokes' theorem on appropriate squares of the plane. The new idea here is that we can use the strict exponential distance distortion (see §8.2.2) of $\mathbb{R}^2$ in G.

In this context we can make the following remark, which, however, is marginal because it is only used in §G.4.2: when G is NB then both left and right translations act trivially on $H(\Lambda_P^*)$ because every closed current in Λ_P^* is polynomially homologous to a closed current of compact support. (Prove that this does it.)

14.3.3 About 0-distorted discrete subgroups

We have already pointed out (see Lemma 4.9) that a discrete central subgroup $\Gamma \subset G$ of the connected Lie group G is finitely generated. As a consequence it admits its intrinsic word distance. When that distance is equivalent (i.e. up to multiplicative constants) to the induced distance as a subset of G (assigned with its left-invariant Lie group distance) we shall say that Γ is 0-distorted (see §§2.14.1, 5.7.1). In view of the reduction that we shall prove in Appendix G (see (14.22) below), the following simple construction will be relevant.

Let G be some connected Lie group and consider $\pi\colon \widetilde{G} \to G$, the simply connected cover so that $\Gamma = \ker \pi \subset \widetilde{G}$ is a discrete central subgroup. Also let $\widetilde{G} = Q \leftthreetimes S$ be the Levi decomposition, where Q is the radical. As a first step, let us factorise π as follows:

$$\begin{aligned} &\widetilde{G} \to \overline{G} = \overline{Q} \leftthreetimes S \underset{\overline{\pi}}{\to} G; \\ &\overline{Q} = Q/Q \cap \Gamma. \end{aligned} \tag{14.20}$$

It is then clear that we have

$$\operatorname{Ker} \overline{\pi} = \Gamma/Q \cap \Gamma = \overline{\Gamma}, \quad \overline{Q} \cap \overline{\Gamma} = \{0\}, \tag{14.21}$$

and if we denote by $\overline{\theta}\colon \overline{G} \to S$ the canonical projection, then $\operatorname{Ker} \overline{\theta} \cap \overline{\Gamma} \subset \overline{Q} \cap \overline{\Gamma} = \{0\}$, and therefore $\overline{\theta}$ is one-to-one when restricted to $\overline{\Gamma}$.

On the other hand, since $\overline{\theta}(\overline{\Gamma}) \subset S$ is central it has to be discrete by the semisimplicity of S, and therefore, as we saw in §5.7.4, this image is 0-distorted in S. Therefore the original group $\overline{\Gamma}$ is also 0-distorted in $\overline{G}$.

The bottom line is this: For an arbitrary connected Lie group G we can find $\overline{G}$, some connected Lie group, and $\overline{\Gamma} \subset \overline{G}$, some 0-distorted central subgroup, such that $\overline{G}/\overline{\Gamma} \simeq G$ and $\overline{G} \simeq \overline{Q} \leftthreetimes S$ splits into a semidirect product, where $\overline{Q}$ is connected soluble and S is semisimple and simply connected.

The reason why this construction is useful in the 'reduction game' that we are playing in this chapter (see §14.1) is that it allows us to give a slightly different approach to the one that we described in the previous subsection.

Indeed, let $S = NAK$ be the Iwasawa decomposition (in the generalised sense that we explained in §4.6) of the semisimple simply connected cofactor S. This means that in the decomposition $\overline{G} = \overline{Q}NAK$ the soluble closed group $R = \overline{Q}NA$ is NC if and only if $\overline{G}$ is NB. Now, in Appendix G we shall show that quite generally we have that if $\pi\colon G_1 \to G_2 = G_1/\Gamma$ with $\Gamma \subset G_1$ 0-distorted, then

$$H(G_1)<+\infty \iff H(G_2)<+\infty, \tag{14.22}$$

with polynomial homology of course.

When we make the additional assumption that K is compact, the above, together with the fact that the group $\overline{G} \approx R \times K$ (quasi-isometry; see Example 11.9) finishes things. Explicitly, the information $H(G) < +\infty$ implies that $H(R) < +\infty$ (here the remark in §14.1.3 is used) and, since R is soluble from the reduction §14.3.1, it follows that R is NB.

Things are a little more tricky in the general case when K is not assumed to be compact. In that case we proceed as follows. We saw in Lemma 4.9 that there exists $\Gamma_S \subset S$, a closed discrete 0-distorted subgroup that is central in $\overline{G}$, and of finite index in $Z(S)$, the centre of S. We have therefore

$$\overline{G} \to \overline{Q} \leftthreetimes \overline{S} = \overline{\overline{G}} = R \cdot \overline{K}, \tag{14.23}$$

where $S/\Gamma_S = \overline{S} = NA\overline{K}$ is the generalised Iwasawa decomposition with compact $\overline{K}$. Now by using (14.22) twice we have

$$H(G)<+\infty \iff H(\overline{G})<+\infty \iff H(\overline{\overline{G}})<+\infty. \tag{14.24}$$

On $\overline{\overline{G}}$ we can argue as before because $\overline{\overline{G}} \simeq R \times \overline{K}$, a quasi-isometry. The required reduction is now done in two steps because using (14.24) we can pass from G to $\overline{G}$ and then from $\overline{G}$ to $\overline{\overline{G}}$.

Note that the use of the finite subgroup F of §11.1.5 and the exercise in §14.3.2 can be avoided with this approach (but of course this exercise is a special case of (14.22)).

About this alternative approach What are the advantages of using the covering mapping $G \to G/\Gamma$ and (14.22) in our reductions? The answer is not many. And had it been just for this, Appendix G would not have been worth writing. So, despite the fact that it would be a good thing to avoid the use of presheaf and Čech cohomology, this unfortunately cannot be done just with (14.22). It turns out, however, that it can be done by adopting a radically different approach, though at a very high price; we shall say a few words about it in §G.6.

14.4 Algebraic Tools

Ideally it is desirable in this chapter to have some familiarity with the use of homological techniques. Many readers may not have this familiarity, so the aim of this section is to provide such readers with a condensed and informal account of these algebraic tools. For more details see standard books such as Cartan and Eilenberg (1956), McCleary (2001), Godement (1958), Hilton and Wylie (1960), Cartan (1948), Greub et al. (1976), Bott and Tu (1982). This is a long list of references and they are the sources that the author has used in his effort to learn something about spectral sequences and the two geometric applications that we shall need in §§14.5–14.6. The reader will make their own choice in this list.

Warning To understand how spectral sequences work is quite tricky but it is possible to 'cheat your way' through the subject with little cost. This is exactly what we intend to do in this section.

14.4.1 A double complex

Everybody should be familiar with the notion of a differential graded module. Just as in Chapter 12, we shall simply refer to such a creature as a complex: $A = \sum_{n\geqslant 0} A_n$ is the grading and for $p < 0$, all the A_p are zero. Throughout, all modules here are just real vector spaces. The differential d is then just a linear mapping such that $\mathrm{d}^2 = 0$ and $\mathrm{d}\colon A_n \to A_{n+1}$ for $n \in \mathbb{Z}$. We shall give the outline that follows for complexes of cohomological type, that is, the differential increases the index, but of course nothing changes when the arrows go the other way round.

A double complex will be a similar creature with a double grading, $A = \sum_{n,m\geqslant 0} A_{n,m}$; as before we assume $A_{p,q} = 0$ if either p or q are negative (see Cartan and Eilenberg, 1956, §IV.4), and for simplicity we consider only 'first quadrant' complexes. Here two differentials $\mathrm{d}', \mathrm{d}''$ are defined: $\mathrm{d}'\colon A_{n,m} \to A_{n+1,m}$ and $\mathrm{d}''\colon A_{n,m} \to A_{n,m+1}$ with $(\mathrm{d}')^2 = (\mathrm{d}'')^2 = 0$; this qualifies them to be differentials. Furthermore, we impose the condition $\mathrm{d}' \circ \mathrm{d}'' + \mathrm{d}'' \circ \mathrm{d}' = 0$. The reason for this anticommutation relation is the construction of the associated simple grading that turns A into an ordinary complex: $A = \sum_{p\geqslant 0} A_p$ with $A_p = \sum_{m+n=p} A_{m,n}$, where the differential is $\mathrm{d} = \mathrm{d}' + \mathrm{d}''$ (verify that $\mathrm{d}^2 = 0$). This sometimes is called the total complex of A.

14.4.2 The two spectral sequences associated with a double complex

What we shall describe in this subsection is the beginning of the construction of a special spectral sequence that is attached to the double complex of the previous subsection. The merit of this special case is that it is simple to describe.

If we had wanted to give a general definition, however, we would have to have taken a different point of view and started from a 'filtered differential module'; this we do not intend to do. For that we shall refer to the extensive literature (as suggested at the start of this section).

Be this as it may, this spectral sequence $E = (E_r)$ is a sequence, $r = 0,1,2,\ldots$, of double complexes with $E_0 = (A_{p,q})$ as in the previous subsection. To construct $E_1 = (E_1^{p,q})$, we simply take the homology with respect to the first differential d': $E_1 = H(E_0, \mathrm{d}')$. Explicitly, this says

$$E_1^{p,q} = \mathrm{Ker}\left(A_{p,q} \xrightarrow{\mathrm{d}'} A_{p+1,q}\right) / \mathrm{Im}\left(A_{p-1,q} \xrightarrow{\mathrm{d}'} A_{p,q}\right). \tag{14.25}$$

Then the second differential d'' induces a differential on (E_1) which for reasons that will become clear will be denoted d_1. The homology of this gives $E_2 = (E_2^{p,q}) = H(E_1; \mathrm{d}_1)$. Explicitly,

$$E_2^{p,q} = \mathrm{Ker}\left(E_1^{p,q} \xrightarrow{\mathrm{d}_1} E_1^{p,q+1}\right) / \mathrm{Im}\left(E_1^{p,q-1} \xrightarrow{\mathrm{d}_1} E_1^{p,q}\right). \tag{14.26}$$

From here we continue with the construction of a sequence of double complexes $E_r = (E_r^{p,q})$, $r = 0,1,2,\ldots$, that we shall call the spectral sequence. In the literature this is called the 'second' spectral sequence attached to the double complex A.

For the construction of the 'first' spectral sequence, we interchange the roles of the two indices and do exactly the same things but start with d'' first and use d' on the corresponding E_1. (To make the distinction, Cartan and Eilenberg, 1956, §XV.6 uses the notation (I_r) and (II_r) for these two sequences.)

We shall not explain how the construction of E_r, $r = 3,\ldots$ is done. The reader can find that construction in the references that we indicated. Instead, what we shall give in the next few lines are some of the principal properties of this spectral sequence. It will turn out that by using just these properties and without necessarily knowing how the construction is done, we shall achieve our goal. This goal will be results on a number of specific double complexes that arise naturally on the geometry of a manifold. This is an unorthodox, but convenient and economical, way of proceeding! On the other hand, it should be stressed that if we really wanted to define E_r, $r = 3,\ldots$, we would be better off starting with the case of a general filtered differential module rather than with the special case of the above double complex to which one assigns its natural filtrations (see §14.4.4 below). In fact, if one is interested in the general definitions of a spectral sequence, as we have already said, what we are doing here amounts to a muddle of what the definition should be. The readers who know the subject will no doubt also say that the way we went about things is misleading. But let it be!

14.4.3 Definition and properties of spectral sequences

A spectral sequence is a sequence $E_r = (E_r^{p,q})$ of double complexes for $r = 1, 2, \ldots$ (it is convenient here to start from $r = 1$ and not $r = 0$). A differential operator is also defined by

$$\mathrm{d}_r \colon E_r \to E_r, \quad \mathrm{d}_r \colon E_r^{p,q} \to E_r^{p+r,q-r+1}; \quad \mathrm{d}^2 = 0; \tag{14.27}$$

drawing arrows in the first quadrant to indicate these mappings we see that $r = 0$ is special: see Figure 14.1.

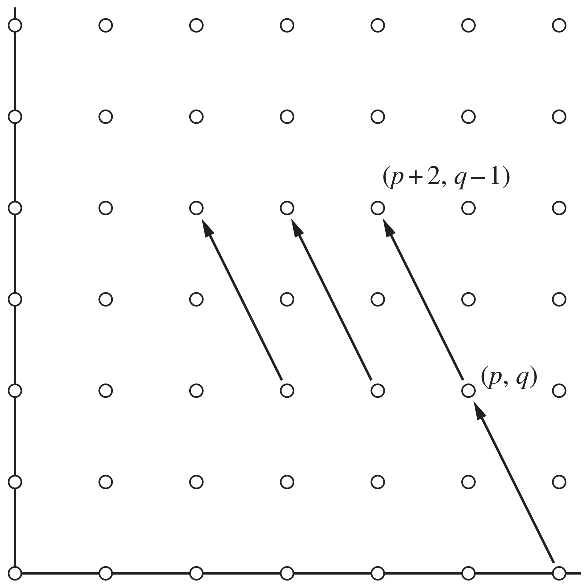


Figure 14.1 The arrows indicate d_2. Some authors (e.g. McCleary, 2001, figure on p. 29) use the transposed notation and interchange the lines and columns. See also Cartan and Eilenberg (1956, XV6, p. 331) for a formal transposition that shows that there is 'divine providence' that protects you if you get mixed up in this point.

The basic prerogative of the spectral sequence is that if we compute the homology of (E_r, d_r) we pick up E_{r+1}, that is, $E_{r+1} = H(E_r; \mathrm{d}_r)$,

$$E_{r+1}^{p,q} = \mathrm{Ker}\left(E_r^{p,q} \underset{\mathrm{d}_r}{\longrightarrow} E_r^{p+r,q-r+1}\right) / \,\mathrm{Im}\left(E_r^{p-r,q+r-1} \underset{\mathrm{d}_r}{\longrightarrow} E_r^{p,q}\right). \tag{14.28}$$

With this terminology, what we have asserted in the previous subsection is that the double complex A gives rise to a spectral sequence and we have defined the first two terms E_1, and E_2 and a mapping d_1 (see (14.25) and (14.26)). The d_2 and E_3 etc. were not constructed in §14.4.2. These have additional properties that we shall explain in the next subsection.

We shall consider only first-quadrant spectral sequences here; that is, we shall assume that $E_r^{p,q} = 0$ if either $p, q < 0$. Also, for simplicity, and because this is the only case that we shall need in the setting of differential forms, we shall assume that $E_1^{p,q} = 0$ for $q \geqslant n$ for some finite integer (which will be the dimension of the manifold in question). Things could have been the other way round and we would have then imposed instead that $E_1^{p,q} = 0$ for $p \geqslant n$ for some finite n. The point is that in either of the two cases it is clear that $\mathrm{d}_r = 0$ if $r \geqslant r_0$, say for $r_0 = n + 10$, because we then spill out of the (p,q)-band in the first quadrant on which $E_r^{p,q}$ has a chance of not vanishing. In this respect observe that it follows from (14.28) that for any p, q, r, the term $E_{r+1}^{p,q} = 0$ automatically if $E_r^{p,q} = 0$. What we have seen is that this happens when the spectral sequence lives in a band as above without necessarily requiring that the original double complex $A_{n,m}$ lives in the first quadrant.

In general, if it happens that $\mathrm{d}_{r_0} = \mathrm{d}_{r_0+1} = \cdots = 0$ we say that the spectral sequence *degenerates* at r_0. We then set $E_{r_0} = E_{r_0+1} = \cdots = E_\infty = (E_\infty^{p,q})$ and say that the spectral sequence $E_r \;\Rrightarrow\; E_\infty$ converges to E_∞.

Remark (This remark will not be needed here but it illustrates well the 'games' one plays with general spectral sequences.) The same argument shows that a first-quadrant spectral sequence always converges in the sense that for fixed $p, q \geqslant 0$, if r is large enough (say $r > p + q + 100$), then $d_r = 0$ (restricted on $E_r^{p,q}$). So we have again the convergence $E_r^{p,q} = E_{r+1}^{p,q} = \cdots = E_\infty^{p,q}$. This convergence is not 'uniform' as before, but for the applications this usually makes no difference.

14.4.4 The limit of the spectral sequence of a double complex

We return to the double complex A and explain an additional property of the spectral sequence of §14.4.2.

For this we consider the total complex $A = \sum_{p \geqslant 0} A_p$ with total differential $\mathrm{d} = \mathrm{d}' + \mathrm{d}''$, and on this we impose the 'filtration'; that is, a sequence of subcomplexes $A = F_0 \supset F_1 \supset \cdots$ defined by $F_p = \sum_{r \geqslant p} \sum_{q \geqslant 0} A_{r,q}$. This is compatible with the total differential d on A in the sense that $\mathrm{d}F_j \subset F_j$. From this it follows that $H = H(A; \mathrm{d})$, the homology of the total complex, is filtered by the induced filtration $H = H(F_0) \supset H(F_1) \supset \cdots$, where $H(F_j)$ is the canonical image of $H(F_j; \mathrm{d})$, the homology of the module F_j, in H. To this filtered module we associate the corresponding graded module GH (here it is a vector space over the reals) which is the direct sum of the modules $H(F_r)/H(F_{r+1})$.

The additional essential property of the first spectral sequence is that this graded homology is isomorphic to the total limit complex. Explicitly,

$$GH^n = \sum_{p+q=n} E_\infty^{p,q}, \tag{14.29}$$

for the homology of degree n. Equation (14.29) is important for applications below but we do not need to know how the isomorphism is done. In fact, here we are only dealing with vector spaces and the only thing that counts is that the dimension (possibly infinite) of $H(A)$ is the sum of the dimensions of the spaces $E_\infty^{p,q}$. Analogous facts hold for the second spectral sequence where in the filtration the rows and columns are interchanged.

Remarks (i) Let us pause a minute to demystify what we have done (or rather not done but claim that it can be done). In the applications (e.g. fibre bundles) it is $H = H(A;\mathrm{d})$ that we want to work out. But that homology is usually difficult to compute. The $E_r^{p,q}$, and especially E_2, on the other hand are easier to compute and since $E_r \rightrightarrows E_\infty$, we can approximate H by things that we know something about.

(ii) In some sense all the above is just linear algebra, and barely more when one considers spectral sequences of Abelian groups or more general modules. However, what makes the subject tricky and not friendly to 'working mathematicians' is the presence of triple indices and all the arrows that indicate the mappings that 'fly all over the place'. These tend to put people off. But once one is convinced that the subject is useful one is usually capable of digesting things.

14.4.5 Spectral sequences that degenerate on the second step

There are important examples of these. This, for instance, obviously happens if $E_2^{p,q} = 0$, with $p = 0, 1, \ldots$ and $q = 1, 2, \ldots$, that is, when the only non-zero terms of E_2 are to be found on the first column $q = 0$. To see this, it suffices to look at bi-degrees in (14.28). We could also switch the indices and demand that these, possibly non-zero, terms lie in the first line: see the six figures of Bott and Tu (1982, pp. 136–138), or you can draw your own figure! What we are saying is that then

$$E_2^{p,0} = E_\infty^{p,0} \simeq H^p(A;d), \tag{14.30}$$

where the $\simeq$ is a special case of (14.29). This observation is obvious and it will be used below.

Exercise 14.7 Marginal and inessential use of this exercise is made in §14.6. To do it one needs to know what the definitions of $\mathrm{d}_2, \mathrm{d}_3, \ldots$ (which we have not given in §14.4.2) actually are, so verifying the facts below will require a lot of dedication and work on the part of the reader.

Let $B = (B_n, \mathrm{d}_B)$ and $C = (C_n, \mathrm{d}_C)$ be two (simple) complexes with $B_n = C_n = 0$ for $n < 0$. We can then define $A = (B \otimes C)$, the double complex with $A_{n,m} = B_n \otimes C_m$ and $\mathrm{d}' = d_B \otimes (\mathrm{Id} C)$, $\mathrm{d}'' = \pm(\mathrm{Id} B) \otimes \mathrm{d}_C$ where the $\pm$ is there to guarantee that $\mathrm{d}'\mathrm{d}'' + \mathrm{d}''\mathrm{d}' = 0$; see, for example, Bott and Tu (1982, p. 91). What now happens is that the spectral sequence (the first or the second) that we can construct from this double complex degenerates at $r = 2$ and E_2 gives the homology of the total complex of A; this is the one that is sometimes referred to as the tensor product of B and C. The above says that $E_2 = \cdots = E_\infty = H(B \otimes C) = H(B) \otimes H(C)$ and more explicitly, $E_2^{p,q} = H^p(B) \otimes H^q(C)$. More on this can be found in Hilton and Wylie (1960, §§10.1.9, 10.3.12) or Bott and Tu (1982, Exercise 14.23).

Exercise 14.8 This exercise is easy and a special case of well-known observations in the subject. We shall make essential use of it in §14.6. Verify the following facts about the finiteness of the dimensions (more on these can be found in McCleary, 2001, §5.2 and Cartan, 1948, Exposé 10, J.-P. Serre). Let $A = (A_{n,m})$ be a double complex and assume that we have $\dim E_2 < +\infty$. By the subquotient property (14.28), it follows that the same holds for the dimensions of all the $E_3, \ldots, E_\infty$. We shall need the following partial converse of this.

We shall make the assumption that $E_2^{p,q} = B^p \otimes C^q$ for the two specific (simple) complexes B and C, as in the previous exercise. We shall assume that $\dim E_\infty < +\infty$ and that $0 \neq \dim B < +\infty$. The apparently artificial situation crops up naturally in geometric applications; see Bott and Tu (1982, §14.18) and (14.41) below. The conclusion is that $\dim C < +\infty$. (*Hint*. Consider the first p for which $B^p \neq 0$ and the first q, if such a q exists, for which $\dim C^q = \infty$. We have then $\dim E_2^{p,q} = \infty$ and also by (14.28) and the choice of p, q we see that the same thing has to hold for all the $E_r^{p,q}$ ($r \geqslant 2$). A contradiction can thus be obtained from this. Therefore no such q exists.)

14.5 The Čech–de Rham Complex

To a large extent, the way we have presented this section comes straight out of Bott and Tu (1982, §8). The aim is to give the proof of the fundamental fact that for any manifold M, and any good cover $\mathfrak{U}$ of M, the de Rham and the Čech cohomologies are identical: $H^*_{\mathrm{DR}}(M) \simeq H^*(\mathfrak{U})$; see §14.1.6. In Bott and Tu (1982) it is the classical case that is treated of course. Here we shall reproduce some of these arguments in a condensed but essentially self-contained manner, and then adapt them to make them work for polynomial cohomology in the Riemannian case of §14.2.5. The reader is urged to fall back on the book of Bott and Tu, or any other text on algebraic topology for additional reading.

14.5.1 The double complex

Let M be some C^∞ manifold. We shall fix some open, not necessarily good, cover $\mathfrak{U} = (U_\alpha)$. All the notation and in particular $U_{\underline{\alpha}} = U_{\alpha_0 \cdots \alpha_p}$ will be as in §14.2. We shall consider the double complex $K = (K^{p,q})$ with

$$K^{p,q} = C^p(\mathfrak{U}, \Omega^q) = \prod_{\alpha_0 < \cdots < \alpha_p} \Omega^q(U_{\alpha_0 \cdots \alpha_p}), \tag{14.31}$$

that is, Čech p-chains as in Example 14.3, with values on the presheaf of q-differential forms on the open sets. In this notation, the product $\prod$ is over all increasing $(p+1)$-multi-indices and, for any open set, $\Omega^q(U)$ indicates the differential forms of degree q on U. If you are unable to study the relevant sections of Bott and Tu (1982), then Figures 14.2, 14.3 (which are essentially reproduced from that work, pp. 95,97) will be helpful. Note that to conform with the notation of this book, the indices p, q for rows and columns, have been interchanged from what we had in Figure 14.1, which was closer to the notation of Cartan and Eilenberg (1956, §IV.4, p. 61).

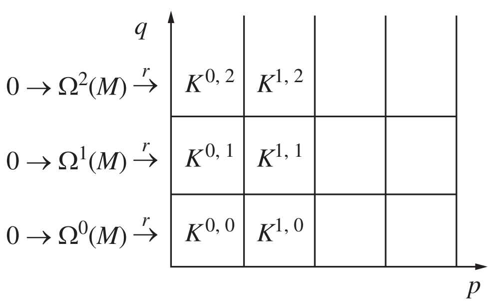


Figure 14.2 From Bott and Tu (1982). Reproduced with permission of Springer Nature.

Now on this double complex, the two differentials of §14.4.1 can be defined as follows. For the first, which in the notation there was actually denoted d′, we take the δ-Čech differential (14.6) and this acts on the rows; for example,

$$\to 0 \to \Omega^m(M) \underset{r}{\to} K^{0,m} \underset{\delta}{\to} K^{1,m} \underset{\delta}{\to} \cdots \tag{14.32}$$

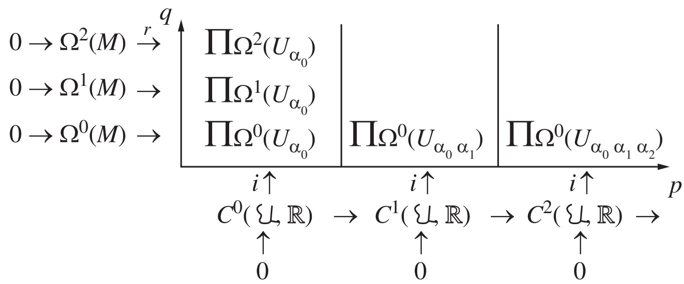


Figure 14.3 From Bott and Tu (1982). Reproduced with permission of Springer Nature.

for the *m*th row. In (14.32) the first mapping

$$\Omega^m(M) \underset{r}{\to} \prod_{\alpha} \Omega^m(U_\alpha)$$

is the restriction mapping coming from the inclusions $U_\alpha \subset M$. We have seen in §14.2.2.1 that the complex (14.32) is acyclic. In fact, here the homology of (14.32) vanishes identically.

The other differential, which is denoted d″ in §14.4, will be taken to be

$$(-1)^p \mathrm{d}\colon K^{p,q} \to K^{p,q+1},$$

for the exterior differential d on the differential forms $\Omega(U_{\alpha_0 \cdots \alpha_p})$. As already explained in §14.4.1, the $(-1)^p$ is needed for the construction of the total differential, which, to avoid confusion, we now write as

$$\mathrm{D} = \mathrm{D}' + \mathrm{D}'' = \delta + (-1)^p \mathrm{d}; \quad \mathrm{D}^2 = 0;$$

this is the notation in Bott and Tu (1982, §8), and in that book, this double complex is called the Čech–de Rham complex and is denoted by $C^*(\mathfrak{U};\Omega^*)$.

With the notation of §14.4 the acyclicity of the complex (14.32) says that $H(K;\mathrm{D}') = (E_1^{p,q})$ with $E_1^{0,q} = \Omega^q(M)$ and $E_1^{p,q} = 0$ for $p \neq 0$. We can therefore apply (14.30) and conclude that

$$H^q_{\mathrm{DR}}(M) = E_2^{0,q} = E_\infty^{0,q} \simeq \begin{pmatrix} \text{the } H^q \text{ homology} \\ \text{of the total complex} \end{pmatrix}, \tag{14.33}$$

which is a way of writing (abusing notation somewhat) (14.29) in the case at hand. We shall ignore the grading and, following Bott and Tu (1982), we shall denote the homology of the total complex by $H_{\mathrm{D}}\big(C^*(\mathfrak{U};\Omega^*)\big)$. The above can therefore be expressed by saying that

$$H^q_{\mathrm{DR}}(M) \simeq H^q_{\mathrm{D}}\big(C^*(\mathfrak{U};\Omega^*)\big), \tag{14.34}$$

where this isomorphism is the one induced by the restriction mapping r of (14.32). We have already urged the reader to study the direct proof of (14.34) – for instance in Bott and Tu (1982), Proposition 8.8 or p. 135 where it comes under the picturesque name of the tic-tac-toe lemma – which is much more instructive than the above and which is done without the use of spectral sequences.

14.5.2 The first spectral sequence

Now, as explained in §14.4, the same thing can be done the other way round on Čech–de Rham double complexes; that is, we can start from the differential $\mathrm{D}'' = (-1)^p \mathrm{d}$. This in some sense is simpler because here we are talking about the ordinary de Rham cohomology on each open set $U = U_{\alpha_0 \cdots \alpha_p}$ in (14.31). There is no presheaf involved here and for every fixed U we have the complex

$$0 \to \mathbb{R} \underset{\varepsilon}{\to} \Omega^0(U) \underset{\pm \mathrm{d}}{\longrightarrow} \Omega^1(U) \underset{\pm \mathrm{d}}{\longrightarrow} \cdots ; \tag{14.35}$$

in the jargon of the area, ε is called the augmentation, and sends 1 to the constant function on U. Here we shall impose the additional condition that $\mathfrak{U}$ is a good cover, and then $U \simeq \mathbb{R}^n$ a diffeomorphism. It follows from de Rham (1960, §19), or Bott and Tu (1982, §4), that the complex(14.35) is acyclic and has vanishing homology.

If we apply this to every $U = U_{\underline{\alpha}}$ we conclude for the double complex (14.31) that $E_1^{p,0} = C^p(\mathfrak{U})$ is the Čech complex defined in §14.2.1, and $E_1^{p,q} = 0$ for $q \neq 0$. This spectral sequence is the one that in §14.4.2 we called the first spectral sequence of the double complex (14.31).

The conclusion from §14.4 is therefore the analogue of (14.34), that is, that the Čech cohomology of M with respect to the cover $\mathfrak{U}$ is

$$H^p(\mathfrak{U}) = E_2^{p,0} = E_\infty^{p,0} = H_{\mathrm{D}}^p\big(C^*(\mathfrak{U};\Omega^*)\big). \tag{14.36}$$

Combining (14.34) and (14.36) we finally obtain the required result:

$$H^*_{\mathrm{DR}}(M) \simeq H^*(\mathfrak{U};M). \tag{14.37}$$

As already pointed out, the exact shape of this isomorphism is irrelevant. In fact, the only thing that counts here is that the dimensions of the two sides of (14.37), possibly infinite, are the same. The proof of (14.37) that we gave can be found in Bott and Tu (1982, §§12.1, 14.16). Note also that the same notation $\simeq$ will be used down the road a number of times and it will always stem from (14.29) and will always be interpreted as saying that the dimensions on the two sides are the same.

14.5.3 The polynomial Čech–de Rham complex

All the notation and in particular that of the Čech–de Rham complex $C^*(\mathfrak{U};\Omega^*)$ will be preserved and the following additional features will be introduced. First of all, a Riemannian structure will be assigned on the manifold M (and some base point $O \in M$ will be fixed) and the condition $\sup_\alpha \operatorname{diam}(U_\alpha) < +\infty$ and the local finiteness of Exercise 14.6 and the uniform boundedness of the gradient for the corresponding partition of unity, as in §14.2.4.3, will be imposed on the cover. Additional conditions will be imposed on this cover $\mathfrak{U}$ later on, but already with these conditions and in the spirit of what we have already done, it is possible to define a subcomplex of the Čech–de Rham double complex. This will be denoted by $C_P^*(\mathfrak{U};\Omega^*) = (K_P^{p,q}) = K_P$, and it will be the subcomplex (i.e. $K_P^{p,q} \subset K^{p,q}$) of the Čech–de Rham complex where $X = (\omega_{\underline{\alpha}}) \in \prod_{\underline{\alpha}} \Omega^q(U_{\alpha_0 \cdots \alpha_p})$ belongs to $K_P^{p,q}$ if there exist constants C that depend on X for which

$$\|\omega_{\underline{\alpha}}\| \leqslant C\big(1+\operatorname{dist}(U_{\underline{\alpha}}, O)\big)^C \quad \text{for all } \underline{\alpha}, \tag{14.38}$$

where $\|\omega_{\underline{\alpha}}\| = \sup\big(|\omega_{\underline{\alpha}}(x)| + |\mathrm{d}\omega_{\underline{\alpha}}(x)|\big)$ for the $|\omega|$ defined in §14.1 and the sup is taken as $x \in U_{\underline{\alpha}}$. This will be called the subcomplex of polynomial growth, denoted $C_P^*(\mathfrak{U};\Omega^*)$. The point of this definition is of course that the two differentials D', D'' induce differentials on K_P. Notice that the indices p, q only assume a finite number of values; as a consequence, the dependence of C on these indices in (14.38) is irrelevant.

It is important to understand that this is a 'global' condition of X and not one on each of the coordinates separately; that is, to decide whether this X belongs to $K_P^{p,q}$ we have to look at all of its coordinates of the Čech chain at the same time. We insist on this in order to stress the point that for some $X \in K_P$ it could be that $X = \mathrm{D}Y$ (or $\mathrm{D}'Y$ or $\mathrm{D}''Y$) for some $Y \in K$ and yet X is not a boundary in K_P (i.e. this Y cannot be chosen to belong to K_P). To deal with this aspect of our definition, we shall use the following.

The pivotal new observation This is the fact that if the good cover $\mathfrak{U}$ that we use satisfies the above two properties, and if we have a uniform bound on the gradients of the partition of unity as in §14.2.4.3, then the acyclicity of the polynomial versions of (14.32) and (14.35) is guaranteed. By this we mean the acyclicity of the complex (14.32), where we rewrite (14.32) with Ω_P^m and $K_P^{j,m}$. To see this we use the homotopy operator of §14.2.2 and the *gradient estimates* of the partition of unity in §14.2.4.3. Observe that $\mathfrak{U}$ does not have to be a good cover for this.

Similarly, for (14.35) the acyclicity is uniform with respect to U in the sense that when $p \geqslant 1$ for $\omega \in \Omega^p(U)$ with $\mathrm{d}\omega = 0$ we can solve $\mathrm{d}\theta = \omega$ with $\|\theta\|_U \leqslant C\|\omega\|_U$ for the L^∞ norms $\|\omega\|_U = \sup_{x\in U} |\omega(x)|$, and some constant C that is independent of $U = U_{\underline{\alpha}}$. For this the full thrust of the construction of §14.2.4.3 is needed; see Exercise 14.5.

From the above we see that when the cover has all the ‘strong properties’ of §14.2.4.3, the issue that we pointed out is no longer a problem. It follows that the proof of (14.37) that we gave generalises verbatim at a stroke to the polynomial case. The conclusion is summarised as follows.

With the same notation as before, let $\mathfrak{U}$ be a, not necessarily good, open cover for which $\sup_\alpha \operatorname{diam}(U_\alpha) < +\infty$, and for which the local finiteness of Exercise 14.6 and the boundedness of the gradients of the partition of unity of §14.2.4.3 holds. Then we have

$$H^*_{\mathrm{DR}}(M;\mathrm{Pol.}) \simeq H_{\mathrm{D}}\big(C^*_P(\mathfrak{U};\Omega^*)\big).$$

We shall now assume in addition that the cover is good and has all the above properties. To fix ideas we shall assume that the manifold is a Lie group, or at least a homogeneous space, and the covering is the one we explicitly constructed in §14.2.4.3. Then we can assert that

$$H^*_P(\mathfrak{U}) \simeq H_{\mathrm{D}}\big(C^*_P(\mathfrak{U},\Omega^*)\big) \simeq H^*_{\mathrm{DR}}(M;\mathrm{Pol.}).$$

14.6 Proof of Proposition 14.2

The classical de Rham cohomology As in the previous section, we shall start with the classical cohomology and a fibre bundle $F \hookrightarrow E \underset{\pi}{\to} B$ and no Riemannian structure or polynomial growth will be involved at first. In this section we shall follow closely the exposition in Bott and Tu (1982, §14.18). The modifications that have to be made to that proof for the cohomology of polynomial growth and for the proof of Proposition 14.2 from §14.1.4 will be given in the second part of this section. To avoid irrelevant complications we will assume throughout that the fibre F is *connected*.

Questions of notation The results that we shall prove, under appropriate conditions, hold for general bundles. Here, however, we shall concentrate on a connected Lie group G and the corresponding homogeneous space $G|K$, where K is a connected and compact subgroup. Depending on the point we wish to emphasise we shall use *interchangeably* the following notation on the bundle: E or G; F or K; B or $G|K$.

14.6.1 The de Rham cohomology

We start from an arbitrary fibre bundle $F \hookrightarrow E \underset{\pi}{\to} B$ and some good cover $\mathfrak{U} = (U_\alpha)$ on B. This cover will be assumed to give a chart for the bundle, as in §14.1.5. For this, it suffices of course that the cover is sufficiently fine. It is also clear that in the case of Lie groups that we consider in this section, the construction that we made in §14.2.4.3 can be made fine enough for this to hold. To this we associate $\pi^{-1}\mathfrak{U} = (\pi^{-1}U_\alpha)$ which is a cover (in general not a good cover) of E. With this cover we construct the Čech–de Rham double complex:

$$C^p(\pi^{-1}\mathfrak{U};\Omega^q) = \prod_{\alpha_0<\cdots<\alpha_p} \Omega^q(\pi^{-1}U_{\alpha_0,\ldots,\alpha_p}). \tag{14.39}$$

The differentials on this complex are denoted, as in §14.5, by $\mathrm{D} = \delta + (-1)^p\mathrm{d} = \mathrm{D}' + \mathrm{D}''$. Then, as in §14.5, we take the homology with respect to D'' and obtain the E_1 term of the corresponding spectral sequence

$$E_1^{p,q} = \prod_{\alpha_0<\cdots<\alpha_p} H^q(\pi^{-1}U_{\alpha_0,\ldots,\alpha_p}). \tag{14.40}$$

We saw in §14.2 that the coefficients under the $\prod$ signs belong to a locally constant presheaf defined on the cover $\mathfrak{U}$ and the presheaf is constant if the bundle is principal (see §14.2.3; or for a general bundle, provided that B is simply connected, see Bott and Tu, 1982, §§13, 14, p. 169). When this presheaf is constant, (14.40) can be written $E_1^{p,q} = C^*\big(\mathfrak{U};H^q(F)\big)$ and in what follows the cohomology of the fibre will be assumed to be finite (i.e. have finite Betti numbers) so that $H^q(F) = \mathbb{R}\times\cdots\times\mathbb{R}$, m copies of the reals, and $E_1^{p,q} = C^*(\mathfrak{U};\mathbb{R}^m)$. In an alternative notation, we can therefore write

$$E_2^{p,q} = H^p(\mathfrak{U};\mathbb{R}^m) = H^p(B)\otimes H^q(F). \tag{14.41}$$

On the other hand, from the results of §14.5 it follows that $E_\infty^{*,*}$ is some grading of $H^*_{\mathrm{DR}}(E)$ – observe that for this we do not need the cover $\pi^{-1}U_\alpha$ to be good; see §14.5.1. We then invoke Exercise 14.8, and the assertions of §14.1.5 follow where we are in the classical situation and we are talking about standard de Rham cohomology. More explicitly, we can recapitulate:

Bottom line We use presheaf cohomology and spectral sequence as follows:

The set-up $F \hookrightarrow E \to B$ is a fibre bundle (on the C^∞ category, i.e. manifolds and smooth mappings).

The hypothesis

(i) $H^*(F) < +\infty$, that is, the Betti numbers are finite. This holds for example if F is compact.

(ii) The homology presheaf of the bundle is constant. Since we have assumed that $\mathfrak{U}$ is a good cover this is guaranteed by either of the following two conditions:

  (a) B is simply connected;

  (b) $F \simeq$ some connected Lie group and the bundle is principal.

This is the case in particular when the bundle is given by a homogeneous space of the connected Lie group G and $G \to G|K$ for some compact connected subgroup K.

From the facts in §14.5.1 applied to the cover $\pi^{-1}\mathfrak{U}$ on E we deduce that $H^*_{\mathrm{DR}}(E) = E^{**}_\infty$. We now use (14.41) as explained above, and then Exercise 14.8 to obtain the required conclusion.

The conclusion In the above set-up, $H^*(E)$ is finite if and only if $H^*(B)$ is finite. This is a result that was stated in §14.1.5.

14.6.2 The polynomial cohomology

The strategy As before we shall use the Riemannian structure of the total space E of the bundle and for some good cover on B we define the subcomplex $C^*_P(\pi^{-1}\mathfrak{U};\Omega^*)$ of (14.39) of the chains of polynomial growth. Using this new double complex, the spectral sequence $E_1, E_2, \ldots, E_\infty$ can be defined as in §14.5, and the proof for this polynomial cohomology, and in particular for Proposition 14.2, follows exactly as for the classical de Rham cohomology verbatim.

Some care is, however, needed because we are dealing with very abstract creatures and one can very easily go wrong. To make sure that we have a real proof we shall therefore, in the rest of this subsection, give precise definitions and details. Notice, however, straight away that for this approach to go through, the compactness of the fibre $F = K$ is essential for otherwise the cover $\pi^{-1}\mathfrak{U}$ cannot be used in §14.5.3.

The presheaf of the fibre bundle revisited We are working in the C^∞ category (manifolds and smooth mappings) with $F \subsetneq E \xrightarrow[\pi]{} B$ some fibre bundle, and for simplicity we shall assume that F is a connected Lie group and the bundle is principal. In fact, it is only the case $G \to G|K$ of a homogeneous space (with G a connected Lie group and K a compact connected subgroup) that we shall use.

We shall consider all $V \subset B$ open subsets that are diffeomorphic to $\mathbb{R}^n$ and, in addition, V is contained in some chart of the bundle (see §14.1.5). For each such set we shall define $H^*(\pi^{-1}V) = \mathfrak{F}(V)$. In other words, the functor is only defined in this subcategory of $\mathrm{Open}(B)$. We saw in §14.2.3 that we can define an isomorphism

$$\gamma\colon H^*(\pi^{-1}V) \to H^*(F). \tag{14.42}$$

This is induced (for a principal bundle) by the injection $F \to \pi^{-1}V$, $F \ni \varphi \to g_x \cdot \varphi$, where $x \in V$, $g_x \in \pi^{-1}x$, and the multiplication is the group action (written as a right action) on the principal bundle. This definition of γ is independent of the choice of x and g_x. The details can be found in §14.2.3.

A trivialisation over V Again, V is open in B and $V \simeq \mathbb{R}^n$. Let us consider now the space $V \times F$, which, we note, admits the natural structure of a principal bundle; then just as before we can define

$$\beta : H^*(V \times F) \to H^*(F). \tag{14.43}$$

Also, if we fix some trivialisation $\tau_V : \pi^{-1}V \underset{\simeq}{\to} V \times F$, we pick up isomorphisms

$$\tau\colon H^*(V \times F) \to H^*(\pi^{-1}V). \tag{14.44}$$

These do not depend on the particular trivialisation because, by their construction, these mappings satisfy

$$\tau = \gamma^{-1} \circ \beta, \quad \gamma = \beta \circ \tau^{-1}. \tag{14.45}$$

14.6.2.1 The (quasi-)norms We shall place ourselves in a situation where natural quasi-norms will be defined on $H^*(F)$, $\mathfrak{F}(V) = H^*(\pi^{-1}V)$ and $H^*(V \times F)$. Indeed, F, $\pi^{-1}V$ and $V \times F$ are manifolds, and observe that once a Riemannian structure is assumed on the manifold M, we can assign the norm

$$\|\omega\| = \|\omega\|_\infty + \|\mathrm{d}\omega\|_\infty; \quad \omega \in \Omega^*(M), \tag{14.46}$$

for the uniform norms, as in §14.1. Then since $H^*(M) = \mathrm{Ker\,d}/\mathrm{Im\,d}$ is a subquotient of $\Omega^*(M)$, this space is also assigned canonically with a quasi-norm ('quasi' because $\mathrm{Im\,d}$ is not necessarily closed; we shall, however, see that in our case all these will be proper norms). This quasi-norm will be denoted $q_M(x)$, with $x \in H^*(M)$.

For the definition of the norm (14.46), the issue is to decide what Riemannian structures we shall use on F, $\pi^{-1}V$, $V \times F$. Here for simplicity we shall not examine the problem for general fibre bundles but restrict ourselves to $G|K$, the homogeneous spaces that we described in §§14.1–14.2.

The Riemannian metric on G is then some G-left-invariant and K-right-invariant metric as explained in §14.2.4.1, and this induces Riemannian metrics on K, $G|K$ and on each $V \subset G|K$, $\pi^{-1}V$ and $V \times K$ with the product Riemannian structure. We have therefore the required quasi-norms on $H^*(F)$, $\mathfrak{F}(V) = H^*(\pi^{-1}V)$, $H^*(V \times F)$. Without giving formal definitions one could say that with this norm on $\mathfrak{F}(V)$ we obtain a 'normed presheaf', that is, a functor with values in the category of quasi-normed vector spaces.

It is also clear that β is a quasi-norm isomorphism, that is,

$$q_{V\times F}(x) = q_F(\beta(x)); \quad x \in H^*(V \times F). \tag{14.47}$$

We shall give the details in the exercise below.

Exercise To see this use the canonical mappings $F \to V \times F \to F$ defined by the product structure. Then the induced vector space isomorphisms

$$H^*(F) \underset{i}{\to} H^*(V \times F) \underset{p}{\to} H^*(F)$$

are norm decreasing, with $p \circ i =$ Identity simply because the corresponding mappings on the spaces of differential forms $\Omega^*(\cdot)$ are norm decreasing. Notice, however, that the product Riemannian structure is used in an essential way here and therefore the way we defined the Riemannian structures matters both here and in what follows.

More details Whether q_F in (14.47) is a genuine norm or only a quasi-norm is for us inessential because in the proof below we can in fact get round that difficulty; see the remark at the end of the section. In the next few lines we shall, however, see that we actually have a norm. By the definition of the norm (14.46) (with $M = F$) it is clear of course that $\operatorname{Ker} d \subset \Omega^*$ is a closed subspace. But a priori, the same thing is not clear for $\operatorname{Im} d$. In our case, F is a compact Lie group (or even a torus, and in particular $H^*(F)$ is finite-dimensional; see Bott and Tu, 1982, §5.3.1).

To show that $\operatorname{Im} d$ is closed for the normed topology of the norm (14.46) we shall fall back on the previous notation that we used in Chapters 12 and 13 (i.e. de Rham's and Schwartz's notation; de Rham, 1960, §§9, 10) and set $\Omega^*(F) = \mathscr{D}$ so that the dual space, which is the space of currents on F, is denoted by $\mathscr{D}'$. The two spaces $(\mathscr{D}, \mathscr{D}')$ are in duality and the mapping that is dual to d of $\mathscr{D}$ is denoted by b on $\mathscr{D}'$ (see §12.3.6 and de Rham, 1960, §11).

Now let $\omega_j \in \mathscr{D}$, for $j = 1, \ldots$, be such that $d\omega_j \to \omega \in \mathscr{D}$ for the uniform topology. We need to show that there exists $\theta \in \mathscr{D}$ such that $\omega = d\theta$. But from the above duality it is clear that there exists $T \in \mathscr{D}'$ such that $\omega = \mathrm{b}T$ (here we identify $\mathscr{D} \subset \mathscr{D}'$ as a subspace (see §12.3.5 and de Rham, 1960, §8). Indeed, $\mathrm{d}\omega_j \in \mathrm{Im}\,\mathrm{b}$ and therefore also $\omega \in \mathrm{Im}\,\mathrm{b}$ because $\mathrm{Im}\,\mathrm{b} = (\ker \mathrm{d})^{\perp}$ (see (12.128): this is the non-trivial 'Poincaré duality) is closed for the weak topology $\sigma(\mathscr{D}', \mathscr{D})$ on $\mathscr{D}'$ (see Bourbaki, 1953).

The fact that T can be chosen to belong to $\mathscr{D}$ is not formal and it relies on the regularisation of currents that we used in §12.10. In the classical reference (de Rham, 1960, §15, especially p. 77), one sees that in the case of the Lie group F this regularisation is obtained simply by convolution.

What this regularisation says is that we can construct two operators R and A of $\mathscr{D}'$ (see §12.10) such that $T - RT = \mathrm{b}AT + A\mathrm{b}T$, for $T \in \mathscr{D}'$, and such that $R\mathscr{D}' \subset \mathscr{D}$ and $A\mathscr{D} \subset \mathscr{D}$. For the above ω we obtain therefore $\omega = \mathrm{b}(RT + A\omega) = \mathrm{d}\theta$ with $\theta \in \mathscr{D}$, as needed.

Note that all the above is a proof of the 'elementary' fact that $\mathrm{Im}\,\mathrm{d}$ is closed for, say, F a torus. The proof that we suggested is perhaps not the simplest, and certainly not the more elementary, since Poincaré duality is used. The reader is invited to give a more direct proof for themselves. *Hint*. Use the fact that the homology $H(F)$ is finite-dimensional and the observation on Banach spaces in the exercise of §12.12.1. To be able to use the Banach space Λ_P induced by the norm (14.46), the same regularisation of §12.10 has to be exploited. This will show that the dimension of the homology of Λ_P is no larger than that of the homology of Ω^*. Some work is needed here.

The choice of the trivialisation This subsection is pivotal. We consider $\mathfrak{U} = (U_\alpha)$, the good cover that was constructed in §14.2.4.3 on the base space $B = G|K$ of the bundle, and restrict further the category of open sets $V \subset B$ on which the functor $\mathfrak{F}(V) = H^*(\pi^{-1}V)$ is defined. For this subcategory we demand that $V \simeq \mathbb{R}^n$ (diffeomorphisms) and also $V \subset U_\alpha$ for some open set of the cover. Notice that by the special properties of the cover, for each fixed V only finitely many indices α can be used. This particular property in the construction of the cover, as defined in §14.2.4.3, is, however, not essential. What is essential is to show that we can *define* trivialisations

$$\tau_\alpha \colon \pi^{-1}U_\alpha \to U_\alpha \times K \tag{14.48}$$

that induce Riemannian quasi-isometries (for the Riemannian structures that we defined in §14.6.2.1) uniformly in α.

To see this, we shall fix some trivialisation $\pi^{-1}U \simeq U \times K$ for some small ball $U \subset G|K$ centred at $\dot{e} \in G|K$ (the class of the identity – note that this is the identity element of the group G/K where K is normal). Then since $U_\alpha = g_\alpha U$ for some $g_\alpha \in G$ (the G-actions on $G|K$) we have $\pi^{-1}U_\alpha = g_\alpha \pi^{-1}U$ (group multiplication). This induces the required trivialisations

$$\pi^{-1}U_\alpha \simeq \pi^{-1}U \simeq U \times K \simeq U_\alpha \times K, \tag{14.49}$$

where the final identification in (14.49) which is induced by $U \to g_\alpha U = U_\alpha$ is a Riemannian isometry on $G|K$. This construction depends of course on the choice of the representatives g_α.

Finally, the following additional choice has to be made before we complete our construction. For every V in our subcategory we shall choose one (out of the finitely many) U_α such that $V \subset U_\alpha$ and then the trivialisation (14.49) that we have chosen will induce a trivialisation

$$\tau_V : \pi^{-1}V \underset{\simeq}{\rightarrow} V \times K. \tag{14.50}$$

This trivialisation will now be fixed for each V. This τ_V is a Riemannian quasi-isometry uniformly in V (induced Riemannian structure on the left of (14.50) and the product Riemannian structure on the right).

Now we go back to (14.44) and define

$$\tau_V = H^*(V \times F) \to H^*(\pi^{-1}V). \tag{14.51}$$

This, we have seen, is uniquely determined, is independent of the previous choices, and the fact that (14.50) is a Riemannian quasi-isometry implies that (14.51) is a (quasi-)norm equivalence, that is,

$$q_{V\times F}(x) \approx q_{\pi^{-1}V}(\tau_V(x)); \quad x \in H^*(V \times F). \tag{14.52}$$

Now, from $\gamma = \beta \circ \tau_V^{-1}$ in (14.45) and from the fact that τ_V and β preserve, up to equivalence, the corresponding $q(\cdot)$ seminorms (see (14.47)), we conclude that

$$q_{\pi^{-1}V}(x) \approx q_K(\gamma(x)); \quad x \in H^*(\pi^{-1}V). \tag{14.53}$$

Furthermore, the equivalences in (14.52) and (14.53) are uniform in V.

A digression (Presheafs of quasi-normed vector spaces.) We shall leave it to the reader to give the formal definitions if they so wish but what we have done can be summarised as follows. We constructed a presheaf of quasi-normal vector spaces that is isomorphic (in the sense of §14.2.1.3, i.e. up to quasi-norm equivalence) to the trivial presheaf $H(F)$ assigned with the quasi-norm that we defined in this subsection.

Combining the previous observations Here we shall put things together and use freely the terminology of presheafs. For this, with $\mathfrak{U}$ a good cover of B as in §§14.2.4.3, 14.6.1 we shall go back to the double complex of (14.40):

$$X = \prod_{\alpha_0 < \cdots < \alpha_p} H^q(\pi^{-1} U_{\alpha_0,\dots,\alpha_p}) \tag{14.54}$$

and consider a subcomplex $X_P \subset X$ – the polynomial complex.

To define X_P we decree that $x = (x_{p,q}) \in X$ belongs to X_P if there exist constants $C > 0$ such that the coordinates $x_{p,q}(\alpha_0, \dots, \alpha_p) \in H^q(\pi^{-1} U_{\alpha_0,\dots,\alpha_p})$ satisfy

$$q_{\pi^{-1} U_{\underline{\alpha}}}\big(x_{p,q}\big(\alpha_0, \dots, \alpha_p\big)\big) \leqslant C\big(1 + \mathrm{dist}(\dot{e}, U_{\alpha_0,\dots,\alpha_p})\big)^C; \tag{14.55}$$

here we recall $U_{\alpha_0,\dots,\alpha_p} \subset B = G|K$ and $\mathrm{dist}(\cdot,\cdot)$ refers to the distance on the homogeneous space $G|K$, where $\dot{e}$ is the identity class.

Another way of writing the double complex of (14.54) is

$$X = \prod_{\alpha_0 < \cdots < \alpha_p} \mathscr{H}^q(U_{\alpha_0,\dots,\alpha_p}), \tag{14.56}$$

where $\mathscr{H}^q$ denotes the constant presheaf of §14.2.3 defined by the cover $\mathfrak{U}$ with values on the finite-dimensional vector space $H^q(F) = \mathbb{R} \times \cdots \times \mathbb{R}$ (cf. (14.42)).

Now taking the D' differential (see §§14.4, 14.5.1) we obtain

$$X_2 = H^p(\mathfrak{U}; H^q(F)) = H^p(\mathfrak{U}) \otimes H^q(F). \tag{14.57}$$

This is exactly as we did in §14.6.1. We shall now restrict the action of D' to the polynomial subcomplex X_P. Because of the quasi-norm equivalence (14.53), we obtain in this way the corresponding $(X_2)_P$ which is

$$(X_2)_P = H_P(\mathfrak{U}) \otimes H(F). \tag{14.58}$$

Dénouement We shall now go back to the Čech–de Rham complex (14.39) that is associated with the bundle

$$C^p(\pi^{-1}\mathfrak{U}; \Omega^q) = \prod_{\alpha_0 < \cdots < \alpha_p} \Omega^q(\pi^{-1} U_{\underline{\alpha}}), \tag{14.59}$$

and consider the associated polynomial subcomplex for which $\omega = (\omega_{p,q})$, an element of (14.59), has coordinates $\omega_{p,q}(\underline{\alpha}) \in \Omega^q(\pi^{-1} U_{\underline{\alpha}})$ that satisfy

$$\|\omega_{p,q}(\underline{\alpha})\| \leqslant C(1 + \mathrm{dist}(\dot{e}, U_{\underline{\alpha}}))^C; \quad \underline{\alpha} = (\alpha_0, \dots, \alpha_p), \tag{14.60}$$

for the norm (14.46) and where the constant C depends on ω (notice that $\omega_{p,q} \neq 0$ only for a finite number of p, q so, at least here, there is no point speculating about whether C could also depend on these indices). The compactness of F

and the fact that the distance of $U_{\underline{\alpha}}$ from $\dot{e}$ is approximately the same as the distance of $\pi^{-1}U_{\underline{\alpha}}$ from e, are here essential. This subcomplex of (14.59) will be denoted by $C_P^*(\pi^{-1}\mathfrak{U};\Omega^*)$.

As in §14.5 we shall apply the D$''$ differential on this polynomial Čech–de Rham subcomplex to get the first term of the spectral sequence $E_1 = \left(E_1^{p,q}\right)$ and this, after a moment's thought, is none other than the X_P that we defined in (14.55). In that 'moments reflexion' [*sic*], use the notation of (14.55) to represent each $x_{p,q}(\underline{\alpha})$ by a cycle $\omega_{p,q}(\underline{\alpha}) \in \Omega^q(\pi^{-1}U_{\underline{\alpha}})$ and in such a way that (14.60) holds. The other way round, the fact that $E_1 \subset X_P$ is even more obvious.

The second term of the spectral sequence, as explained in (14.58), is therefore

$$E_2^{p,q} = H_P^p(\mathfrak{U};B) \otimes H^q(F). \tag{14.61}$$

On the other hand, by the abstract theory of §14.4, we know that

$$E_\infty = H_{\mathrm{D}}\left(C_P^*(\pi^{-1}\mathfrak{U};\Omega^*)\right) \tag{14.62}$$

for some appropriate graduation of the right-hand side.

Now the complex $C_P^*(\pi^{-1}\mathfrak{U};\Omega^*)$ is exactly the polynomial complex on the space E for the cover $\pi^{-1}\mathfrak{U}$ that we considered in the previous section; see (14.38). We can therefore apply the results of §14.5.3 and deduce that the right-hand side of (14.62) is $H_{\mathrm{DR}}^*(E;\mathrm{Pol.})$. Ignoring the graduations we deduce that the two vector spaces are isomorphic:

$$E_\infty \simeq H_{\mathrm{DR}}^*(E;\mathrm{Pol.}). \tag{14.63}$$

To make (14.63) work we use the partition of unity subordinated to $\pi^{-1}\mathfrak{U}$ (cf. §14.2.2) that we pick up by pulling back the partition of unity of the original cover $\mathfrak{U}$ on B. The condition on the gradients of this partition of unity (see §14.2.4.3) is therefore satisfied simply because this holds for the good cover $\mathfrak{U}$ on B.

It remains now to combine (14.61), (14.62) and (14.63) with Exercise 14.8 and we conclude the required result:

$$\dim H_{\mathrm{DR}}^*(E;\mathrm{Pol.}) < +\infty \iff \dim H_P^*(\mathfrak{U};B) < +\infty. \tag{14.64}$$

Given that on B , $H_P^*(\mathfrak{U};B)$ and $H_{\mathrm{DR}}^*(B;\mathrm{Pol.})$ are the same by the previous subsection, it follows that Proposition 14.2 has been proved, and by §14.3.2, this completes the proof of (14.5) and we are done.

Remark Throughout the proofs we have used, at least implicitly, the fact that q_F is a norm. But even in situations where we have problems guaranteeing this (e.g. $\dim H_P^* < +\infty$ for F, but F is not compact) the proof can be adapted.

Indeed, we have in general $H_P^* = H' \oplus H''$, a direct decomposition and q_F restricts to a proper norm on H' and vanishes on H''. What is then obtained, instead of (14.61), is

$$E_2 = \big(H_P^*(B) \otimes H'\big) \oplus \big(H^*(B) \otimes H''\big). \tag{14.65}$$

Notice that both the polynomial and the standard cohomologies of B are involved here. (*Warning*. The author confesses that he has not written out the details of this approach, and it is in this light that (14.65) should be viewed.)

Exercise Now that Proposition 14.2 has been proved, the reader is invited to do the following 'retrospective' and figure out why the following very simple (but incorrect!) 'proof' of (14.16) does not work: repeat the argument of the exercise in §14.3.2 with the finite group F replaced by the central compact torus of (14.16). There is after all no problem in replacing $\frac{1}{|F|}\sum_{x\in F}$ by the normalised integral $\int_F$. What goes wrong is subtle (Greub et al., 1973, §6.3). It is possible to straighten things out. The procedure for doing this is deep and interesting (cf. §G.6 below).

Appendix: Künneth's formula

A particularly simple bundle is the one where $E = B \times F$ is a direct product. Spectral sequences can be used to prove that

$$H^*(E) = H^*(B) \otimes H^*(F)$$

and the analogous formula for the polynomial cohomology (cf. Bott and Tu, 1982, §14.19). This is what is called Künneth's formula. Here, for simplicity, we shall assume as before that F is compact.

One way to do things is to start from the simple observation that in this special case in the corresponding Čech–de Rham complex of (14.59) we can specify the following subcomplex:

$$\mathbb{A} = \mathbb{F} \otimes \mathbb{B} = \Omega^*(F) \otimes C^*(\mathfrak{U}; B) \subset C^*(\pi^{-1}\mathfrak{U}; \Omega^*) = \mathbb{E}. \tag{14.66}$$

As we observed in Exercise 14.7, the E_2 of $\mathbb{A}$ is just $H(\mathbb{F}) \otimes H(\mathbb{B})$ and furthermore this spectral sequence degenerates so that $E_2 = E_3 = \cdots = E_\infty$.

Now we have seen that the E_2 term of the spectral sequence of the Čech–de Rham complex $\mathbb{E}$ is exactly the same. By very simple, purely algebraic considerations, it follows that the same thing holds for the successive terms $E_2 = \cdots = E_\infty$ and these are identical for $\mathbb{A}$ and $\mathbb{E}$. This is the content of Cartan and Eilenberg (1956, Theorem 15.3.2) or McCleary (2001, §3.4). You do not have to know how the spectral sequences are constructed to use this reference here. What we did in §14.4 suffices provided that you adopt the correct point of view (which we did not!) so that the correspondence that assigns the spectral sequence $E(A)$ to A is functional. As we have already said, this viewpoint involves just a module A assigned with differentiation and a compatible filtration.

Therefore $H^*(\mathbb{E})$, which is isomorphic to E_∞, is $H(\mathbb{F}) \otimes H(\mathbb{B})$, as stated in the Künneth formula.

Exercise Assume that $H_P^*(F) < +\infty$ (but that F is not necessarily compact). The polynomial Čech–de Rham complex can again be defined. Here, for the coordinates $\omega_{p,q} \in \Omega^q(\pi^{-1}U_{\alpha_0,\ldots,\alpha_p})$ we demand that $|\omega_{p,q}(x)| + |\mathrm{d}\omega_{p,q}(x)| \leqslant C(1+|x|)^C$, with $x \in B \times F$ and $|x|$ denotes the distance from the base point. Adapt the arguments of this subsection to show that $H_P^*(E) = H_P^*(B) \otimes H_P^*(F)$.

This exercise, while not important in our considerations, is interesting for several reasons. In particular, the spaces $\Omega^q(\pi^{-1}U_{\underline{\alpha}})$ are not now normed spaces but have instead an inductive limit topology (see §12.13). Nonetheless, natural quasi-norms can be induced on the corresponding homologies because, by our hypothesis, these are finite-dimensional.

This is also the moment, if ever there was one, for the reader to check out the proof of Künneth's formula given by A. Grothendieck that uses the theory of nuclear (locally convex) topological vector spaces (see Schwartz, 1953). For more on this see Appendix G next. For the author, that proof was particularly inspiring but this may well reflect a subjective view.

An extension to more general fibre bundles

The reader is invited to extend the results of §14.6 to the setting of a general fibre bundle $F \hookrightarrow E \to B$ where the following hypothesis is made: Riemannian structures are assigned to the three connected manifolds F, E, B, and a trivialisation (cf. §14.1.5) exists such that $\varphi_\alpha\colon \pi^{-1}U_\alpha \to U_\alpha \times F$ are quasi-isometries uniformly in α. Notice also that this situation occurs naturally in our theory: $E = G$ a connected Lie group and $\Gamma \subset G$ is some discrete central subgroup. We then take F to be some closed connected Abelian subgroup such that F/Γ is compact (see Hochschild, 1965, Chapter XVI). An interesting spectral sequence as in §14.6 crops up from this. Furthermore, we feel that with a bit of luck this could settle the 'ultimate reduction' of §14.1.4. There is a problem to all of this however. (What is the problem? *Hint*. Look at (G.1) in Appendix G.). On the other hand, it is very likely that this formalism can be used to provide an alternative proof of (14.24) without the use of the Cartan–Leray spectral sequence of §G.2.3. The reader is invited to explore this set-up further.

Appendix G
Discrete Groups

The aim of this appendix is to introduce the reader to some interesting problems related to our theory as applied to discrete groups (cf. §1.9). To be able to do this in a few pages I had to recall some of the background material in a rather sketchy way (in particular, in §§G.1.2, G.3.1, G.5 below). Those readers that get interested in these problems will be able to find additional information on this necessary background in the literature and in the references that will be supplied in the text.

Throughout this appendix, Γ will denote some discrete, finitely generated group and $|\gamma| = d_\Gamma(e, \gamma)$, for $\gamma \in \Gamma$, will, as in §1.1, denote the word distance from the identity element.

G.1 Group Action on a Metric Space

G.1.1 The set-up

Here $\widetilde{X}$ will be some metric space on which Γ acts isometrically and discretely; that is, every $x \in \widetilde{X}$ has a neighbourhood U such that $d(U, \gamma U) > 0$, for all $\gamma \in \Gamma$, with $\gamma \neq e$, where d is the distance function of $\widetilde{X}$. The quotient mapping $\pi\colon \widetilde{X} \to X = \widetilde{X}/\Gamma$ induces then a distance on X (see Exercise 2.6).

Although most of the considerations of the appendix can be made in this general setting, we shall tacitly concentrate on just the following two cases:

(a) **The covering map of §14.2.4.4.** Here $\widetilde{G}$ is some connected Lie group, Γ is some central discrete subgroup and $\pi\colon \widetilde{G} \to \widetilde{G}/\Gamma = G$ is the canonical homomorphism. The distances are the left-invariant distances on these groups and Γ is assumed to be 0-distorted (see §14.3.3).

(b) **The uniform lattice.** Here $\widetilde{G}$ is, as before, a connected Lie group and $\Gamma \subset \widetilde{G}$ is some discrete subgroup such that the homogeneous space $\widetilde{G}|\Gamma = X =$

$(\Gamma g; g \in \widetilde{G})$ is compact. The canonical projection is denoted by $\pi\colon \widetilde{G} \to \widetilde{G}|\Gamma$ as before. The left action of Γ on $\widetilde{G}$ and the left-invariant distance on $\widetilde{G}$ are used here.

The 0-distortion In both of these cases the action of Γ on $\widetilde{X}$ is 0-distorted. In case (b) this means that $\mathrm{d}(x, \gamma x) \approx |\gamma|$, that is, that there exist fixed uniform constants such that $\mathrm{d}(x, \gamma x) \leqslant C|\gamma|$ and $c_0|\gamma| \leqslant \mathrm{d}(x, \gamma x)$, as long as x lies in some fixed compact subset whose image by π is the whole of X; see §2.14.2.

G.1.2 Covering spaces

We shall use the same notation, $\mathfrak{U} = (U_\alpha)$ as in §14.2.1, to denote an open cover of X, and assume throughout that $\mathrm{diam}(U_\alpha) \leqslant c_1$ is sufficiently small (depending on c_0). In that case $\pi^{-1}U_\alpha$ breaks into disjoint pieces $\bigcup_{\gamma\in\Gamma} \gamma\widetilde{U}_\alpha$, where $\widetilde{U}_\alpha \subset \widetilde{X}$ is open, and $\mathrm{diam}\,\widetilde{U}_\alpha \leqslant c_1$. The projection π reduces to a bijective isometry on each of these pieces.

By fixing once and for all some representative $\widetilde{U}_\alpha$ we can index $\widetilde{\mathfrak{U}} = \pi^{-1}\mathfrak{U} = (\Gamma \times \mathfrak{U})$. In this indexing it will be convenient to specify the representative $\widetilde{U}_\alpha$ further. We shall fix $\widetilde{O} \in \widetilde{X}$, some base point on $\widetilde{X}$ (the identity of G in cases (a) and (b)) and set $O = \pi(\widetilde{O}) \in X$. For $U_\alpha \in \mathfrak{U}$ we shall then choose some $\widetilde{U}_\alpha \in \widetilde{\mathfrak{U}}$ such that $\mathrm{dist}(\widetilde{O}, \widetilde{U}_\alpha) \approx \mathrm{dist}(O, U_\alpha)$ (several such choices exist). With this choice then, in the identification $\pi^{-1}\mathfrak{U} = \Gamma \times \mathfrak{U}$, for any $\widetilde{U} = (\gamma, U)$ we have

$$\mathrm{dist}(\widetilde{O}, \widetilde{U}) \approx |\gamma| + \mathrm{dist}(O, U), \tag{G.1}$$

where $\approx$ has the obvious meaning (namely, if X_i are the two sides of (G.1) then $X_i \leqslant CX_j + C$, with $i, j = 1, 2$; see Definition 11.4). With this fixed, and if the choice of c_1 for the diameters is small enough, the following holds: if $U_\alpha \cap U_{\alpha'} \neq \emptyset$ and $\gamma \in \Gamma$ then there exists one and only one $\gamma' \in \Gamma$ such that $\gamma\widetilde{U}_\alpha \cap \gamma'\widetilde{U}_{\alpha'} \neq \emptyset$. The second term in the right-hand side of (G.1) is of course only relevant in case (a).

Čech complexes With the covers $\widetilde{\mathfrak{U}}, \mathfrak{U}$ on $\widetilde{X}$ and X we can associate the Čech complexes $C^*(\widetilde{\mathfrak{U}}), C^*(\mathfrak{U})$ as in §14.2.4.4. The differentials will be denoted δ_U (in either case). The group Γ acts naturally on $C^*(\widetilde{\mathfrak{U}})$ and the projection induces $\pi^*\colon C^*(\mathfrak{U}) \to C^*(\widetilde{\mathfrak{U}})$. And it is clear that the differentials intertwine the Γ-action and the projection. Explicitly,

$$\delta_U \circ \gamma = \gamma \circ \delta_U, \quad \delta_U \circ \pi^* = \pi^* \circ \delta_U.$$

To see this think of c_1, the mesh of the cover, satisfying $c_1 \ll c_0$. Let us write $C_I^*(\widetilde{\mathfrak{U}}) = [c \in C^*(\widetilde{\mathfrak{U}});\ \gamma c = c,\ \gamma \in \Gamma]$. Then it is clear that $C_I^*(\widetilde{\mathfrak{U}}) = \pi^* C^*(\mathfrak{U})$.

Exercise Let $U = U_1 \cap \cdots \cap U_m \neq \emptyset$ for $U_j \in \mathfrak{U}$, let $\widetilde{U}_j \in \pi^{-1} U_j$ be appropriate pieces in the above decomposition chosen such that $\widetilde{U} = \widetilde{U}_1 \cap \cdots \cap \widetilde{U}_m \neq \emptyset$, and let also $c \in C^*(\mathfrak{U})$. Then $\pi^* c(\widetilde{U}) = c(U)$, where we use (14.11) to take care of the order of the indices. Use (14.11) to verify the above formulas (see Bott and Tu, 1982, Exercise 8.4 if you get stuck). The fact that the differential is independent of the ordering of the indices is something that is standard in simplicial homology, where the order of the vertices of a simplex gives the *orientation* of the simplex (see standard references on algebraic topology, e.g. Hilton and Wylie, 1960).

This way of using the 'orientation' (i.e. the ordering of the index set of the covering) by invoking (14.11) is 'cavalier' to say the least and so also therefore are the above explanations about the differential. This is already the attitude adopted in Bott and Tu (1982) and probably it is the best that can be done if one is not prepared to devote a whole section to the use of the boundary operator in oriented (abstract) simplicial complexes in algebraic topology. The reader who is not happy with this should consult standard references in algebraic topology, for example Hilton and Wylie (1960, §2.3) or Godement (1958, Chapter 3).

Čech complexes of polynomial growth These complexes, $C^*_{\mathrm{P}}(\widetilde{\mathfrak{U}}), C^*_{\mathrm{P}}(\mathfrak{U})$ (alternatively written as $C^*(\widetilde{\mathfrak{U}}; \mathrm{Pol.})$), can be defined exactly as in §14.2.5. Similarly, we can define $C^*(\Gamma \times \mathfrak{U}; \mathrm{Pol.})$ for the product distance on $\Gamma \times \mathfrak{U}$, as the one used in (G.1). The equivalence (G.1) then shows that $C^*_{\mathrm{P}}(\widetilde{\mathfrak{U}}) = C^*_{\mathrm{P}}(\Gamma \times \mathfrak{U})$.

Good covers The cover $\widetilde{\mathfrak{U}}$ will be throughout a good cover of the group $\widetilde{G}$ with all the properties of §14.2.2. In case (a), this was explained in §14.2.4.4. In case (b), we use the Riemannian structure on $G|\Gamma$ induced by the left-invariant Riemannian structure on $\widetilde{G}$ and the compactness to make a construction as we did in §14.2. We shall skip the details.

G.1.3 Informal description of the problem

Everything that will be done in this appendix stems directly from the relations that exist between the homologies of the two complexes $C^*(\widetilde{\mathfrak{U}})$ and $C^*(\mathfrak{U})$. It was for this purpose that H. Hopf introduced the notion of homology of the discrete group Γ. (This incidentally has since become a subject of its own; see, for example, Brown, 1982.)

Here we shall examine exactly the same problem in terms of the polynomial homologies and see how these relate to the homology of Γ. Few things can be proved and many more interesting problems naturally arise. We believe there is scope for further development.

To describe the new difficulty that arises in two lines, think of the dual situation where we consider $\widetilde{F} = F(\widetilde{\mathfrak{U}})$ and $F = F(\mathfrak{U})$, the corresponding subcomplexes of the finite chains. Then the key algebraic fact that one exploits in the classical theory is that $\widetilde{F} = F \otimes \Lambda(\Gamma)$, where $\Lambda = \mathbb{R}\Gamma$ is the group algebra and is defined in §G.2.2 below, that is, that $\widetilde{F}$ is a free Λ-module and that we can take as a Λ-basis an $\mathbb{R}$-basis of F.

The analogous fact for the two complexes S and $\widetilde{S}$ of Čech chains on G and $\widetilde{G}$ of superpolynomial decay (this is the analogue of the finite complexes; precise definitions will be given below) no longer holds.

The fascinating thing here, however, is that we can save the day by using the topological tensor product and we have $\widetilde{S} = S \hat{\otimes} \mathscr{S}(\Gamma)$ where $\mathscr{S}(\Gamma)$ is the algebra of functions on Γ of superpolynomial decay (see §G.3.1 below), that is, the completion of $\Lambda(\Gamma)$ (in the sense of topological vector spaces) for an appropriate metrisable topology.

The development of the topological tensor product, as well as applications to algebraic topology, is the subject matter of A. Grothendieck's thesis (see Grothendieck, 1952; Schwartz, 1953) and as I have already said, Grothendieck's work in this area has been of great inspiration to me.

G.2 The Group $\Gamma = \mathbb{Z}$

In this section Γ will be alternative notation for the group of integers $\mathbb{Z}$.

G.2.1 Elementary complexes

Let C^0, C^1 be two copies of the vector space of real sequences indexed by the integers $\prod_{i\in\Gamma} \mathbb{R}_i$. More precisely, we shall denote $C^0 = (c_i;\ i \in \Gamma)$, $C^1 = (c_{i,i+1};\ i \in \Gamma)$, where the notation used is suggested by the identification that we can make of $C^p = C^p(\mathfrak{U};\ \mathbb{R})$ with the Čech chains of §14.2.1 for the cover of the manifold $\mathbb{R}$ given by a succession of congruent intervals $\ldots, c_0, c_1, \ldots$ with the intersections $c_i \cap c_{i+1} = c_{i,i+1}$. The notation c_i, $c_{i,i+1}$ will be used to denote the coordinates of the chains c in C^0 or C^1 respectively (see Figure G.1).

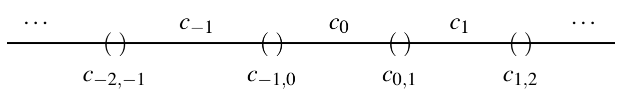


Figure G.1

We shall then define the mapping

$$\delta = \delta_C : C^0 \to C^1; \quad (\delta c)_{i,i+1} = c_{i+1} - c_i; \quad c \in C^1,$$

which is the Čech differential of §14.2.1. This gives rise to the complex

$$C^* : 0 \to C^0 \xrightarrow{\delta} C^1 \to 0. \tag{G.2}$$

Since $\operatorname{Ker}\delta$ is the one-dimensional space of constant sequences, we have the homology $H_0 = \mathbb{R}$. Also $H_p = 0$ for $p \neq 0$.

Exercise Show that $H_1 = 0$. The other values of p are clear. Let

$$C^1_+ = (\ldots, 0, 0, c_{0,1}, c_{1,2}, \ldots), \quad C^1_- = (\ldots, c_{-2,-1}, c_{-1,0}, 0, 0, \ldots)$$

be subspaces so that $C^1 = C^1_- \oplus C^1_+$. We define

$$C^0 \ni h_+(c) = (\ldots, 0, 0, c_{0,1}, c_{0,1} + c_{1,2}, c_{0,1} + c_{1,2} + c_{2,3}, \ldots)$$

for $c \in C^1_+$, where the 'first non-zero' term is at the coordinate $i = 1$, and

$$C^0 \ni h_-(c) = (\ldots, c_{-2,-1} + c_{-1,0}, c_{-1,0}, 0, 0, \ldots)$$

for $c \in C^1_-$, where the 'final non-zero' term is at $i = -1$. If we set $h\colon C^1 \to C^0$, the linear mapping that coincides with $\pm h_\pm$ on $C^0_\pm$ (abusing notation, this could be 'thought of' as $h = h_+ - h_-$), we obtain the required inverse operator $\delta \circ h =$ Identity that proves the assertion. In the terminology of homological algebra, h is a homotopy operator.

Notice finally that the group Γ acts on the complex C^* by the shift in the coordinates, for example $(\gamma c)_i = c_{i+\gamma}$, for $c \in C^0$. This action intertwines with the differential δ, that is, $\gamma \circ \delta = \delta \circ \gamma$, but not of course with h.

Remark (Generalisation to $\Gamma = \mathbb{Z}^n$) We can obtain the same type of complex with a tessellation of $\mathbb{R}^n$ by translates of the unit cube as in Figure G.1. We shall denote by $\mathfrak{U}$ the corresponding open cover as in Figure G.1. Doing the construction of the homotopy directly is now more involved. One can on the other hand use the de Rham cohomology of $\mathbb{R}^n$, namely $H_{\mathrm{DR}}(\mathbb{R}^n) \simeq H^*(\mathfrak{U})$, and use the homotopy supplied by the Poincaré lemma for the de Rham complex (see Exercise 14.5). Then one can use the general theory to pass from this to the homotopy for the Čech complex (see Bott and Tu, 1982, remark in §9.7). None of this, however, will be of any use to us here.

The following subcomplexes will also have to be considered:

$$\begin{aligned} F_* = F_*(\Gamma) : 0 \longrightarrow F_1 \xrightarrow{\delta} F_0 \longrightarrow 0; \quad \delta = \delta_F, \\ \mathscr{E}^* : 0 \longrightarrow \mathscr{E}^0 \xrightarrow{\delta} \mathscr{E}^1 \longrightarrow 0; \quad \delta = \delta_{\mathscr{E}}, \end{aligned} \tag{G.3}$$

with $F_1, \mathscr{E}^0 \subset C^0$, $F_0, \mathscr{E}^1 \subset C^1$, and with F_i consisting of the sequences of finite support (i.e. all but finitely many coordinates are zero) and $\mathscr{E}^i$ are the sequences of polynomial growth (e.g. $|c_n| = O(|n|^C)$. The notation $\mathscr{E}$ comes from distribution theory; see Schwartz (1957) or §G.5.1.2 below. Also we have switched the degrees 0 and 1 in the F_* complex to make it correspond to the dimensions of a simplicial decomposition of $\mathbb{R}$ that the reader should visualise for themselves.

What is clear is that the homotopy operator of the above exercise operates on $\mathscr{E}^*$; that is, $h(\mathscr{E}^1) \subset \mathscr{E}^0$, and therefore, as in the complex C^*, we also have $H^0(\mathscr{E}^*) = \mathbb{R}$ and $H^p(\mathscr{E}^*) = 0$ for $p \neq 0$.

For the complex F_* we have $H_0(F_*) = \mathbb{R}$ and $H_p(F_*) = 0$ for $p \neq 0$. The next exercise shows this.

Exercise Let $\varepsilon(c) = \sum_i c_{i,i+1} \in \mathbb{R}$ for $c \in F_0$. Then $\operatorname{Im}\delta_F = \operatorname{Ker}\varepsilon$.

This is seen by a different homotopy operator: $h_F(\operatorname{Ker}\varepsilon) \subset F_1$ with $h_F(c) = (\sum_{i \geqslant p} c_{i,i+1};\ p \in \Gamma)$, and from this it follows that $\operatorname{Im}\delta_F = \operatorname{Ker}\varepsilon$. On the other hand, clearly $\operatorname{Ker}\delta_F = 0$, and our assertion follows.

G.2.2 A short digression: homology of the discrete group Γ

Let $\Lambda = \mathbb{R}\Gamma$ be the group ring with real coefficients; that is, $\Lambda = \sum_{\gamma\in\Gamma} \alpha_\gamma \gamma$, with $\alpha_\gamma \in \mathbb{R}$ finite sums and where the group multiplication and the linearity over $\mathbb{R}$ induces the multiplication. For every Λ-module N (this is a real vector space that admits a Γ-action) we can define the new complex $F_*(\Gamma) \otimes_\Lambda N$ ($= 0 \to F_1 \otimes_\Lambda N \to F_0 \otimes_\Lambda N \to 0$) and the homology of this complex is written $H_*(\Gamma;\ N)$: the homology of the group Γ (cf. Hilton and Wylie, 1960, §10.7), with values in N. Clearly, here $H_p = 0$ if $p \neq 1, 0$.

Assume now that the action of Γ on N is trivial, that is, $\gamma n = n$, with $\gamma \in \Gamma$, $n \in N$, and denote $H_*(\Gamma) = H_*(\Gamma;\ \mathbb{R})$ with trivial action of Γ on $\mathbb{R}$. Then $H_*(\Gamma \otimes_\Lambda N) = H_*(\Gamma) \otimes N$; the $\otimes$ without the index means $\otimes$ over $\mathbb{R}$. This is seen by the use of a basis $N = \sum_i \mathbb{R}_i$, countable or not, where $\mathbb{R}_i \cong \mathbb{R}$. In our case it is also clear that $\dim H_1(\Gamma) \leqslant 1$ (it actually is 1 but we do not need this). We also have $H_0(\Gamma) = \mathbb{R}$.

Exercise We can verify this directly without recycling abstract definitions from homological algebra (see Cartan and Eilenberg, 1956, §X.4) as follows. It is clear that $F_\alpha \otimes_\Lambda N = N$ ($= (\ldots, 0, 1, 0, \ldots) \otimes N$)) for $\alpha = 0, 1$ because we can use the Γ-action to shift the non-zero coordinates to $i = 0$. On the other hand, $\delta_F(\ldots, 0, 1, 0, \ldots) = (0, 0, 1, -1, 0, 0, \ldots)$ and therefore

$$\delta_F \otimes \operatorname{Id}\big((\ldots, 0, 1, 0, \ldots) \otimes n\big) = (\ldots, 0, n, -n, 0, \ldots) = 0 \text{ in } F_0 \otimes_\Lambda N.$$

This implies that $\mathrm{Im}(\delta_F \otimes \mathrm{Id}) = 0$ and $H_0(F_* \otimes_\Lambda N) = N$.

G.2.3 Covering spaces with group $\Gamma = \mathbb{Z}$ and the Cartan–Leray spectral sequence

Here we shall go back to the covering map $\widetilde{G} \to \widetilde{G}/\Gamma = G$ (where $\Gamma = \mathbb{Z}$) of case (a) of §G.1.1. The cover $\mathfrak{U}$ of G will be as in §14.2.4.3 and it will be sufficiently fine for all the considerations of §G.1.2 to hold. The corresponding cover on $\widetilde{G}$ will be $\widetilde{\mathfrak{U}}$; cf. also §14.2.4.4.

The double complex that we shall consider is $M = F_* \otimes_\Lambda C^*(\widetilde{\mathfrak{U}})$, where the Γ-action on $C^*(\widetilde{\mathfrak{U}})$ is as in §G.1.2. From this we shall construct the two spectral sequences as in §14.4. The fact that the arrows in F go the wrong way only makes notational differences in §14.4 and anyway we can put the arrows straight by setting $F^{-i} = F_i$, but then we obtain second quadrant spectral sequences. (In (G.8) below we shall refer to Bott and Tu, 1982, Exercise 12.12.1 and there one finds a similar reversal of the arrows which happens for the same reason.)

We consider first the 'first' spectral sequence of §14.4.2, that is, the one where we start by taking the $\mathrm{d}'' = \delta_U$ on M. We clearly obtain $''E_1^{**} = F_*(\Gamma) \otimes_\Lambda H^*(\widetilde{\mathfrak{U}})$. But we have already seen that the Γ-action on $H^*_{\mathrm{DR}}(\widetilde{G})$ and therefore also on $H^*(\widetilde{\mathfrak{U}})$ is trivial (see the exercise in §14.3.2). From §G.2.2 it therefore follows that

$$''E_2^{**} = H_*(\Gamma) \otimes H^*(\widetilde{\mathfrak{U}}). \tag{G.4}$$

Now for the 'second' spectral sequence, where we start with d' which is induced by δ_F. We shall use the fact that $\Lambda \otimes_\Lambda N = N$ for any Γ-module N; this is automatic by the definition of $\otimes_\Lambda$. It will also be convenient to use §G.1.2 and identify $\widetilde{\mathfrak{U}}$ with $\Gamma \times \mathfrak{U}$. This allows us to identify $C^*(\widetilde{\mathfrak{U}})$ with C^* which is a space of double arrays $c(\gamma, U)$, where $\gamma \in \Gamma$ and U is a finite intersection of open sets of $\mathfrak{U}$. The double complex M consists of just the first two rows $C^*_{(0)}$ and $C^*_{(1)}$, each identical to C^*; and d' induces $C^*_{(0)} \xrightarrow{\delta_F \otimes \mathrm{Id}} C^*_{(1)}$. With this differential we have

$$C^*_{(0)} \ni \big(c(\gamma, U)\big) \to \big(c(\gamma+1, U) - c(\gamma, U)\big) \in C^*_{(1)}.$$

Therefore $\mathrm{Ker}(\delta_F \otimes \mathrm{Id})$ consists of all the $\big(c(\gamma, U)\big)$ that are independent of γ. Hence this kernel and so also $'E_1^{0,*}$ can be identified with $C^*(\mathfrak{U})$, the Čech complex of $\mathfrak{U}$ on G (see the definition of C^*_{I} in §G.1.2). But $'E_1^{1,*} = 0$ (and also, trivially, $'E_1^{p,*} = 0$ for $p \neq 0, 1$), because the homotopy operator h of the exercise from §G.2.1 can be used on just the first coordinate γ of $c(\gamma, U)$.

Furthermore, from §G.1.2 it follows that the second differential $\mathrm{d}'' = \delta_U$ on $C^*(\widetilde{\mathfrak{U}})$ induces on ${}'E_1^{0*} = C^*(\mathfrak{U})$ the Čech differential δ_U with respect to $\mathfrak{U}$.

The bottom line is that we are in the situation where (14.29), (14.30) apply so that with a slight abuse of notation we have

$$
{}'E_2^{0,n} = \cdots = {}'E_\infty^{0,n} = H^n(\mathfrak{U}) = \sum_{p+q=n} {}''E_\infty^{p,q} \tag{G.5}
$$

for some filtration of the cohomology. For more on this Cartan–Leray spectral sequence see McCleary (2001, §8$^{\mathrm{bis}}$.9).

If we combine (G.4) and (G.5) and the abstract considerations of Exercise 14.8 in §14.4.5 we obtain the required conclusion. Namely that in the covering $\pi\colon \widetilde{G} \to \widetilde{G}/\Gamma = G$ the cohomology $H^*(\widetilde{\mathfrak{U}};\widetilde{G})$ is finite-dimensional if and only if the cohomology $H^*(\mathfrak{U};G)$ is finite-dimensional. This point is elaborated on further in the example of §G.5 below, where we deal with a situation that is more involved.

In the above we have worked with the ordinary Čech cohomology or, equivalently, the ordinary de Rham cohomology, and, because of (12.1), we hardly need to do all this work to prove the finiteness of the cohomology. On the other hand, by the above consideration, we are only a step away from the equivalence (14.22) on polynomial cohomologies. This was the main geometric motivation of this appendix.

The polynomial cohomology For this we apply exactly the same steps on the polynomial version of the previous double complex $M_P = F_* \otimes_\Lambda C_P^*(\widetilde{\mathfrak{U}})$, using the notation of §G.1.2 for the polynomial complex. The constructions that we have made are verbatim the same here, and we shall leave to the reader the task of writing out the details. The following simple observation has to be taken into account: since $C_P^*(\widetilde{\mathfrak{U}}) = C_P^*(\Gamma \times \mathfrak{U})$, as explained at the end of §G.1.2, it follows that the homotopy which we used to show that ${}'E_1^{1,*} = 0$ can be used again. This holds because the explicit formula that implies that homotopy gives $h(\mathscr{E}) \subset \mathscr{E}$: see §G.2.1.

G.3 Discrete Groups

G.3.1 Notation and free resolutions

From here until the end of the appendix, Γ will denote some finitely generated group and $|\gamma| = \mathrm{d}_\Gamma(\mathrm{e},\gamma)$ will be the word distance from the identity (see §1.1). Together with the group ring $\Lambda = \mathbb{R}\Gamma$ of §G.2.1, we shall use $\Lambda_s = \mathscr{S}(\Gamma)$, the larger ring of all sums $x = \sum_{\gamma\in\Gamma} \alpha_\gamma \gamma$ for which

$$\|x\|_C = \sum_\gamma |\alpha_\gamma|\ |\gamma|^C < +\infty \quad \text{for all } C > 0.$$

Exercise Verify that $\mathscr{S}(\Gamma)$ is closed under multiplication because $|\gamma_1\gamma_2| \leqslant |\gamma_1| + |\gamma_2|$. But for our purposes what counts is that it is a module over Λ. We shall call this the space of functions of rapid decay on Γ.

In the homology theory for discrete groups one says that Γ is an (F)-group if it admits a free resolution of finite type. More explicitly, there exists

$$F_* = F_*(\Gamma):\ \cdots \longrightarrow F_n \xrightarrow{\delta} F_{n-1} \xrightarrow{\delta} \cdots F_0 \longrightarrow 0; \quad \delta = \delta_F,\ \delta^2 = 0, \quad \text{(G.6)}$$

where $F_j = V_j \otimes \Lambda$ for *finite-dimensional* real vector spaces $V_0, V_1, \dots,$ and such that the homology of (G.6) satisfies $H_p = 0$ for $p \neq 0$, and $H_0 = \mathbb{R}$. To stay closer to the terminology used in homological algebra, the F_j are free Λ-right modules and the Λ-action on $H_0 = \mathbb{R}$ is trivial, that is, $x\gamma = x$, $\gamma \in \Gamma$, $x \in H_0$. One then says that (G.6) is a Γ-free resolution of $\mathbb{R}$. We shall say that Γ is an $(\mathscr{S})$-group if it admits a resolution as in (G.6) for which in addition the complex

$$S_* = S_*(\Gamma):\ \cdots S_n \xrightarrow{\delta} S_{n-1} \longrightarrow \cdots S_0 \longrightarrow 0; \quad \delta = \delta_S, \qquad \text{(G.7)}$$

with $S_n = F_n \otimes_\Lambda \Lambda_s = V_n \otimes \Lambda_s$ and the induced differential, also satisfies, for its homology, $H_0 = \mathbb{R}$, and $H_p = 0$ for $p = 0$. Here Λ_s is considered a left Λ-module and, in the terminology of homological algebra, we are simply saying that $H(\Gamma;\Lambda_s) = \mathbb{R}$.

Example The resolutions (G.3) for $\mathbb{Z}$ satisfy these properties because the mapping ε and the homotopy h_F in the exercise of §G.2.1 adapt in an obvious way to the corresponding S_*.

Using algebraic or topological methods, many things can be proved about (F)-groups. For example, finite groups, finitely generated Abelian groups etc. are (F)-groups (see Serre, 1970; Brown, 1982). We say that a group is strictly polycyclic (Ragunathan, 1972) if a finite composition series $\Gamma \supset \Gamma_1 \supset \cdots$ exists such that $\Gamma_i/\Gamma_{i+1} \cong \mathbb{Z}$. These groups are (F)-groups.

In the next subsection we shall put together some observations on the above notions that are related to the subject of the book. We shall be brief and for the proofs in the subsequent section we shall be even briefer. As they stand, these results are mere curiosities. One can, however, speculate whether these $(\mathscr{S})$-groups can be used as a substitute to obtain a B–NB classification for discrete groups that is related to random walks, as is the case for Lie groups.

G.3.2 The B–NB classification for lattices

Now we shall concentrate on case (b). We shall denote by $\Gamma \subset \widetilde{G}$ some uniform lattice in some simply connected soluble Lie group; $\widetilde{G}$ is a model. Recall that $\widetilde{G}|\Gamma$ is compact and so the above condition implies that $\widetilde{G}$ is unimodular because $\widetilde{G}|\Gamma$ carries an invariant measure (see Bourbaki, 1963, Chapter 7, §2.6; Ragunathan, 1972, §§1.4, 1.11, 3.1). Then we have the following *dichotomy*:

(i) *If* $\widetilde{G}$ *is an* NC-*group then* Γ *is an* $(\mathscr{S})$-*group*.
(ii) *Conversely, when* Γ *is a uniform lattice in some model, then this model is* NC *provided that the homology* $H(\Gamma;\mathscr{S}(\Gamma))$ *is finite-dimensional*.

Many things can be said about lattices in models (see Ragunathan, 1972). Here are a few of the facts:

(1) Every such lattice is strictly polycyclic and therefore finitely generated and torsion-free.
(2) (Mal'cev) Every finitely generated torsion-free nilpotent group N can be realised as a uniform lattice on some simply connected nilpotent Lie group. It follows that N is an $(\mathscr{S})$-group.
(3) There is a converse to (2). Indeed, if Γ is a lattice in some NC-model $\widetilde{G}$, then by the unimodularity of $\widetilde{G}$ and §B.6 it follows that $\widetilde{G}$ is an R-group and therefore that Γ is a finite extension of a nilpotent group (by Gromov's theorem, see §1.2, on the growth function $\gamma(n)$ for Γ among other things, but of course here Γ is actually soluble and the whole thing is much easier).
(4) All lattices as above are (F)-groups and by the above we see that they are not all $(\mathscr{S})$-groups; cf. Ragunathan (1972, §§4.28–4.29). In fact, putting everything together we see that for a polycyclic group to be $(\mathscr{S})$ it has to be *nilpotent by finite* and this 'essentially' characterises the condition.
(5) $\dim H(\Gamma;\mathscr{S}) < +\infty$ holds for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ or if Γ is a free group. The proof of this will not be given. See, however, Serre (1970) and §G.4.2 below.

Recall that the condition on the homology $H(\Gamma;\mathscr{S})$ says this: there exists a free resolution of Γ as in (G.6) – but we do not necessarily assume that the spaces V_j are finite-dimensional – for which the corresponding complex (G.7) has finite-dimensional homology. Notice that then, this also holds for every other free resolution of Γ. The fact that we can pass from one free resolution to another is an easy standard, but critical, fact in homological algebra (see Cartan and Eilenberg, 1956, §V.1).

Problem (Difficult) How about random walks? How about $H(\Gamma;\mathscr{S})$ when Γ is some lattice (not necessarily uniform) in a semisimple Lie group (see Ragunathan, 1972)?

G.4 Outline of the Proofs

We give only an outline, and this in a style that would probably only be right for seminar notes. However, many new ideas are introduced and the prospects for further research and development are obvious. It is this that justifies the writing of this section. But it is this that also explains the style of writing because to present these new ideas in a more acceptable manner would have forced me to introduce a tremendous amount of background material (that would have amounted to a new chapter).

G.4.1 Čech homology

This is a covariant functor, as opposed to the Čech cohomology of §14.1 which is a contravariant functor. (In the ‘specialist literature’ such creatures are referred to as ‘pre-cosheafs’. In this book, however, we follow the lead of Bott and Tu, 1982, and ignore several aspects of the specialised literature!)

For general definitions see Eilenberg and Steenrod (1952) and also Bott and Tu (1982, Exercise 12.12.1) for C^∞ manifolds M. In what follows we briefly recall the definition when the manifold M is a Lie group (and so is orientable, for people who know what that means). For the case of rapid decay, in §G.4.2 below we shall even restrict ourselves to the models $\widetilde{G}$ that admit a lattice $\Gamma \subset \widetilde{G}$ as in §G.3.2.

For these manifolds a constant presheaf can then be defined by $\mathscr{H}(U) = H_c^n(U)$, which is the de Rham cohomology with compact support, where $U \subset M$ runs through all open subsets that are diffeomorphic to $\mathbb{R}^n$. This is similar to what we did in §14.1; the main difference is that for $U \subset V$ the isomorphisms $i_U^V\colon \mathscr{H}(U) \to \mathscr{H}(V)$ are covariant (it is the direction of the arrow that does that). Using this presheaf, which is constant (analogous to the definitions in §14.2.1) because of the orientation, we shall define $C_*(\mathfrak{U}, \mathscr{H}) = \sum_{p\geqslant 0} C_p$, the Čech homology complex, where $\mathfrak{U}$ is a good cover of the Lie group that has all the properties from §14.2.4.3. The differential $\delta\colon C_{p+1} \to C_p$ is defined here by

$$(\delta c)_{\alpha_0,\ldots,\alpha_p} = \sum_{\alpha} c_{\alpha,\alpha_0,\ldots,\alpha_p}; \quad c \in C_{p+1},$$

and by the properties of the cover this sum is always finite; under the summation sign we have suppressed the obvious i_U^V mapping and used (14.11).

G.4.2 Chains of rapid decay

As we said we shall restrict ourselves to $\widetilde{G}$ of §G.3.2. In that case $\widetilde{\mathfrak{U}}$ and $\mathfrak{U}$ (which is finite) will be (good) covers of $\widetilde{G}$ and $\widetilde{G}|\Gamma$ respectively, and the notation $\widetilde{\mathfrak{U}} = \pi^{-1}\mathfrak{U}$ is as in §G.1.2. Furthermore, $C_*(\widetilde{\mathfrak{U}}, \mathscr{H}) = C(\Gamma) \otimes C_*(\mathfrak{U})$ where $C(\Gamma)$ is the space of functions on Γ and $C_*(\mathfrak{U})$ is the Čech complex on $G|\Gamma$ with respect to the finite cover $\mathfrak{U}$. Note that $C_*(\mathfrak{U})$ is a finite-dimensional space. The chains of rapid decay will then, by definition, be the subspace $C_*(\widetilde{\mathfrak{U}}, \mathscr{H}, \mathscr{S}) = \mathscr{S}(\Gamma) \otimes C_*(\mathfrak{U})$.

A general definition (for any manifold) can be given for these rapid decay chains, by imitating the work that was done in §14.2.5, but we shall leave it to the reader to do this.

Proof of the dichotomy With the above notation the following Poincaré duality holds (this needs proving: see Bott and Tu, 1982, Exercise 12.12.1 for the proof in the classical case with compact cohomology H_c on the left and Čech chains of finite support; the proof extends to the case of chains of rapid decay as in Definition G1 below):

$$H^p{}_{\mathrm{DR}}(\widetilde{G}, \mathscr{S}) = H_{n-p}\big(C_*(\widetilde{\mathfrak{U}}, \mathscr{H}, \mathscr{S})\big). \tag{G.8}$$

To clarify the meaning of the left-hand side when $\widetilde{G}$ is NB we recall that then $\widetilde{G} \approx \mathbb{R}^n$ (polynomial equivalence as in Appendix F) because it is an R-group and the left-hand side of (G.8) stands for the cohomology of the complex $\Omega^*(\mathbb{R}^n, \mathscr{S})$, that is, of smooth differential forms ω on $\widetilde{G} \approx \mathbb{R}^n$ that together with $\mathrm{d}\omega$ decay rapidly (i.e. $|\omega(x)|$, $|\mathrm{d}\omega(x)| = O\big((1+|x|)^{-c}\big)$ for all $c > 0$). By the same proof (see Bott and Tu, 1982, §4.7.1) as for the case of compact cohomology $H_c^p(\mathbb{R}^n)$ for the forms of compact support $\Omega_c^*(\mathbb{R}^n)$, we can then easily show that the left-hand side of (G.8) is $\mathbb{R}$ for $p = n$ and zero otherwise.

Now, whether $\widetilde{G}$ is NB or not, the complex C_* on the right-hand side of (G.8) is $\mathscr{S}(\Gamma) \otimes F_*$ where F_* is the subcomplex of $C_* = C_*(\widetilde{\mathfrak{U}}, \mathscr{H}, \mathscr{S})$ of the chains of finite support. Furthermore, by classical Poincaré duality (see Bott and Tu, 1982), the fact that topologically $\widetilde{G} \approx \mathbb{R}^n$ implies that F_* is a Γ-free resolution of $\mathbb{R}$ as in (G.6) for which the corresponding (G.7) is none other than the complex C_*. Putting things together, part (i) of the dichotomy follows directly from (G.8).

This also proves part (ii) of the dichotomy because when Γ satisfies the condition in (ii), (G.8) shows that (see the end of §G.3.2)

$$\text{dimension of the homologies in (G.8)} = \dim H^*_{\mathrm{DR}}(\widetilde{G}; \mathscr{S}) < +\infty \tag{G.9}$$

for the de Rham homology of the complex of smooth forms of rapid decay $\Omega^*(\widetilde{G}; \mathscr{S})$; see (G.8) and Definition G1 below.

The final step is to deduce from (G.9) that $\widetilde{G}$ is an NC-group. This is contained in Chapter 12 and is the hardest and longest step in this chain of arguments. In Chapter 12 we did not work on the complex of smooth forms of rapid decay $\Omega^*(\widetilde{G};\mathscr{S})$, but on the corresponding complex of currents (see §12.7). To pass to smooth forms we have to use some smoothing procedure as in §12.10.

Definition G1 We could adopt several definitions for $\Omega^*(\widetilde{G};\mathscr{S})$. For instance, we could demand that the forms in that space satisfy

$$\int_{\widetilde{G}} \big(|\omega(g)| + |\mathrm{d}\omega(g)|\big)\,|g|^c\,\mathrm{d}g < +\infty; \quad c > 0,$$

the integral being taken with respect to left Haar measure and the left Riemannian structure is used in the definition of $|\omega|$, $|\mathrm{d}\omega|$ and $|g|$ (note that here $\widetilde{G}$ is unimodular).

Several variants of this definition could also be given. Notice, for instance, that this is not the same as the one we gave for $\Omega^*(\mathbb{R}^n;\mathscr{S})$ in the first part of the proof. The only thing that counts in the definition is that $H^p_{\mathrm{DR}}(\widetilde{G};\mathscr{S}) = H_{n-p}(\widetilde{\mathfrak{U}};\mathscr{H},\mathscr{S})$ as in (G.8) and that we can use the results of Chapter 12. Notice that in full generality (G.8) should be compared with (14.37), but to prove that this works the analogous verifications of §14.2.4.2 are needed. The discussion on Poincaré duality of §12.11.2 and Remark 12.53 could then be seen in this light.

G.5 A Variation on the Same Theme

We shall outline here a variant of the circle of ideas that we have developed in this appendix. In order not to interrupt the presentation, terminology from the theory of topological vector spaces will only be recalled at the end of the section. Even so, only those readers who are comfortable with functional analysis will enjoy these variations.

We shall place ourselves in the context of case (a) in §G.1.1 with $\pi\colon \widetilde{G} \to \widetilde{G}/\Gamma = G$ for some central 0-distorted discrete subgroup Γ that is therefore a finitely generated Abelian group (see §4.6.1). The notation for the Čech complexes are as before and the covers $\mathfrak{U}$, $\widetilde{\mathfrak{U}}$ are as in §G.1 and satisfy the good properties of §14.2.4.3. We then have

$$C_*(\widetilde{\mathfrak{U}},\mathscr{S}) = C_*(\widetilde{\mathfrak{U}},\mathscr{H},\mathscr{S}) = \mathscr{S}(\Gamma)\widehat{\otimes} C_*(\mathfrak{U},\mathscr{H},\mathscr{S}) = \mathscr{S}\widehat{\otimes} C_*(\mathfrak{U},\mathscr{S}). \quad \text{(G.10)}$$

The new feature here is that on the spaces $\mathscr{S}$ and $C_*(\cdot,\mathscr{H},\mathscr{S})$ we have assigned the Fréchet topology induced by the seminorms $\|x\|_C$ of §G.3.1 and their analogue for C_*. In (G.10) we have completed the tensor product for the projective topology $\otimes_\pi$ (hence the hat '$\widehat{\ }$'). The proof of (G.10) – it is the middle relation that needs proving as the other two amount to notation – is then an easy consequence of the definitions. (This holds because the $\mathscr{S}$-topology is a projective limit of L^1 topologies and projective limits behave well under $\otimes_\pi$; see Schwartz, 1953, Exposée 7, Proposition 5 or Jarchow, 1981).

We shall also assume in the free resolution $F_*(\Gamma)$ of (G.6) that not are only the dimensions of the V_j finite, but also the resolution is of finite length (i.e. $V_j = 0$ for j large enough). This is certainly the case when $\Gamma \simeq \mathbb{Z}^k$ (cf. fact (2) in §G.3.2). We shall consider now a variant of the double complex of §G.2.3 which can be written

$$M = F_*(\Gamma) \otimes_\Lambda C_*(\widetilde{\mathfrak{U}}, \mathscr{S}) = \sum_{p \geqslant 0} S_p \widehat{\otimes} C_*(\mathfrak{U}, \mathscr{S}) = S_*(\Gamma) \widehat{\otimes} C_*(\mathfrak{U}, \mathscr{S}) \quad \text{(G.11)}$$

and on this we shall play the same game that we did twice before (in §§G.2 and G.4).

Explicitly, we use the first expression in (G.11) to work out the first spectral sequence (where we start with the second differential $\mathrm{d}'' = \delta_U$). We pick up

$$''E^1_{p,q} = F_p \otimes_\Lambda H_q(\widetilde{\mathfrak{U}}, \mathscr{S}),$$

and the action of Γ on $H(\widetilde{\mathfrak{U}}, \mathscr{S})$ is trivial (see the exercise in §14.3.2 and §G.4). It follows that (cf. §G.2.2)

$$''E^2_{p,q} = H_p\big(F_*(\Gamma) \otimes_\Lambda H_q(\widetilde{\mathfrak{U}}, \mathscr{S})\big) = H_p(\Gamma) \otimes H_q(\widetilde{\mathfrak{U}}, \mathscr{S}). \quad \text{(G.12)}$$

Now we pass to the second spectral sequence $'E^*_{*,*}$, that is, the one we get by starting with the first differential $\mathrm{d}' = \delta_F$. This amounts to applying the δ_S differential in the third expression of M in (G.11). What we obtain for this is

$$'E^1_{0,*} = C_*(\mathfrak{U}, \mathscr{S}); \quad 'E^1_{p,*} = 0 \text{ for } p \neq 0. \quad \text{(G.13)}$$

This needs proving. It follows from a lemma due to Grothendieck (see §G.5.1 below) which asserts that as long as $\mathscr{S}(\Gamma)$ is a nuclear topological vector space (which is the case here) we have

$$H(S_* \widehat{\otimes} C_*(\mathfrak{U}, \mathscr{S}); \delta_S) = H(S_*; \delta_S) \widehat{\otimes} C_*(\mathfrak{U}, \mathscr{S}) = C_*(\mathfrak{U}, \mathscr{S}), \quad \text{(G.14)}$$

where for simplicity let $\Gamma \simeq \mathbb{Z}$ and use §G.2.1 to see that this is an $(\mathscr{S})$-group. This proves (G.13).

Back in (G.11), the δ_U (for $\widetilde{\mathfrak{U}}$ on $\widetilde{G}$) on the left-hand side induces a mapping on M that commutes with δ_F and when we pass to the homology, as in (G.13), it induces the Čech differential of the right-hand side of (G.14). To put it differently, δ_U induces on $'E^1_{0,*}$ in (G.13) the Čech differential δ_U of $C_*(\mathfrak{U}, \mathscr{S})$. It is obvious that this 'has to be that way' but the actual verification is a *non-trivial* exercise in topological vector spaces and homological algebra. Grothendieck stuff! We shall say more on this in the next subsection.

Having done all of this, together with the abstract algebraic considerations of §14.4, we finally conclude

$$'E^2_{0*} = \cdots = 'E^\infty_{0*} = H_*(\mathfrak{U}) = H_*(C_*(\mathfrak{U}, \mathscr{S}); \delta_U) \tag{G.15}$$

for some filtration of the homology.

This is the end of this story because for the first spectral sequence we have the E^2 from (G.12) and the limit E^∞ is the same as in (G.15). From this, the rest of the work is abstract and from these spectral sequences we can draw the same conclusions exactly as before.

Example For instance, let q_0 be the last integer (if any) for which $H_q(\widetilde{\mathfrak{U}})$ is infinite, that is, dim $= \infty$. As in §G.2.3, the arrows in the spectral sequence go the 'wrong' way because in (G.7) we decrease dimensions. We must therefore resort to the same trick as in §G.2.3 and change the index from n to $-n$ (cf. Cartan and Eilenberg, 1956, §V.i and Bott and Tu, 1982, §12.2.1), and then (14.27) still holds but for a second quadrant spectral sequence. Now $H_0(\Gamma) = \mathbb{R}$ in (G.12) (cf. §G.2.2) so using therefore Exercise 14.8 again, we deduce that $E^2_{0,q_0}, E^3_{0,q_0}, \ldots$ are all infinite. Therefore E^∞ would also have to be infinite. As a consequence we see that $\dim H_*(\mathfrak{U}) < +\infty \implies \dim H_*(\widetilde{\mathfrak{U}}) < +\infty$.

Further prospects It is clear from the previous three sections that for any finitely generated discrete group Γ, the homology $H = H(\Gamma; \mathscr{S}(\Gamma))$ could turn out to be an interesting invariant and our wild speculation is that this may provide some kind of analogue, for discrete groups, of the B–NB classification. That aside we could try to look at the following concrete problems.

The question of whether $\dim H < +\infty$ was answered in §G.3.2(4) for polycyclic groups using geometric methods. Similarly, geometric methods are used for the groups of (5) in §G.3.2. Purely algebraic proofs seem to be difficult to devise. As a consequence, deciding what happens for general soluble finitely generated groups, for example, does not seem to be that easy (see McCleary, 2001,§8^{bis}.12 for more on this).

Note though that by the very definition of $\mathscr{S}(\Gamma)$ we always have

$$H_0(\Gamma; \mathscr{S}(\Gamma)) = \mathbb{R}$$

(prove this! hint: $\varphi \in \mathscr{S}$ with $\int \varphi = 0$ can be written as a sum of 'bipolars' $\lambda(\delta_x - \delta_y)$ as in §13A.3) so if $\dim H = \infty$ we have to look at higher dimensions and to a large extent it is this that makes the problem non-trivial and interesting.

Topological vector spaces A Fréchet space is a complete, metrisable, locally convex topological vector space. The topology is then defined by a countable sequence $p_i(\cdot)$ of seminorms as in §12.14. Such a space is $\mathscr{S}(\Gamma)$ and the seminorms are then given by $\|x\|_C$ in §G.4. In the esoteric terminology of the subject, $\mathscr{S}(\Gamma)$ is a Köthe sequence space (see Jarchow, 1981) and is easy to see that with this definition $\mathscr{S}(\Gamma)$ is a nuclear space if and only if Γ is a group of polynomial growth. So for our spaces $\mathscr{S}(\Gamma)$, this could be taken as the definition of nuclearity.

The *projective* topology on the tensor product $F \otimes_\pi E$ is the topology induced by the seminorms $p_i \otimes q_j$ for the seminorms p_i and q_j of F and E respectively. Here $p \otimes q(u) < 1$ if it is possible to write $u = \sum x_j \otimes y_j$ as a finite sum so that $\sum p(x_j) q(y_j) < 1$. There is another topology that can be defined on the tensor product, called the *injective* topology. The completion under that topology is denoted by $\hat{\hat{\otimes}}$ and if one of the two spaces is nuclear the projective and the injective topologies on the tensor product coincide. For all that, see the references below.

Relevant references See Grothendieck (1958) or Bourbaki (1953) for topological vector spaces, Schwartz (1953) for the theory of topological tensor products or the more recent reference Jarchow (1981) for both.

G.5.1 The Grothendieck lemma and all that…

The lemma The spaces A, B, C, E, F below are Fréchet spaces and B, E are assumed to be nuclear. Furthermore, a differential $\mathrm{d}\colon E \to E$ is given (i.e. there is a linear continuous map with $\mathrm{d}^2 = 0$) for which $B(E)$ is closed. (Here we adopt standard notation and in the presence of a differential d on the space X we denote $Z(X) = \operatorname{Ker} \mathrm{d}$, the space of cycles; $B(X) = \operatorname{Im} \mathrm{d}$; $H(X) = Z/B$ = the homology.) In Grothendieck's thesis (see Schwartz, 1953, Exposée 27, Grothendieck, 1952) one finds the proofs of the following results. Let

$$0 \to A \underset{i}{\to} B \underset{\pi}{\to} C \to 0$$

be an exact sequence (topologically, i.e. A is a closed subspace of B and $C = B/A$ with the quotient topology). Then by tensoring with F we also obtain a topological exact sequence

$$0 \to A\hat{\otimes}F \xrightarrow{i\otimes 1} B\hat{\otimes}F \xrightarrow{\pi\otimes 1} C\hat{\otimes}F \to 0.$$

(The reason for this is simple: $\hat{\otimes}$ behaves well under the projection π, and $\hat{\hat{\otimes}}$ well under the inclusion i, and by our hypothesis $\hat{\otimes} = \hat{\hat{\otimes}}$.) This can be applied to the following exact sequences:

$$0 \to Z(E) \to E \to B(E) \to 0,$$
$$0 \to B(E) \to Z(E) \to H(E) \to 0,$$

and we obtain

$$Z(E\hat{\otimes}F) = Z(E)\hat{\otimes}F, \quad B(E\hat{\otimes}F) = B(E)\hat{\otimes}F, \quad H(E\hat{\otimes}F) = H(E)\hat{\otimes}F.$$

Furthermore, if $F = F_1 \oplus F_2$ is a direct topological vector space sum, both the left- and right-hand sides naturally 'split' into the F_1 and F_2 components.

We can specialise the above to the case $H(E) = \mathbb{R}$ and to the splitting $F = \{c\} \oplus F'$, where $\{c\}$ is the one-dimensional space spanned by some $c \in F$ and F' is some complement (Hahn–Banach). The bottom line in that case is that in the identification $H(E\hat{\otimes}F) = F$ the element $c \in F$ can be represented by the cycle $z \otimes c \in Z(E) \otimes F$, where $z \in Z(E)$ is any cycle that represents the identity $1 \in H(E) = \mathbb{R}$. (We write $[z] = 1$ where $[\ ]$ denotes the homology class.)

G.5.1.1 The 'non-trivial' exercise after (G.14) The above facts will be used with $F = C_*(\mathfrak{U}, \mathscr{S})$ – the space will be denoted by C_* in what follows – and $E = S_*(\Gamma)$ which, as we have pointed out, is nuclear. Together with the Čech differential δ_U on C_* we shall define another mapping $\delta_H : C_* \to C_*$ as follows. Denoting by 1 the identity mapping, we start with the mapping $1 \otimes \delta_U$ on $\widetilde{M} = F_*(\Gamma) \otimes C_*(\widetilde{\mathfrak{U}}, \mathscr{S})$ – the $\otimes$ product over $\mathbb{R}$ – and use this to induce a mapping $\delta_M : M \to M$ (use the left-hand side of (G.11) to do this). It is easy to see that δ_M commutes with the mapping $\delta_S \otimes 1$, where we use the right-hand side in formula (G.11). This simply happens because $1 \otimes \delta_U$ commutes with $\delta_F \otimes 1$ on the original tensor product $\widetilde{M}$. It follows that δ_M induces a mapping on the homology of $\delta_S \otimes 1$ which is C_* (see (G.13), (G.14)). This is the mapping that we call δ_H. What will be proved below is that $\delta_H = \delta_U$, which was used in (G.15).

To give a concrete expression for $\delta_H c$, with $c \in C_*$, we can use the previous considerations and represent c by a cycle $z \otimes c \in M$, with $z \in Z_\Gamma(S_*(\Gamma))$, that is, for the differential δ_S, where the z has been chosen as explained above. We then have $\delta_H c = [\delta_M(z \otimes c)] \in C_*$. Note also that this cycle z can even be chosen to belong to F_0 of (G.6). This additional information has a certain 'psychological' value in the argument below but is not essential. At any rate, the reason for this is simple enough: $\operatorname{Im} \delta_S \subset S_0$ is of codimension 1 and it is also closed by the continuity of δ_S and Banach's theorem (see the exercise in §12.12.1). We then use the density of F_0 in S_0.

Be this as it may, from this concrete definition it is easy to verify that when c is a Čech chain of $\mathfrak{U}$ of finite support then $\delta_H c = \delta_U c$. From this we can then in turn verify the required conclusion that $\delta_H = \delta_U$ because the chains of finite support are dense in C_*. To be able reach that conclusion it suffices to verify that both the mappings δ_U and δ_H are continuous.

For the continuity of δ_U we use the definition of the seminorms (the analogues of the $\| \ \|$ of §G.3.1) that induce the Fréchet topology of C_*. But we also need to use the uniform local finiteness of the cover $\mathfrak{U}$ (see §14.3). So this is easy but not formal. The verification that δ_H is continuous uses the same principles but is more involved. Proceed as follows.

Let $C_* \ni c_n \to 0$ and represent, as explained earlier, the c_n by the cycles $z \otimes c_n$ where $z \in Z_\Gamma(S_*(\Gamma))$ is fixed. We can then write z as a finite sum $z = \sum e_k \otimes s_k$ where e_k are the basis elements of $\sum V_j$ (see (G.7)) and $s_k \in \mathscr{S}(\Gamma)$. It suffices therefore to show that $\delta_M(e \otimes s \otimes c_n) \to 0$ for every basis element e as above and every fixed $s \in \mathscr{S}(\Gamma)$. This, by the definition of δ_M, amounts to showing that $\delta_U(s \otimes c_n) \to 0$ where δ_U refers to $\widetilde{\mathfrak{U}}$ and where the $s \otimes c_n$ are identified with elements of $C_*(\widetilde{\mathfrak{U}}; \mathscr{S})$; see §G.1.2. For this we proceed on the covering $\widetilde{\mathfrak{U}}$ as we did just above for $\mathfrak{U}$.

The above proof of the continuity of δ_H is rather ad hoc. A more canonical way of proceeding would have been to start by proving that the mapping δ_M that we defined on M is already continuous. The proof of this follows the same lines and is an interesting exercise in topological vector spaces. For that proof one can use the fact that in the $\hat{\otimes}$-product of two Fréchet spaces E and F, one of which is nuclear, the bounded sets are just the tensor products of bounded sets in E and F (after we take closed convex bounded envelopes; see Schwartz, 1953, Exposé 19, Théorème 2, Corollaire). We then use the fact that Fréchet spaces are bornological, that is, that a linear mapping is continuous as soon as it takes bounded sets into bounded sets (cf. Bourbaki, 1953; Grothendieck, 1958).

G.5.1.2 On the role of the topological vector space With hindsight one understands why difficult aspects, and also to a large extent ‘forgotten aspects’, of topological vector spaces have to be used in the theory.

In short, homology for complexes that are topological vector spaces has to be considered in our theory because one could almost say that this is what the theory is all about. As a consequence the tensor product is forced into the game and it *has* to be the ‘topological tensor product’.

This is what Grothendieck developed in his thesis (before he moved on to algebraic geometry). That kind of mathematics was very fashionable in the early 1960s.[1] Since then, as often happens with mathematical fashions (and

[1] I was a student in Cambridge and I studied a lot of that!

other fashions as well!), much of this has been forgotten and is now familiar only to a few specialists in the subject.

While writing the third part of this book I came across a problem in topological vector spaces (and distribution theory) that I found intriguing and that I was not able to do – but of course it is a long time since I was an eager graduate student in Cambridge! Here it is:

Problem With the notation used in distribution theory, $\mathscr{D}(\mathbb{T}) \cong \mathscr{S}(\mathbb{Z})$ and $\mathscr{D}'(\mathbb{T}) \cong \mathscr{E}(\mathbb{Z})$. Here $\mathbb{T}$ is the one-dimensional torus, $\mathscr{D}$ are smooth functions, $\mathscr{D}'$ is the space of distributions and $\mathscr{E}$ represents the sequence of polynomial growth. The above isomorphisms are done using Fourier transforms. The problem is this: *do we have* $(\mathscr{D}\hat{\otimes}\mathscr{D}')^* = \mathscr{D}'\hat{\otimes}\mathscr{D}$ *for the dual space*? Incidentally, the problems on TVS that we encountered in §12.13.4 are of the same nature: very specialised, 'screwy' and probably quite difficult.

Maybe some 'eager graduate student' of today will be able to give the answer to this and also see how this connects with some of the things we have done in this appendix (for example, homology of Γ with values in $\mathscr{E}(\Gamma)\ldots$).

G.6 Connections, Curvature and Cohomology

This is not 'variations' but an entirely different 'theme'. It is also the title of the three-volume work by Greub et al. In that work, a theory is developed that is based on the cohomology theory of Lie algebras. Presheaf cohomology is not used in that theory but the spectral sequences cannot be avoided. The 'Koszul complexes' are the *magic tool* that one uses here.

One of the important results of this theory (see Greub et al., 1976, §IX.2), the so-called *fundamental theorem of Chevalley*, can be specialised to the homogeneous spaces $G|K$ of §14.1.4 and one immediately obtains from this the proposition of that section. From this the whole program of Chapter 14 folds up at a stroke. So much so, that we can do this in half a page rather than a chapter! (See the exercise at the end of §14.6.)

So why did we not adopt this approach rather than the presheaf cohomology of Chapter 14?

As a matter of fact, this approach was originally adopted in a first attempt to present the program of Chapter 14. But then we felt that we had to help the reader navigate the vast technical material of Greub et al. (1976) and in particular we had to explain why the Chevalley theorem *does* specialise to give §14.1.4.

This turned out to be as long and elaborate and incomparably less satisfactory than what we did.

To finish, let me throw in a few more *magic words*: 'Serre's thesis', 'cubical singular homology', 'fibre spaces in the sense of Serre', etc. To the experts, these describe a circle of ideas that can be used to prove the results of §14.1.4 without the use of presheaf cohomology. The price one pays for this alternative approach is, on the other hand, quite high. Nonetheless, this approach, which does not avoid the use of spectral sequences, can be used here also. More precisely, after quite a bit of hard work one can adapt this approach and what one obtains is an equivalent version of our theorem expressed in terms of currents of superpolynomial decay of §12.7.

Epilogue

Part III of the book was to a large extent devoted to the proof of the homological B–NB classification (see §§1.7, 1.8). This was quite involved and uses deep basic tools from geometry and topology together with ad hoc constructions and ideas.

As I explained in Chapter 1, this homological classification was not viewed in the book as an aim itself. It was rather an intermediate link in the global B–NB classification of Lie groups.

With hindsight, I now feel that this homological classification deserves a better, cleaner and more systematic treatment in its own right than the one I gave in the book. It would take a new program to achieve this and it would require a separate examination of each of the $H_0, H_1, \dots$, the homology groups of polynomial growth of dimension $0, 1, \dots$. Depending on the linear geometry of the roots, one would want to decide, for each of these, whether it is of finite or infinite dimension. We would thus be refining the classification of §2.2.

I did not consider this problem until the writing of the book was finished and even then my main concern was to find a way to bypass the geometric constructions of Chapters 9 and 10 which in places become very 'hairy'. Had it not been for this I do not think I would have pursued this problem for almost a year, since in addition this proved to be an arduous and 'screwy' (to say the least) game to play and in an area that is totally ad hoc and has nothing to do with the rest of the book.

This epilogue is not the place to tell readers what I now know on this problem, other than 'yes, with different methods we can eliminate most of the difficult geometric constructions of Chapters 9 and 10.' In these new methods we use only calculus and algebraic manipulations on differential forms; that is, things that are more digestible to people with an analytical background. However, these manipulations are not straightforward, and in fact I now think that it would be quite difficult to finalise this project. To the futile question of whether I would have written Part III of the book differently had I known about these methods earlier, I can only give yes as an answer. The reason is that the geometric constructions of Chapters 9 and 10 are replaced by different constructions that may be more subtle but are considerably easier and simpler to describe. I try not to regret this fact too much and say to myself that in any mathematical theory, important simplifications are bound to happen with time.

Having said all this, in the next few lines I shall give readers a taste of the things that happen here. For simplicity I shall stick to the special semiproduct group $\mathbb{R}^n \leftthreetimes \mathbb{R} = G$ (see §2.3.2). The Euclidean coordinates of $\mathbb{R}^n$ are denoted by $(x_1, \ldots, x_n)$ and $y \in \mathbb{R}$ acts by $(e^{L_1 y}x_1, \ldots, e^{L_n y}x_n)$, where $L_1, \ldots, L_n$, the roots, are now real numbers. In what follows we can split off the zero roots so that $G = G_1 \times \mathbb{R}^{n-k}$, where G_1 is similar, with $L_1, \ldots, L_k$ all non-zero. On the other hand, since G_1 is a polynomial retract of G (see §12.2) our problem reduces to G_1. (Incidentally, there is scope here to use Grothendieck's version of Künneth's formula; see Appendix G.5, but let us not get into this.)

The most convenient complex to consider here is that of currents ω such that the coefficients of both ω and $\mathrm{d}\omega$ are measures (these are called *normal* currents; see §10.2) and such that the mass decay of both ω and $\mathrm{d}\omega$ are superpolynomial at infinity (see §12.7, especially (12.52)–(12.54), for a formal definition). Then H_p is the pth homology of this simplex. Here p indicates the *dimension* of the currents that we consider, so that when, for instance, ω is represented by integration on a cell, then p is the dimension of the cell. We have of course $H_0 = \mathbb{R}$ (one can use the ideas of Exercises (i) and (iii) in §13A.3 to see this) and $H_{n+1} = 0$ (since $\mathrm{Ker}\,\mathrm{d} = 0$ in that dimension) and, as we said, all the roots $L_1, \ldots, L_n$ are now assumed to be non-zero.

In this simple situation our question admits a very natural and satisfactory answer, and we have the following criterion. Surprisingly, even in this case, which is as simple as it gets, the proofs are not easy.

Criterion *Let $p = 1, 2, \ldots, n$. Then $H_p = 0$ if and only if all partial sums of p distinct roots, $L_{i_1} + \cdots + L_{i_p}$ are all $\geqslant 0$ or all $\leqslant 0$.*

Clearly the change of direction $y \mapsto -y$ makes these two formulations equivalent. We could then say that the roots satisfy the NC_p-condition. Conversely, in the C_p case, where there exist two such partial sums, $L_{i_1} + \cdots + L_{i_p} > 0$ and $L_{j_1} + \cdots + L_{j_p} < 0$, we have $\dim H_p = \infty$. To convince yourself that the proof of this cannot be easy, observe that it implies that if $H_p = 0$ then so is H_{p+1}. Try to prove this directly! From the nature of the criterion, one also sees that it is essential to separate the zero roots before we even start thinking of the problem.

For anyone who tries to prove the above criterion for themselves it is worth pointing out that the *non-degenerate* case, where we assume in addition that all partial sums $L_{\alpha_1} + \cdots + L_{\alpha_p} \neq 0$ (distinct summands) do not vanish, is much easier to handle than the general case. This is one of the many strange things that one comes across in this area and it illustrates the complexity of the problem, and this already in a very special class of 'rank 1' groups. In fact, the difficulty I had in proving this criterion left me with no doubt that any general theorem in this project is bound to be quite challenging even when it comes to making a 'reasonable' guess.

Whether I shall pursue this program to its 'bitter end', let alone be successful, I do not know. At this point it is not even clear to me that I shall ever get round to writing down in full detail what I know. But what I have said in this epilogue may inspire other people to pursue this matter further.

References

Alexopoulos, G. (1992). An application of homogenization theory to harmonic analysis: Harnack inequalities and Riesz transforms to groups of polynomial growth. *Canad. J. Math.*, **44**(4), 691–727.

Borel, A. and Tits, J. (1965). Groupes reductifs. *Publ. Math. Inst. Hautes Etudes Sci.*, **27**, 55–150.

Bott, R. and Tu, L. W. (1982). *Differential Forms in Algebraic Topology*, Springer.

Bougerol, P. (1981). Theorem central limit local sur certain groupes de Lie. *Ann. Sc. E.N.S., série 4*, **14**, 403–32.

Bourbaki, N. (1953). *Espaces Vectoriels Topologiques, Livre V*, Hermann.

Bourbaki, N. (1963). *Eléments de Mathématiques, Livre VI, Integration*, Hermann.

Bourbaki, N. (1972). *Groupes et Algèbres de Lie*, Hermann.

Brown, K. S. (1982). *Cohomology of Groups*, Springer.

Bruhat, F. and Tits, J. (1972). Groupes réductifs sur un corps local. *Publ. Math. Inst. Hautes Etudes Sci.*, **41**, 5–252.

Cartan, H. (1948). *Séminaire ENS, 1948–1963*, Tomes 1–16. Faculté de Sciences, Paris.

Cartan, H. and Eilenberg, S. (1956). *Homological Algebra*, Princeton University Press.

Cassels, J. W. S. (1986). *Local Fields*, LMS Student Texts **3**, Cambridge University Press.

Cheeger, J. and Ebin, D. G. (1975). *Comparison Theorems in Riemannian Geometry*, North-Holland.

Chevalley, C. (1951). *Théorie de Groupes de Lie, Tome II*, Hermann.

Chevalley, C. (1955). *Théorie de Groupes de Lie, Tome III*, Hermann.

Chung, K. L. (1982). *Lectures from Markov Processes to Brownian Motion*, Springer.

Coifman, R. and Weiss, G. (1977). *Transference Methods in Analysis*, Regional Conference Series in Mathematics **31**, American Mathematical Society.

de Rham, G. (1960). *Variétés Differentiables*, Hermann.

Dixmier, J. (1957). L'application exponential dans les groupes de Lie resolubles. *Bill. Soc. Math. France*, **85**, 113–21.

Dubrovin, B. A., Fomenko, A. T. and Novikov, S. P. (1990). *Modern Geometry – Methods and Applications. Part III* , Springer.

Dunford, N. and Schwartz, J. T. (1958). *Linear Operators, Part I*, Wiley-Interscience.

Eilenberg, S. and Steenrod, N. (1952). *Foundations of Algebraic Topology*, Princeton University Press.

Epstein, D. B. A., Cannon, J. W., Holt, D. F., Paterson, M. S. and Thurston, W. P. (1992). *Word Processing in Groups*, A. K. Peters/CRC Press.

Federer, H. (1969). *Geometric Measure Theory*, Springer.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vols I, II*, Wiley.

Gangoli, R. and Varadarajan, V. S. (1980). *Harmonic Analysis of Spherical Functions on Real Reductive Groups*, Springer.

Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*, Addison–Wesley.

Godement, R. (1958). *Topologie Algébrique et Théorie de Faisceaux*, Hermann.

Greene, R. E. and Wu, H. (1979). *Function Theory on Manifolds Which Possess a Pole*, Lecture Notes in Mathematics **699**, Springer.

Greenleaf, F. (1969). *Invariant Means on Topological Groups and Their Applications*, Van Nostrand.

Greub, W., Halperin, S. and Vanstone, R. (1973). *Connections, Curvature and Cohomology, Volume II*, Academic Press.

Greub, W., Halperin, S. and Vanstone, R. (1976). *Connections, Curvature and Cohomology, Volume III*, Academic Press.

Gromov, M. (1981). Groups, polynomial growth and expanding maps. *Publ. Math. Inst. Hautes Etudes Sci.*, **53**, 53–78.

Gromov, M. (1991). Asymptotic invariants of infinite groups. In G. Niblo and M. Roller, eds., *Geometric Group Theory, 2*, LMS Lecture Notes Series **182**, Cambridge University Press.

Grothendieck, A. (1952). Résumedés resultats essentiels dans la théorie des produits tensoriels topologiques et des espaces nucléaires. *Ann. l'Inst. Fourier*, **4**, 73–112.

Grothendieck, A. (1958). *Espaces Vectoriels Topologiques*. Publicação da Sociedade de Matemática de Sao Paulo.

Grünbaum, B. (1967). *Convex Polytopes*. Springer.

Guivarc'h, Y. (1973). Croissance polynômiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, **101**, 333–79.

Hall, M. (1959). *The Theory of Groups*. Chelsea.

Hebisch, W. (1992). Estimates on the semigroup generated by left-invariant operators on Lie groups. *J. reine angew. Math.*, **423**, 1–45.

Helgason, S. (1978). *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press.

Helgason, S. (1984). *Groups and Geometric Analysis*, Academic Press.

Hicks, N. J. (1971). *Notes on Differential Geometry*, Van Nostrand.

Hilton, P. J. (1953). *Introduction to Homotopy Theory*, Cambridge University Press.

Hilton, P. J. and Wylie, S. (1960). *Homology Theory*, Cambridge University Press.

Hirsch, M. W. (1976). *Differential Topology*, Springer.

Hochschild, G. (1965). *The Structure of Lie Groups*, Holden–Day.

Hörmander, L. (1967). Hypoelliptic second order differential equations. *Acta Math.*, **119**, 147–71.

Hörmander, L. (1983). *The Analysis of Linear Partial Differential Operators, Volume I*, Springer.

Hörmander, L. (1985). *The Analysis of Linear Partial Differential Operators, Volume III*, Springer.

Humphreys, J. E. (1975). *Linear Algebraic Groups*, Springer.

Jacobson, N. (1962). *Lie Algebras*, Wiley–Interscience.

Jacobson, N. (1989). *Basic Algebra*, 2nd ed., vols. 1 and 2, Freeman.

Jarchow, H. (1981). *Locally Convex Spaces*, Teubner.

Jenkins, J. (1973). Growth of connected locally compact groups. *J. Funct. Anal.*, **12**, 113–27.

Jikov, V. V., Kozlov, S. M. and Oleinik, O. A. (1991). *Homogenization of Differential Operators and Integral Functionals*, Springer.

Katznelson, Y. (1968). *Introduction to Harmonic Analysis*, Wiley. (3rd ed. (2004), Cambridge University Press).

Kawada, Y. and Ito, K. (1940). On the probability distribution on a compact group. *Proc. Phys. Soc.*, **22**, 977–99.

Knapp, A. W. (1986). *Representation Theory of Semisimple Groups. An Overview Based on Examples*, Princeton University Press.

Kobayashi, S. and Nomizu, K. (1963). *Foundations of Differential Geometry, Volume 1*, Wiley–Interscience.

Kobayashi, S. and Nomizu, K. (1969). *Foundations of Differential Geometry, Volume 2*. Wiley–Interscience.

Magnus, W., Karass, A. and Solitas, D. (1965). *Combinatorial Group Theory*, Dover.

Massey, W. S. (1991). *A Basic Course in Algebraic Topology*, Springer.

McCleary, J. (2001). *User's Guide to Spectral Sequences*, 2nd ed., Cambridge University Press.

Montgomery, D. and Zippin, L. (1955). *Topological Transformation Groups*, Wiley–Interscience.

Naimark, M. A. (1959). *Normed Rings*, Noordhoff.

Onischik, A. L. and Vinberg, E. B. (1988). *Lie Groups and Algebraic Groups*, Springer.

Paterson, A. T. (1988). *Amenability*, AMS Mathematical Surveys and Monographs **29**, American Mathematical Society.

Pier, J.-P. (1984). *Amenable Locally Compact Groups*, J. Wiley.

Pontrjagin, L. (1939). *Topological Groups*, Princeton University Press.

Ragunathan, M. S. (1972). *Discrete Subgroups of Lie Groups*, Springer.

Reiter, H. (1968). *Classical Harmonic Analysis and Locally Compact Groups*, Oxford University Press.

Sagle, A. A. and Walde, R. E. (1973). *Introduction to Lie Groups and Lie Algebras*, Academic Press.

Schaefer, H. H. (1974). *Banach Lattices and Positive Operators*, Springer.

Schwartz, L. (1953). *Séminaire Schwartz, 1953–1954*. Produits tensoriels topologiques d'espaces vectoriels topologiques, Espaces vectoriels topologiques nucléaires. Applications. Faculté des Sciences de Paris.

Schwartz, L. (1957). *Théorie des Distributions, Tome 1*, Hermann.

Serre, J.-P. (1965). *Lie Algebras and Lie Groups*, Benjamin.

Serre, J.-P. (1970). Cohomologie de groupes discrets. *Séminaire Bourbaki* #399, 1970/71, pp. 337–50.

Steenrod, N. (1951). *The Topology of Fiber Bundles*, Princeton University Press.

Szegö, G. (1939). *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ., XXIII.

Varadarajan, V. S. (1974). *Lie Groups, Lie Algebras and Their Representations*, Prentice-Hall.

Varopoulos, N. Th. (1990). Small time Gaussian estimates of heat diffusion kernels, II. The theory of large deviations. *J. Funct. Anal.*, **93**, 1–33.

Varopoulos, N. Th. (1994a). Diffusion on Lie groups. *Canad. J. Math.*, **46**, 438–46.

Varopoulos, N. Th. (1994b). Diffusion on Lie groups, II. *Canad. J. Math.*, **46**, 1073–93.

Varopoulos, N. Th. (1996a). The heat kernel on Lie groups. *Rev. Mat. Iberoamericana*, **12**, 147–86.

Varopoulos, N. Th. (1996b). Analysis on Lie groups. *Rev. Mat. Iberoamericana*, **12**, 791–917.

Varopoulos, N. Th. (1999a). Distance distorsion on Lie groups. In M. A. Picardello and W. Woess, eds., *Random Walks and Discrete Potential Theory, Proceedings, Cortona (1997)*, Symposia Mathematica **39**, Cambridge University Press.

Varopoulos, N. Th. (1999b). Diffusion on Lie groups III. *Canad. J. Math.*, **48**, 641–72.

Varopoulos, N. Th. (1999c). Potential theory in conical domains. *Math. Proc. Camb. Phil. Soc.*, **125**, 335–84.

Varopoulos, N. Th. (2000a). Geometric and potential-theoretic results on Lie groups. *Canad. J. Math.*, **52**(2), 412–37.

Varopoulos, N. Th. (2000b). A geometric classification of Lie groups. *Rev. Mat. Iberoamericana*, **16**, 49–136.

Varopoulos, N. Th. (2001). Potential theory in Lipschitz domains. *Canad. J. Math.*, **53**(5), 1057–120.

Varopoulos, N. Th. (2014). The central limit theorem in Lipschitz domains. *Boll. Unione Math. Ital.*, **7**, 103–56.

Varopoulos, N. Th., Saloff-Coste, L. and Coulhon, T. (1992). *Analysis and Geometry on Groups*, Cambridge University Press.

Warner, F. W. (1971). *Foundations of Differentiable Manifolds and Lie Groups*, Scott, Foresman and Co. Reprinted 1983, Springer.

Warner, G. (1970). *Harmonic Analysis on Semisimple Lie Groups, Vol. I*, Springer.

Weil, A. (1953). *L'Integration dans les Groupes Topologiques et Applications*, Hermann.

Weil, A. (1995). *Basic Number Theory*, Springer.

Whitney, H. (1958). Elementary structure of real algebraic varieties. *Annals Math.*, **66**(3), 545–56.

Williams, D. (1991). *Probability with Martingales*, Cambridge University Press.

Woess, W. (2000). *Random Walks on Infinite Graphs and Groups*, Cambridge Tracts in Mathematics **138**, Cambridge University Press.

Yosida, K. (1970). *Functional Analysis*, 5th ed. Springer.

Index