
S E
M M
E S

Henryk Woźniakowski

Erich Novak
Ian H. Sloan
Joseph F. Traub
Henryk Woźniakowski

Essays on the Complexity of Continuous Problems

European Mathematical Society

Authors:

Erich Novak
Friedrich-Schiller-Universität Jena
Mathematisches Institut
07743 Jena
Germany

Joseph F. Traub
Department of Computer Science
Columbia University
1214 Amsterdam Avenue
New York, NY 10027-7003
USA

Ian H. Sloan
School of Mathematics and Statistics
University of New South Wales
Sydney, NSW, 2052
Australia

Henryk Woźniakowski
Department of Computer Science
Columbia University
1214 Amsterdam Avenue
New York, NY 10027-7003
USA

and

Institute of Applied Mathematics and Mechanics
Warsaw University
Banacha 2
02-097 Warszawa
Poland

ISBN 978-3-03719-069-2

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at http://www.helveticat.ch.

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the authors' TEX files: I. Zimmermann, Freiburg
Printed in Germany

9 8 7 6 5 4 3 2 1

Preface

This book contains five contributions on the complexity of continuous problems and, at the same time, is a Festschrift for Henryk Woźniakowski.

A version of these contributions (except for the last one) were presented on June 6, 2008, at the Friedrich Schiller University in Jena, when Henryk was presented with an honorary doctoral degree by the University of Jena. It was for us a very happy occasion, and it was an occasion for talks addressed to a wider audience. We believe that these talks (and now papers) are a good introduction and supplement to more technical research papers and books.

We briefly describe the contents:

- The account "Henryk Woźniakowski and the complexity of continuous problems" by EN describes some of the achievements of Henryk, and it contains remarks about the history of information-based complexity (IBC). It also contains the reports from Mathematical Reviews about the two main monographs on IBC, the two black books that appeared 1980 and 1988, respectively. As an appendix, we list all publications of Henryk.
- The essay "Complexity as a new challenge for mathematicians" by HW discusses the computational complexity of three problems: matrix multiplication, multivariate integration of smooth periodic function, and multivariate approximation of smooth functions. The first two problems are studied in the worst case setting, whereas the third problem is studied in the average case setting with the folded Wiener sheet measure. These three problems serve as an illustration that computational complexity presents a new set of questions, whose answers often require new proof techniques. That is why even well studied mathematical problems need to be revisited when we want to find sharp bounds on their complexity.
- The section "A brief history of information-based complexity" by JFT relates how Henryk Woźniakowski first came to Carnegie Mellon University in 1973. It then flashes back to precursors of IBC as well as the beginning of optimal iteration theory. The rest of the essay is devoted to the history of IBC, from the early 70s and then follows IBC to the present.

- The section "How high is high-dimensional?" by IHS describes the fascinating history of high-dimensional integration. This field changed dramatically over the last 15 years, with much of that change driven by Henryk's persistent question: 'What happens to the error as the dimension goes to infinity?'.
- The purpose of the last essay "What is information-based complexity?" by HW is to introduce information-based complexity in an informal way. The basic notions of IBC for the approximate solution of continuous mathematically posed problems are described there.

We thank the Friedrich Schiller University and the Gesellschaft der Freunde und Förderer der FSU Jena for their support. Joseph F. Traub and Henryk Woźniakowski wish to thank the National Science Foundation for supporting their work. Ian Sloan acknowledges the support of the Australian Research Council under its Centres of Excellence program.

We thank Henryk for many suggestions, and even more importantly for a long and deep friendship.

Contents

Henryk Woźniakowski and the complexity of continuous problems

Erich Novak

Introduction

Henryk Woźniakowski is a fascinating colleague and friend. This short paper can only describe a very small part of what Henryk has done.

Henryk was born on August 31, 1946, in Lublin, Poland. Lublin is about 150 kilometers south east of Warsaw and today has about 350 000 inhabitants. His family moved to Warsaw in 1950. Henryk studied mathematics and computer science at the University of Warsaw and got his diploma in 1969, Ph.D. in 1972 and habilitation in 1976. From 1972 to 1977, he was assistant professor; in 1977 he became an associate professor in Warsaw.

In 1981, Henryk was elected chairman of the Department of Mathematics, Computer Science and Mechanics. He was running as a Solidarity candidate. In the same year, the Senate of the University of Warsaw decided that Henryk should become full professor. However, he had to wait till 1988 that this decision became a reality because of political reasons – Henryk Woźniakowski was for many years one of the leaders of the Solidarność movement at the University of Warsaw. In 1989, after political changes in Poland, Henryk was elected chairman of Solidarity at the University of Warsaw and served two years. Even before, in 1984, Henryk got a position as a full professor at Columbia University in New York. Since then, he has been teaching in both Warsaw and at Columbia.

Henryk received many prizes, such as the Stanisław Mazur prize of the Polish Mathematics Society in 1988. He had long stays in Berkeley (MSRI and ICSI), at the MGU in Moscow, at Carnegie Mellon University, and at the University of New South Wales in Sydney. In 2005 Henryk was awarded the Humboldt Research Award and visited from November 2006 till July 2007 the University of Jena. Henryk is a member of the Polish Academy of Sciences.

The University of Jena (FSU) is quite picky with respect to honorary doctoral degrees. Although the FSU celebrated its 450th anniversary in 2008, only three

colleagues have received an honorary degree because of their work in mathematics:

- Erna Weber, 1897–1988, for her work in statistics,
- Aleksander Pełczyński, born 1932, for his work in functional analysis,
- Boris Trachtenbrot, born 1921, for his work in theoretical computer science.

After Aleksander Pełczyński, Henryk is the second mathematician from Warsaw who is honored by the University of Jena. This is certainly a very good proof of the high quality of Polish mathematics, as well as the good relations between the Universities in Warsaw and Jena. I cite from the diploma:

> In Anerkennung seiner grundlegenden Arbeiten zur Numerischen Mathematik. Besonders hervorgehoben seien die tiefen Einsichten durch die neue Disziplin "Information-Based Complexity" und die Arbeiten zum Fluch der Dimension, mit deren Hilfe man erstmals versteht, welche hochdimensionalen Probleme lösbar sind.[1]

1 Early work

Henryk is an excellent mathematician with a great creative urge and power. His first paper [36] appeared in 1969. Henryk was 23 years old and got his diploma in the same year. During the next ten years Henryk published many papers about the numerical solution of linear and nonlinear equations.

Several of these papers, as well as his Ph.D. dissertation, deal with the maximal order of methods for the solution of nonlinear equations. In particular, Woźniakowski proved a conjecture of Traub and Kung concerning the maximal order of multi-point iterations without memory, see Woźniakowski [37], [38], [39], Traub and Woźniakowski [26], and the paper by Joseph Traub in this book.

Several of his papers about the numerical stability for solving equations appeared in Numerische Mathematik and in BIT in the years 1977 and 1978. These papers [7], [40], [41], [42] are still cited quite often.

[1] Translation: In recognition of his fundamental work in numerical mathematics. We emphasize in particular the deep insights by the new discipline "Information-Based Complexity" and the work on the curse of dimensionality. With this work we understand for the first time which high-dimensional problems are tractable.

2 A general theory of optimal algorithms

Henryk was decisively involved with the creation of two big theories – the second one is his own child, other colleagues collaborating only later.

Together with Joseph Traub, Henryk built a complexity theory for continuous problems around 1977. The discrete world of the Turing machine is too narrow for many applications. We want to understand efficient algorithms for numerical integration, for the solution of differential equations, and for many other problems that involve real- or complex-valued functions on intervals or more complicated domains.

Of course there is a long tradition of studying algorithms for continuous problems and some of the algorithms even bear the name of their inventors, such as Newton's method, Gaussian quadrature formula, or Lagrange interpolation. More recent algorithms include the Metropolis algorithm or the Jenkins–Traub algorithm.

Also the complexity, i.e., the cost of optimal algorithms, was studied, sometimes for a restricted class of algorithms and always only for a specific problem.

Hence there existed something what could be called "pieces of the puzzle" and it is fair to mention many fathers of the theory of optimal algorithms for continuous problems, such as Babushka, Bakhvalov, Kolmogorov, Nikolskij, Smolyak and Sobolev in the east and Golomb, Kiefer, Sard and Weinberger in the west. This list is certainly far from complete!

But the flow of information between the west and the east was sometimes slow and, more important, there only existed somewhat isolated results for specific problems. These results became a part of a comprehensive theory only later.

Hence the "first black book" by Traub and Woźniakowski, *A general theory of optimal algorithms*, was a sensation. It was published in 1980 by Academic Press. This book described for the first time a comprehensive theory for continuous problems. The book also contains, as Part C, a brief history of the field and a long annotated bibliography. There the reader may find all the references that are missing here.

Ko-Wei Lih writes in the Mathematical Reviews an excellent report that ends with "... the authors should be congratulated on their magnificent product which elevates the study of the approximate to a higher dimension." Actually, this report is still very informative and this is why it is reprinted here in full length:

> This monograph is a report on work in progress in the theory of analytic computational complexity which is the study of optimal algorithms for prob-

lems solved approximately. Such a line of investigation had its inception around 1950 with the work of Kiefer, Sard, and Nikolskij on optimal algorithms for locating the maximum, for integration, for approximation, etc. This stream of research generated mainly results concerning specific problems. In 1961 Traub initiated a second stream of research with the study of solutions to nonlinear equations by iterative methods. The possibility of unification of these two streams into one general and necessarily more abstract framework was first shown by the authors in two long reports ["General theory of optimal error algorithms and analytic complexity", parts A and B; per bibl.]. This monograph includes extended and improved material from these two reports. A central concern of the computer scientist is the selection of the best algorithm for solving a problem. However, selection of the best is subject to multivariate criteria such as time and space complexity, ease of implementation, robustness, and stability. The authors only deal with time complexity here. Nevertheless, conclusions could be easily adjusted to work for space complexity. Also, the authors study only problems which cannot be solved exactly with finite complexity or problems which one chooses to solve approximately for reasons of efficiency. The final theory includes algebraic complexity as a special case. The generality and simplicity achieved by this theory has its cornerstone on the notion of information operator. Adversary arguments based on the information used by an algorithm lead to lower bound theorems. This has its practical application in the rationalization of the synthesis of algorithms. Traditional ad hoc algorithms are revealed to be paying high penalty without the use of optimal information. The authors propose 20 general questions to be studied. The following is a sample of some of them. 1. What is a lower bound on the error of any algorithm for solving a problem using given information? 2. In general is there an algorithm which gets arbitrarily close to this lower bound? 3. What is the optimal information for solving a problem? 4. Given a specific problem, how do we characterize and construct an optimal algorithm for its solution? 5. Can it be established that one problem is intrinsically harder than another? 6. Compare the power of adaptive and non-adaptive algorithms. 7. Compare the power of linear and nonlinear information operators. 8. What is the class of all problems which can be solved by iteration using linear information?

This monograph is divided into three parts. Part A has ten chapters and deals with a general information model. The basic concepts are first formalized.

The notions of optimal error algorithm and optimal complexity algorithms are introduced. Then a large portion is devoted to the study of linear problems using linear information. It is shown that adaptive information is not more powerful than non-adaptive information for a linear problem. A linear problem is also constructed to possess no linear optimal error algorithms. However, natural problems are immune from such pathology. Algorithms optimal in the sense of Sard and Nikolskij are shown to be optimal error algorithms. There exist linear problems with essentially arbitrary complexity. So there are no "gaps" in the complexity function.

In Chapter 6, the theory is applied to the solution of many different linear problems including approximation, interpolation, integration, and the solution of linear partial differential equations. Finally, the theory of nonlinear information is developed and applications given. In the general information setting the class of nonlinear information operators is actually too powerful to be of interest. In the last two chapters, a partial hierarchy of complexity is presented and other models of computation are briefly discussed. Part B consists of one chapter with 11 sections. It deals with the iterative information model and is built on some 20 years of research on iterative complexity initiated by Traub. The deepest question studied is: what problems can be solved by iteration using iterative linear information? For one-point stationary iterations using iterative linear information, it is shown that the class of iterative algorithms is empty for a problem unless the "index" of the problem is finite. A conjecture characterizing problems with finite index is posed to the effect that a positive solution implies that only nonlinear equations can be solved by iteration. Part C provides a brief history of the theory of analytic computational complexity and an annotated bibliography of over 300 papers and books covering both the eastern European and the Western literature. The authors supply numerous conjectures and open problems throughout the book. They also recommend eight tracks for various readers with particular interests such as researchers interested in open problems, researchers interested in the literature on history, theoretical computer scientists, mathematicians, numerical analysts, scientists and engineers. Some of these readers will definitely find that the study of this book is a quite strenuous task. However, the authors should be congratulated on their magnificent product which elevates the study of the approximate to a higher dimension. [MR0584446 (84m:680410)]

3 Information-based complexity

There are two more monographs, written jointly with Joseph Traub and Grzegorz Wasilkowski. The third monograph, *Information-based complexity*, is certainly a special highlight.

The first monograph did *not* discuss the average case setting and did *not* study randomized algorithms. These are two major new subjects of the "second black book", that appeared again with Academic Press, see [24]. Of course the book contains many more results, for example also a section on linear PDEs written by Arthur Werschulz. Again I cite the complete report from the Mathematical Reviews, written by M. I. Dekhtyar.

> There are two main branches of computational complexity theory. The first is combinatorial complexity, which considers problems for which the information is complete, exact, and free. The second, which deals with problems for which the information is partial, noisy, and priced and for which solutions are not exact, is called information-based complexity and is the subject of the book under review. The authors summarize and present a number of results that are concerned with various definitions of the cost and the error of algorithms. The book may be viewed as a continuation and extension of two previous books [Traub and Woźniakowski, A general theory of optimal algorithms, Academic Press, New York, 1980; the authors, Information, uncertainty, complexity, Addison-Wesley, Reading, MA, 1983].
>
> The book consists of twelve chapters and two appendices. Chapter 1 is an introduction. In Chapter 2, the basic concepts of information-based complexity are illustrated by the example of continuous binary search.
>
> In Chapter 3 an abstract formulation of an information-based theory is presented. A problem is defined as a solution operator $S\colon F \to G$, where F is a set and G is a normed linear space over the scalar field of real or complex numbers. Elements $f \in F$ are called problem elements, and the $S(f)$ are called solution elements. Computation of an approximation $U(f)$ of $S(f)$ consists of two steps. The first is to obtain information about f:
>
> $$N(f) = [L_1(f), L_2(f; y_1), \ldots, L_{n(f)}(f; y_1, \ldots, y_{n(f)-1})],$$
>
> where $y_i = L_i(f; y_1, \ldots, y_{i-1})$, and L_i is a permissible information operation. Information N is called non-adaptive if $y_i = L_i(f)$. The second step is to evaluate the approximation by $N(f) \mapsto U(f) = \varphi(N(f))$, where φ is a mapping (algorithm): $N(F) \to G$. Then the cost of comput-

ing $U(f)$ is given by $\text{cost}(U, f) = \text{cost}(N, f) + \text{cost}(\varphi, N(f))$. The main results presented in the book deal with the first item of this sum. Three definitions of the error $e(U)$ are considered: (i) the worst case setting: $e(U) = \{\sup \|S(f) - U(f)\| : f \in F\}$; (ii) the average case setting: $e(U) = (\int_F \|S(f) - U(f)\| \mu(dF))^{1/2}$; (iii) the probabilistic setting: let $\delta \in [0, 1]$; then $e(U) = \inf\{\{\sup \|S(f) - U(f)\| : f \in F - A\} : \mu(A) \le \delta\}$. In Chapter 4 theoretical results for the worst case setting are presented. The radius of information is introduced; it is a sharp lower bound on the error of any algorithm using this information. The minimal cardinality of information with radius at most ε is denoted by $m(\varepsilon)$. If c is the cost of one information operation then $c \cdot m(\varepsilon)$ is a lower bound on the ε-complexity. Conditions under which this bound is almost sharp are investigated. Special attention is paid to the class of linear problems. It is shown that the use of adaptive information does not decrease ε-complexity for this class. Chapter 5 contains examples of approximation problems to which the results of Chapter 4 are applied to obtain complexity bounds and optimal algorithms. They include integration, function approximation, optimization, etc. For most of them only short sketches of the results are presented and the authors direct the reader to the references cited for detailed analysis.

Chapters 6 and 7 deal with the average case setting. A Gaussian measure is proposed as the probability measure on the set F. The average radius of information and the average minimal cardinality of error $m^{\text{avg}}(\varepsilon)$ are introduced. As in Chapter 4, $c \cdot m^{\text{avg}}(\varepsilon)$ is a lower bound on the average ε-complexity. It is shown that this bound is tight for linear problems. In Chapter 7 three linear problems are considered and bounds on their average ε-complexity are established.

The probabilistic setting is considered in Chapter 8. The probabilistic radius of information and the probabilistic (ε, δ)-cardinality number $m^{\text{prob}}(\varepsilon, \delta)$ are defined in such a way that $c \cdot m^{\text{prob}}(\varepsilon, \delta)$ is a lower bound on probabilistic complexity. Complexity of linear problems is analyzed. The probabilistic complexity is compared with the average complexity and some relations are developed.

Chapter 9 contains a comparison between different settings for four problems: the integration of smooth functions, the integration of smooth periodic functions, the approximation of smooth periodic functions, and the approximation of smooth non-periodic functions. For each problem a 5×3 table is presented with formulas for the complexity under five error criteria (ab-

solute error, normalized error, and three kinds of relative errors) in three settings.

In Chapter 10 the asymptotic setting is studied. Two approaches to optimal asymptotic algorithms are considered. Under one of them the best speed of convergence is achieved by algorithms which are optimal in the worst case setting. The other approach leads to a close relation between the asymptotic and average case settings.

The main question investigated in Chapter 11 is the extent to which randomization can lower the worst and the average case complexities. The results presented here show that randomization does not help significantly for linear problems.[2] Some results concerning noisy information are given in Chapter 12. Noisy information about $f \in F$ has the form

$$N(f, \bar{x}) = [L_1(f) + x_1, \ldots, L_n(f; z_1, \ldots, z_{n-1}) + x_n],$$

where $\bar{x} = [x_1, \ldots, x_n]$ is noise, $n = n(f, \bar{x})$, and $z_i = L_i(f; z_1, \ldots, z_{i-1}) + x_i$ is the ith observed piece of information. The relationship between adaptive information and non-adaptive information is discussed. These areas are open for further investigation.

Two appendices contain the main definitions and facts concerning functional analysis and measure theory. The extensive bibliography includes more than 450 items. Almost all chapters and sections are followed by notes and remarks that contain some additional results, comments and references. The book is clearly written and may be used as a handbook by specialists in information-based complexity; it may also be recommended as a textbook for those who want to study this area of computer science.

[MR0958691 (90f:68085)]

4 Tractability of multivariate problems

Many results in numerical analysis and approximation theory concern the *optimal order of convergence* for a problem and a class of functions. Also many results in the two black books GTOA and IBC deal with this subject. There is a widespread

[2] Remark of the author: It is true that randomization does not help for *some* linear problems. There are other problems, however, where randomization helps a lot; the most popular is the problem of numerical integration.

belief that we understand the complexity of a problem if we know the optimal order. This belief is wrong; it was Henryk who first studied the following problem seriously since about 1992:

Which multivariate problems can be efficiently solved in high dimensions?

First we formalize this question according to [45], [46]. Assume that we want to solve a problem S_d for functions $f \in F_d$, where F_d is a class of d-variate functions. An example would be the computation of

$$S_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x, \tag{1}$$

for $f \in F_d$, up to an error ε. Assume that the cost of an optimal algorithm for this problem is $n(\varepsilon, d)$. If

$$n(\varepsilon, d) \le C \cdot \varepsilon^{-\alpha} d^{\beta} \tag{2}$$

for certain $C, \alpha, \beta > 0$ then the problem is called (*polynomially*) *tractable*. Observe that C is independent of d in the definition (2) of tractability. Therefore the optimal order of convergence does not say much about tractability. We give three examples. Only the first example uses known results about the order of convergence to decide the tractability problem.

4.1 Integration of C^k-functions

There is a basic result of Bakhvalov from the year 1959 that the optimal order of convergence for the computation of the integral (1) for C^k-functions is $n^{-k/d}$. We conclude that the problem is *not* (polynomially) tractable. Roughly speaking, the cost is exponential in d, this is called the *curse of dimensionality*, after Bellman.

4.2 Integration of smooth periodic functions

We now consider the Korobov space $F_{d,\alpha}$ of complex functions from $L_1([0,1]^d)$, where $\alpha > 1$. This class is defined by controlling the sizes of Fourier coefficients of functions. More precisely, for $h = [h_1, h_2, \ldots, h_d]$ with integers h_j, consider the Fourier coefficients

$$\hat{f}(h) = \int_{[0,1]^d} f(x) e^{-2\pi \mathrm{i} h \cdot x}\,\mathrm{d}x,$$

where $\mathrm{i} = \sqrt{-1}$ and $h \cdot x = \sum_{j=1}^{d} h_j x_j$. Denote $\bar{h}_j = \max(1, |h_j|)$. Then

$$F_{d,\alpha} = \left\{ f \in L_1([0,1]^d) \mid |\hat{f}(h)| \le \left(\bar{h}_1 \bar{h}_2 \dots \bar{h}_d\right)^{-\alpha} \text{ for all } h \in \mathbb{Z}^d \right\}.$$

Again we consider the integration problem

$$S_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x \quad \text{for } f \in F_{d,\alpha}.$$

We consider algorithms A_n that use n function values, the worst case error $e^{\mathrm{wor}}(A_n)$, the nth minimal worst case error $e(n,d)$, and the minimal number

$$n^{\mathrm{wor}}(\varepsilon, S_d, F_{d,\alpha})$$

of function values needed to approximate the integrals to within ε.

The integration problem for the Korobov class $F_{d,\alpha}$ has been studied in a number of papers and books. It is known that

$$e(n,d) = \mathcal{O}\left(n^{-p}\right) \quad \text{as } n \to \infty, \text{ for all } p < \alpha.$$

For $p = \alpha$ we have

$$e(n,d) = \mathcal{O}\left(n^{-\alpha} (\ln n)^{\beta(d,\alpha)}\right)$$

where $\beta(d,\alpha)$ is of order d. Such errors can be obtained by lattice rules of rank 1, i.e., by algorithms of the form

$$A_n(f) = \frac{1}{n} \sum_{j=0}^{n-1} f\left(\left\{ j\,\frac{z}{n} \right\}\right),$$

where n is prime and $z \in \{1, 2, \dots, n-1\}^d$ is a well-chosen integer vector. Here, $\{x\}$ denotes the vector whose jth component is the fractional part of x_j.

Hence, for large α, the optimal order of convergence is also large and roughly equal to α independently of d. This is encouraging, but what can we say about tractability?

The tractability of this integration problem was studied by Sloan and Woźniakowski in [20], where it was proved that

$$e(n,d) = 1 \quad \text{for } n = 0, 1, \dots, 2^d - 1,$$

which implies that

$$n^{\mathrm{wor}}(\varepsilon, S_d, F_{d,\alpha}) \ge 2^d \quad \text{for all } \varepsilon \in (0,1).$$

That is, even for arbitrarily large α, despite an excellent order of convergence, this integration problem is *not* tractable. More about this problem can be found in Henryk's paper in this booklet.

4.3 The star-discrepancy

After two negative examples the reader may have the impression that all "interesting" problems are intractable and the curse of dimension is "always" present. This would be a wrong impression, since there are many tractable problems; actually there are many multivariate problems that can be solved in very high dimension.

Discrepancy is a measure of the deviation from uniformity of a set of points. It is desirable that a set of n points be chosen so that the discrepancy is as small as possible. The notion of discrepancy appears in many fields of mathematics.

We begin with the definition of the star discrepancy. Let $x = [x_1, x_2 \ldots, x_d]$ be from $[0,1]^d$. By the box $[0,x)$ we mean the set $[0,x_1) \times [0,x_2) \times \cdots \times [0,x_d)$, whose (Lebesgue) measure is clearly $x_1 x_2 \ldots x_d$. For given points $t_1, t_2, \ldots, t_n \in [0,1]^d$, we approximate the volume of $[0,x)$ by the fraction of the points t_i that are in the box $[0,x)$. The error of such an approximation is called the *discrepancy function*, and is given by

$$\mathrm{disc}(x) = x_1 x_2 \ldots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i),$$

where $1_{[0,x)}$ is the indicator (characteristic) function, so that $1_{[0,x)}(t_i) = 1$ if $t_i \in [0,x)$ and $1_{[0,x)}(t_i) = 0$ otherwise.

The *star discrepancy* of the points $t_1, \ldots, t_n \in [0,1]^d$ is defined by the L_∞-norm of the discrepancy function disc

$$\mathrm{disc}(t_1, t_2, \ldots, t_n) = \sup_{x \in [0,1]^d} \left| x_1 x_2 \ldots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|.$$

The main problem associated with star discrepancy is that of finding points $t_1, t_2, \ldots, t_n$ that minimize disc, and to study how this minimum depends on d and n.

We now show that the star discrepancy is intimately related to multivariate integration. Let $W_1^{\mathbf{1}} := W_1^{(1,1,\ldots,1)}([0,1]^d)$ be the Sobolev space of functions defined on $[0,1]^d$ that are once differentiable in each variable and whose derivatives have finite L_1-norm. We consider first the subspace of functions that satisfy the

boundary conditions $f(x) = 0$ if at least one component of x is 1, and define the norm

$$\|f\|_{d,1}^* = \int_{[0,1]^d} \left| \frac{\partial^d}{\partial x} f(x) \right| \mathrm{d}x.$$

Here, $\partial x = \partial x_1 \partial x_2 \dots \partial x_d$.

That is, we consider the class

$$F_d^* = \left\{ f \in W_1^{\mathbf{1}} \;\middle|\; f(x) = 0 \text{ if } x_j = 1 \text{ for some } j \in [1,d], \text{ and } \|f\|_{d,1}^* \le 1 \right\}.$$

Consider the multivariate integration problem

$$S_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x \quad \text{for } f \in F_d^*.$$

We approximate $S_d(f)$ by *quasi-Monte Carlo* algorithms, which are of the form

$$Q_{d,n}(f) = \frac{1}{n} \sum_{j=1}^{n} f(t_j)$$

for some points $t_j \in [0,1]^d$. We stress that the points t_j are chosen non-adaptively and deterministically. The name "quasi-Monte Carlo" is widely used, since these algorithms are similar to the Monte Carlo algorithm which takes the same form but for which the points t_j are randomly chosen, usually as independent uniformly distributed points over $[0,1]^d$.

We also stress that we use especially simple coefficients n^{-1}. This means that if $f(t_1), f(t_2), \dots, f(t_n)$ are already computed then the computation of $Q_{d,n}(f)$ requires just $n-1$ additions and one division. Since the points $t_1, t_2, \dots, t_n$ are non-adaptive, $Q_{d,n} f$ can be evaluated very efficiently in parallel since each $f(t_j)$ can be computed on a different processor. Obviously, $Q_{d,n}$ integrates constant functions exactly, even though $1 \notin F_d^*$.

The quality of the algorithm $Q_{n,d}$ depends on the points t_j. There is a deep and beautiful theory about how the points t_j should be chosen. We add that quasi-Monte Carlo algorithms have been used very successfully for many applications, including mathematical finance applications, for d equal 360 or even larger.

We now recall Hlawka and Zaremba's identity, which states that for $f \in W_1^{\mathbf{1}}$, we have

$$S_d(f) - Q_{d,n}(f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\dots,d\}} (-1)^{|\mathfrak{u}|} \int_{[0,1]^{|\mathfrak{u}|}} \mathrm{disc}(x_{\mathfrak{u}}, 1) \frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f(x_{\mathfrak{u}}, 1) \mathrm{d}x_{\mathfrak{u}}.$$

Here, we use the following standard notation. For any subset $\mathfrak{u}$ of $\{1,2,\ldots,d\}$ and for any vector $x \in [0,1]^d$, we let $x_{\mathfrak{u}}$ denote the vector from $[0,1]^{|\mathfrak{u}|}$, where $|\mathfrak{u}|$ is the cardinality of $\mathfrak{u}$, whose components are those components of x whose indices are in $\mathfrak{u}$. For example, for $d = 5$ and $\mathfrak{u} = \{2,4,5\}$ we have $x_{\mathfrak{u}} = [x_2, x_4, x_5]$. Then $\partial x_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \partial x_j$ and $\mathrm{d}x_{\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \mathrm{d}x_j$. By $(x_{\mathfrak{u}}, 1)$ we mean the vector from $[0,1]^d$ with the same components as x for indices in $\mathfrak{u}$ and with the rest of components being replaced by 1. For our example, we have $(x_{\mathfrak{u}}, 1) = [1, x_2, 1, x_4, x_5]$. Note that

$$\mathrm{disc}(x_{\mathfrak{u}}, 1) = \prod_{k\in\mathfrak{u}} x_k - \frac{1}{n}\sum_{j=1}^{n} 1_{[0,x_{\mathfrak{u}})}\left((t_j)_{\mathfrak{u}}\right).$$

For $f \in F_d^*$, the boundary conditions imply that all terms in Hlawka and Zaremba identity vanish except the term for $\mathfrak{u} = \{1,2,\ldots,d\}$. Hence, for $f \in F_d^*$ we have

$$S_d(f) - Q_{d,n}(f) = (-1)^d \int_{[0,1]^d} \mathrm{disc}(x)\frac{\partial^d}{\partial x} f(x)\mathrm{d}x.$$

Applying the Hölder inequality, we obtain that the worst case error of $Q_{n,d}$ is

$$e^{\mathrm{wor}}(Q_{d,n}) = \sup_{f\in F_d^*} \left|S_d(f) - Q_{d,n}(f)\right| = \mathrm{disc}(t_1, t_2, \ldots, t_n),$$

which is the star discrepancy for the points $t_1, t_2, \ldots, t_d$ that are used by the quasi-Monte Carlo algorithm $Q_{d,n}$.

We now remove the boundary conditions and consider the class

$$F_d = \left\{ f \in W_1^{\mathbf{1}} \mid \|f\|_{d,1} \le 1 \right\},$$

where the norm is given by

$$\|f\|_{d,1} = \sum_{\mathfrak{u}\subseteq\{1,2,\ldots,d\}} \int_{[0,1]^{|\mathfrak{u}|}} \left|\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f(x_{\mathfrak{u}}, 1)\right| \mathrm{d}x_{\mathfrak{u}}.$$

The term for $\mathfrak{u} = \emptyset$ corresponds to $|f(1)|$.

We return to the Hlawka and Zaremba identity and again apply the Hölder inequality, this time for integrals and sums, and conclude that the worst case error again is

$$e^{\mathrm{wor}}(Q_{d,n}) = \sup_{f\in F_d} \left|S_d(f) - Q_{d,n}(f)\right| = \mathrm{disc}(t_1, t_2, \ldots, t_n).$$

The multivariate problem is properly scaled for both classes F_d^* and F_d since the initial error is 1. Then

$$n(\varepsilon, d) = \min\{n \mid \text{there are } t_1, t_2, \ldots, t_n \in [0,1]^d \text{ with } \operatorname{disc}(t_1, t_2, \ldots, t_n) \le \varepsilon\}$$

is the same for both classes; this is just the inverse of the star-discrepancy.

Hence, tractability of multivariate problems depends on how the inverse of the star discrepancy behaves as a function of ε and d. Based on many negative results for classical spaces and on the fact that all variables play the same role for the star discrepancy, it would be natural to expect an exponential dependence on d, i.e., intractability. Therefore it was quite a surprise when a positive result was proved in [5]. More precisely, let

$$\operatorname{disc}(n, d) = \inf_{t_1, t_2, \ldots, t_n \in [0,1]^d} \operatorname{disc}(t_1, t_2, \ldots, t_n)$$

denote the minimal star discrepancy that can be achieved with n points in the d-dimensional case. Then there exists a positive number C such that

$$\operatorname{disc}(n, d) \le C d^{1/2} n^{-1/2} \quad \text{for all } n, d \in \mathbb{N}.$$

The proof of this bound follows directly from deep results of the theory of empirical processes. The proof is unfortunately non-constructive, and we do not know points for which this bound holds. The slightly worse upper bound

$$\operatorname{disc}(n, d) \le 2\sqrt{2} n^{-1/2} \left(d \ln \left(\left\lceil \frac{d n^{1/2}}{2 (\ln 2)^{1/2}} \right\rceil + 1 \right) + \ln 2 \right)^{1/2}$$

follows from Hoeffding's inequality and is quite elementary. Also this proof is non-constructive. However, using a probabilistic argument, it is easy to show that many points $t_1, t_2, \ldots, t_n$ satisfy both bounds modulo a multiplicative factor greater than one, see [5] for details.

The upper bounds on $\operatorname{disc}(n, d)$ can be easily translated into upper bounds on $n(\varepsilon, d)$. In particular, we have

$$n(\varepsilon, d) \le \left\lceil C^2 d \left(\frac{1}{\varepsilon} \right)^2 \right\rceil \quad \text{for all } \varepsilon \in (0,1) \text{ and } d \in \mathbb{N}.$$

This means that we have *polynomial tractability*. Furthermore it was also shown in [5] that there exists a positive number c such that

$$n(\varepsilon, d) \ge c\, d \ln \varepsilon^{-1} \quad \text{for all } \varepsilon \in (0, 1/64] \text{ and } d \in \mathbb{N}.$$

In fact, this lower bound holds not only for quasi-Monte Carlo algorithms, but also in full generality for all algorithms.

The theory of tractability of multivariate problems was initiated by Henryk and also mainly developed by Henryk – sometimes together with colleagues and friends. The recent paper [47] of Henryk is an excellent survey and also describes the history of this young field.

We now understand why certain multivariate problems are tractable or not. We also know how intractable problems can be modified to obtain tractable problems. But, again, with a new theory, there are also new problems. Actually, the recent book [15] contains 30 Open Problems that, hopefully, are a challenge for many mathematicians.[3]

5 Why is the work of Henryk so fascinating?

There are two reasons. First, the persistence of Henryk who attacks, from quite different angles, similar questions again and again: How can we describe and find good or even optimal algorithms for different continuous problems of mathematics? What properties do those algorithms have? Henryk wants to understand this by a comprehensive theory. This is visible already in Henryk's early work and gets even more prominent later when Henryk's interests cover the whole range of numerical mathematics, as well as other subjects, such as computational physics and quantum computers.

Secondly, the strength and patience that are needed to study problems in their detail: For this it was necessary to work in many different areas of mathematics.

It is not enough to develop a general theory. Also in mathematics a theory gets thought-provoking only through laborious investigations of many single problems that need many different skills and lots of energy. Henryk got a lot of deep results by studying such problems and examples; he had to study many parts of mathematics to obtain these results. In this way he influenced many fields, as can be seen by studying the Mathematical Reviews.

Most of us publish papers in a relative small field, and probably are happy to work a small amount in a second field. Henryk published results in numerical analysis and in many other fields, such as computability, number theory, lin-

[3] Open Problems 18 and 25 of [15] have been already solved by Stefan Heinrich, and Open Problem 26 by Anargyros Papageorgiou and Iasonas Petras. Partial answers to Open Problems 2 and 5 were found by Novak and Woźniakowski.

ear algebra, measure theory, interpolation and approximation, Fourier analysis, functional analysis, operator theory, probability theory and stochastic processes, statistics and quantum theory.

There do not exist many other mathematicians with a similar versatility. All these excursions into different parts of mathematics are still strongly related to the main basic question that was always studied by Henryk: How can we construct and understand optimal algorithms for numerical problems?

Today, Henryk is a main leader all over the world and is a great communicator who works together with excellent colleagues in many countries.

6 Four of Henryk's papers

It is difficult to select only four papers of Henryk. Probably it would be an interesting game to select the four "most influential" or the four "best" papers. To avoid this intractable problem I mention four papers that are cited most often – as can be checked with the Mathematical Reviews.[4]

Average case complexity of multivariate integration [43], 1991. Henryk studies the average case complexity of multivariate integration for the class $C([0,1]^d)$ equipped with the classical Wiener sheet measure. To derive the average case complexity one needs to obtain optimal sample points. This design problem was open for a long time. In this paper Henryk proves that the optimal design is closely related to discrepancy theory. The respective L_2-discrepancy problem was solved by K. F. Roth (the lower and the upper bounds being published in 1954 and 1980, respectively) and by K. K. Frolov (who also proved the upper bound), who showed that optimal sample points are given by shifted Hammersley points $z_1^*, z_2^*, \ldots, z_n^*$. Henryk showed that $1 - z_1^*, 1 - z_2^*, \ldots, 1 - z_n^*$ are the optimal sample points for the quadrature problem, and that the ε-complexity of the problem is of the order $\Theta(\varepsilon^{-1}(\ln \varepsilon^{-1})^{(d-1)/2})$.[5]

[4] This kind of selection discriminates against older papers as well as very young papers. This is obvious for very fresh papers. But also older papers have a disadvantage since most colleagues do not bother to cite a paper from the early eighties if they can also cite the IBC book. Therefore all four papers were published between 1991 and 2000.

[5] Observe that this paper studies the *optimal order of convergence*, not tractability. The tractability of the L_2-discrepancy problem was studied later.

Hence this paper is in the intersection of complexity theory, stochastic processes, discrepancy theory, number theory and numerical analysis. In particular, the paper proves that this intersection is non-empty.

Explicit cost bounds of algorithms for multivariate tensor product problems [29], with Grzegorz Wasilkowski, 1995. The authors study explicit error bounds for the Smolyak algorithm in the worst case setting and in the average case setting for multivariate tensor product problems.

In 1963, Smolyak introduced an algorithm for tensor product problems and proved bounds of the form

$$n(\varepsilon, d) \le C_d\, \varepsilon^{-\beta_1} (\ln \varepsilon^{-1})^{\beta_2 (d-1)}.$$

The interesting thing is that β_1, the order of convergence, does not depend on d. The constant C_d is, however, not known and therefore this is a typical classical result about the order of convergence. The authors prove bounds of the form

$$n(\varepsilon, d) \le C \left(\beta_1 + \beta_2 \frac{\ln \varepsilon^{-1}}{d-1} \right)^{\beta_3 (d-1)} \varepsilon^{-\beta_4},$$

where all the constants $C, \beta_1, \beta_2, \beta_3$ and β_4 are known and can be computed from error bounds for $d = 1$.

When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? [21], with Ian Sloan, 1998[6]. This paper is of fundamental importance for the understanding of high dimensional problems. It gives a partial answer to why quasi-Monte Carlo algorithms are successful, even in huge dimension. The authors define *weighted spaces* of functions using the idea that, for many applications, the number d of variables is huge, however, not all variables play the same role and some variables are "less important" than others. The idea of *weighted spaces* is central for the recent theory of tractability of multivariate problems.

Integration and approximation in arbitrary dimensions [6], with Fred Hickernell, 2000. The authors study several multivariate integration and approximation problems. They consider algorithms for classes $f \in F_d$ using function values. Let $n(\varepsilon, d)$ be the minimal number of function values needed for a worst

[6] This is the paper of Henryk that recently has been cited most often, according to Mathematical Reviews. By the way, it is also the paper of Ian Sloan that is most often cited.

case error ε in the dimension d for the class F_d. The authors are mainly interested in spaces and problems with the property

$$n(\varepsilon, d) \le C\, \varepsilon^{-p},$$

where C and p do not depend on d. Problems with this property are called *strongly* (*polynomially*) *tractable*. The authors prove that integration and approximation are strongly tractable for certain weighted Korobov and Sobolev spaces.

For the approximation problem the authors also consider algorithms that use arbitrary linear functionals instead of function values. The main result is that (under some assumptions) this much more general information is "not much" better, i.e., the ε-exponents stay the same.

7 Other directions

Here I just mention very few other of Henryk's research directions. Again my choice is very selective.

- **Linear optimization.** Traub and Woźniakowski study in [27] the ellipsoid algorithm and observe that, even if this algorithm is a polynomial-time algorithm within the bit number model, it has unbounded cost in the real number model. The authors conjecture that there does not exist a polynomial-time algorithm for the linear inequalities problem. This problem is still open today, in 2009.
- **Computation of fixed points.** It is known from work of Nemirovskii that it is impossible to improve the efficiency of the simple iteration whenever the dimension of the domain of contractive functions is large. However for a modest dimension, Sikorski, Tsay and Woźniakowski [18] exhibit a fixed point ellipsoid algorithm that is much more efficient than the simple iteration for mildly contractive functions. This algorithm is based on Khachiyan's construction of minimal volume ellipsoids used for solving linear programming.
- **Testing operators.** Together with David Lee, Henryk wrote several papers about testing and verification of linear and nonlinear operators, see, e.g., [12]. For the testing problem, A is an implementation of a specification S, both are mappings from a compact metric space F into a metric space G. Given $\varepsilon > 0$, one is allowed to compute Af and Sf for a finite number of

f and has to decide whether $d(Af, Sf) \le \varepsilon$ for all $f \in F$. It is shown that asymptotically correct sequences of guesses can be arranged. Sharp upper and lower bounds on the number of tests are given in terms of the Kolmogorov entropy of F. Probabilistic testing methods are developed and analyzed.

- **Tractability of path integration.** In [30], Wasilkowski and Woźniakowski analyze the complexity of computing integrals $\int_X f(x)\,\mu(\mathrm{d}x)$, where μ is a Gaussian measure on a Banach space X. For r times differentiable functions on X, the integration problem is tractable for deterministic algorithms iff the covariance operator is of finite rank. Hence for measures μ with infinite-dimensional covariance operator, Monte Carlo integration is superior to deterministic algorithms. For certain classes of entire functions on X, it is shown that the problem becomes tractable in the deterministic setting.

  Plaskota, Wasilkowski and Woźniakowski [17] suggest a new algorithm for the computation of Feynman–Kac path integrals. This algorithm has a very small cost, which gives a dramatic improvement of earlier results. However, there is also a problem since the new algorithm needs a lot of precomputation and therefore can (so far) only be used if the error requirements are moderate.
- **Quantum computers.** Henryk studied quantum computation and wrote several papers about optimal numerical algorithms in the quantum setting. Kwas and Woźniakowski [11] prove sharp error bounds for the Boolean summation problem. Traub and Woźniakowski [28] study path integration on a quantum computer.
- **Weighted problems and finite-order weights.** As already mentioned, the work [21] by Sloan and Woźniakowski has been continued by many colleagues. Dick, Sloan, Wang and Woźniakowski [1] discuss general weights in order to give recommendations for choosing the weights in practice. They defined in [2] finite-order weights. Such weights seem to be appropriate for many applications and they model functions of d variables that can be expressed as a sum of functions of k variables with k independent of d. The authors also prove new lower and upper bounds for the tractability of quasi-Monte Carlo algorithms for the computation of integrals for functions from weighted Sobolev spaces. For finite-order weights, we usually have tractability bounds depending polynomially on $d^{\beta(k)}$ with $\beta(k)$ linearly dependent on k.
- **Lower bounds.** To determine the complexity of a problem, we also need good lower bounds. This is straightforward for linear operators $S : F \to G$

between Hilbert spaces, if we consider algorithms that use arbitrary linear functionals in the worst case setting. Then we have to study the singular values of S. The proof of lower bounds is much more difficult in the randomized setting and/or if we consider algorithms that use function values. Many papers of Henryk deal with lower bounds, we mention [14], [20], [30].

- **Good lattice rules.** By lattice rule algorithms we mean algorithms that are based on function values at (sometimes shifted) lattice sample points. Such algorithms can be used for integration and approximation and in different settings. Some tractability results were first proved in a non-constructive way and today can be proved (by the work of Kuo, Joe, Sloan and others) in a constructive way, see also work of Cools and Nuyens. We mention [2], [8], [9], [22], where the reader can find more references.
- **Smolyak algorithm and generalizations.** The algorithm of Smolyak has been generalized to the concept of weighted tensor product algorithms to prove, in a constructive way, the tractability of many tensor product problems, see [31].
- **The power of standard information.** Often we know upper bounds for algorithms based on arbitrary linear information, based on estimates of singular values. It is then a challenge to prove similar upper bounds for algorithms based on function values – or to prove that such algorithms do not exist. Here we mention the paper [34], which deals with the randomized setting, and the paper [10], which deals with the worst case setting.
- **Generalized tractability.** In this short survey we discussed polynomial tractability that is defined by the requirement

$$n(\varepsilon, d) \le C \cdot \varepsilon^{-\alpha} d^{\beta} \tag{3}$$

for certain $C, \alpha, \beta > 0$ and all $\varepsilon > 0$ and $d \in \mathbb{N}$. There are different notions of tractability, since one might be interested in different tractability domains (for example, only d is large while ε^{-1} is modest) and different tractability functions, instead of polynomials. Then (3) is replaced by

$$n(\varepsilon, d) \le C \cdot T(\varepsilon^{-1}, d)^t,$$

for all $(\varepsilon^{-1}, d) \in \Omega$. This was studied by Henryk together with Gnewuch, see [3], [15].

- **Quasilinear problems.** Many IBC results have been proved for linear problems, for example for linear operator equations, where the solution u depends

linearly on the right hand side f. Together with Arthur Werschulz, Henryk studied the tractability of quasilinear problems. Many problems of mathematical physics belong to this class of problems. The paper [34] is the first paper in a series of papers.

Concluding remarks

I want to conclude with a few personal remarks. I have known Henryk since 1985, when Joseph Traub and Henryk invited me to a conference to New York. We immediately started discussions – which have not ended so far. Our first joint paper appeared in 1992. Henryk was always a good friend – actually, he is the nicest guy you can imagine. I thank him for 1000 suggestions and also for his sympathetic warmth.

References

[1] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Liberating the weights. *J. of Complexity* **20** (2004), 593–623. 19

[2] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights. *Numer. Math.* **103** (2006), 63–97. 19, 20

[3] M. Gnewuch and H. Woźniakowski, Generalized tractability for multivariate problems. I. Linear tensor product problems and linear information. *J. of Complexity* **23** (2007), 262–295. 20

[4] M. Griebel and H. Woźniakowski, On the optimal convergence rate of universal and nonuniversal algorithms for multivariate integration and approximation. *Math. Comp.* **75** (2006), 1259–1286.

[5] S. Heinrich, E. Novak, G. W. Wasilkowski and H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.* **96** (2001), 279–302. 14

[6] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimensions. *Adv. Comput. Math.* **12** (2000), 25–58. 17

[7] M. Jankowski and H. Woźniakowski, Iterative refinement implies numerical stability. *Nordisk Tidskr. Informationsbehandling* (*BIT*) **17** (1977), 303–311. 2

[8] F. Y. Kuo, I. H. Sloan and H. Woźniakowski, Lattice rules for multivariate approximation in the worst-case setting. In *Monte Carlo and quasi-Monte Carlo methods 2004*, Springer-Verlag, Berlin, 2006, 289–330. 20

[9] F. Y. Kuo, I. H. Sloan and H. Woźniakowski, Lattice rule algorithms for multivariate approximation in the average case setting. *J. of Complexity* **24** (2008), 283–323. 20

[10] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for multivariate approximation in the worst case setting. *J. Approx. Theory* **158** (2009), 97–125. 20

[11] M. Kwas and H. Woźniakowski, Sharp error bounds on quantum Boolean summation in various settings. *J. of Complexity* **20** (2004), 669–698. 19

[12] D. Lee and H. Woźniakowski, Testing nonlinear operators. *Numer. Alg.* **9** (1995), 319–342. 18

[13] E. Novak, K. Ritter and H. Woźniakowski, Average-case optimality of a hybrid secant-bisection method. *Math. Comp.* **64** (1995), 1517–1539.

[14] E. Novak and H. Woźniakowski, Intractability results for integration and discrepancy. *J. of Complexity* **17** (2001), 388–441. 20

[15] E. Novak and H. Woźniakowski, *Tractability of multivariate problems.* Volume I: Linear information, EMS Tracts Math. 6, European Mathematical Society Publishing House, Zürich, 2008. 15, 20

[16] A. Papageorgiou and H. Woźniakowski, The Sturm-Liouville eigenvalue problem and NP-complete problems in the quantum setting with queries. *Quantum Inf. Process.* **6** (2007), 101–120.

[17] L. Plaskota, G. W. Wasilkowski and H. Woźniakowski, A new algorithm and worst case complexity for Feynman-Kac path integration. *J. Comput. Phys.* **164** (2000), 335–353. 19

[18] K. Sikorski, C. W. Tsay and H. Woźniakowski, An ellipsoid algorithm for the computation of fixed points. *J. of Complexity* **9** (1993), 181–200. 18

[19] I. H. Sloan, X. Wang and H. Woźniakowski, Finite-order weights imply tractability of multivariate integration. *J. of Complexity* **20** (2004), 46–74.

[20] I. H. Sloan and H. Woźniakowski, An intractability result for multiple integration. *Math. Comp.* **66** (1997), 1119–1124. 10, 20

[21] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. of Complexity* **14** (1998), 1–33. 17, 19

[22] I. H. Sloan and H. Woźniakowski, Tractability of multivariate integration for weighted Korobov classes. *J. of Complexity* **17** (2001), 697–721. 20

[23] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information, uncertainty, complexity*. Addison-Wesley, Reading, MA, 1983.

[24] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-based complexity*. Academic Press, New York, 1988. 6

[25] J. F. Traub and H. Woźniakowski, Convergence and complexity of Newton iteration for operator equations. *J. Assoc. Comput. Mach.* **26** (1979), 250–258.

[26] J. F. Traub and H. Woźniakowski, *A general theory of optimal algorithms*. Academic Press, New York, 1980. 2

[27] J. F. Traub and H. Woźniakowski, Complexity of linear programming. *Oper. Res. Lett.* **1** (1981/82), 59–62. 18

[28] J. F. Traub and H. Woźniakowski, Path integration on a quantum computer. *Quantum Inf. Process.* **1** (2002), 365–388. 19

[29] G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems. *J. of Complexity* **11** (1995), 1–56. 17

[30] G. W. Wasilkowski and H. Woźniakowski, On tractability of path integration. *J. Math. Phys.* **37** (1996), 2071–2086. 19, 20

[31] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor product algorithms for linear multivariate problems. *J. of Complexity* **15** (1999), 402–447. 20

[32] G. W. Wasilkowski and H. Woźniakowski, On the power of standard information for weighted approximation. *Found. Comput. Math.* **1** (2001), 417–434.

[33] G. W. Wasilkowski and H. Woźniakowski, Polynomial-time algorithms for multivariate linear problems with finite-order weights: worst case setting. *Found. Comput. Math.* **5** (2005), 451–491.

[34] G. W. Wasilkowski and H. Woźniakowski, The power of standard information for multivariate approximation in the randomized setting. *Math. Comp.* **76** (2007), 965–988. 20, 21

[35] A. G. Werschulz and H. Woźniakowski, Tractability of quasilinear problems I. General results. *J. Approx. Theory* **145** (2007), 266–285.

[36] H. Woźniakowski, Some remarks on Bairstow's method. *Zastos. Mat.* **11** (1969/70), 207–214. 2

[37] H. Woźniakowski, On nonlinear iterative processes in numerical methods. Doctoral Thesis, University of Warsaw, 1972 (in Polish). 2

[38] H. Woźniakowski, Maximal stationary iterative methods for the solution of operator equations. *SIAM J. Numer. Anal.* **11** (1974), 934–949. 2

[39] H. Woźniakowski, Maximal order of multipoint iterations using n evaluations. In *Analytic computational complexity* (Proc. Symp., Carnegie-Mellon Univ., Pittsburgh, 1975), ed. by J. F. Traub, Academic Press, New York, 1976, 75–107. 2

[40] H. Woźniakowski, Numerical stability for solving nonlinear equations. *Numer. Math.* **27** (1977), 373–390. 2

[41] H. Woźniakowski, Numerical stability of the Chebyshev method for the solution of large linear systems. *Numer. Math.* **28** (1977), 191–209. 2

[42] H. Woźniakowski, Round-off error analysis of iterations for large linear systems. *Numer. Math.* **30** (1978), 301–314. 2

[43] H. Woźniakowski, Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* (*N.S.*) **24** (1991), 185–194. 16

[44] H. Woźniakowski, Average case complexity of linear multivariate problems. *Bull. Amer. Math. Soc.* (*N.S.*) **29** (1993), 70–76.

[45] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems. *J. of Complexity* **10** (1994), 96–128. 9

[46] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems. *J. Comput. Inform.* **4** (1994), 1–19. 9

[47] H. Woźniakowski, Tractability of multivariate problems. In *Foundations of computational mathematics* (Hong Kong 2008), ed. by F. Cucker, A. Pinkus and M. Todd, London Math. Soc. Lecture Note Ser. 363, Cambridge University Press, Cambridge, 236–276. 15

Publications of Henryk Woźniakowski

Books

1. (with J. F. Traub) *A general theory of optimal algorithms*. Academic Press, New York, 1980; Russian translation by MIR, Moscow, 1983.
2. (with J. F. Traub and G. W. Wasilkowski) *Information, uncertainty, complexity*. Addison-Wesley, Reading, Ma., 1983; Russian translation by MIR, Moscow, 1988.
3. (with J. F. Traub and G. W. Wasilkowski) *Information-based complexity*. Academic Press, New York, 1988.
4. (editor with B. Bojanov) *Optimal recovery*. Nova Science, New York, 1992
5. *Festschrift for Joseph F. Traub*. Editor, Academic Press, New York, 1993.
6. (with E. Novak) *Tractability of multivariate problems*. Volume I: Linear information. EMS Tracts Math. 6, European Mathematical Society Publishing House, Zürich, 2008.

Articles

1. Some remarks on Bairstow's method. *Zastos. Mat.* 11 (1969/1970), 207–214.
2. Polynomial decomposition into quadratic factors with controlled accuracy by Bairstow's method. *Zastos. Mat.* 11 (1969/1970), 215–220.
3. Modification of the alternating direction implicit method and connections with von Neumann's method. *Zastos. Mat.* 12 (1971/1972), 397–411.
4. On Nonlinear iterative processes in numerical methods. Doctoral Thesis, University of Warsaw, 1972 (in Polish).
5. (with A. Kiełbasinski and G. Woźniakowska) Algorithmization of the best strategy methods for large linear systems with positive definite matrix. *Mat. Stos.* (3) 1 (1973), 47–68 (in Polish).
6. Rounding Error analysis for the evaluations of a polynomial and some of its derivatives. *SIAM J. Numer. Anal.* 11 (1974), 780–787.
7. Maximal stationary iterative methods for the solution of operator equations. *SIAM J. Numer. Anal.* 11 (1974), 934–949.

8. Properties of maximal order methods for the solution of nonlinear equations. *Z. Angew. Math. Mech.* 55 (4) (1975), T268–T271.

9. (with G. Woźniakowska) Algorithmization of the me-T method. *Mat. Stos.* (3) 5 (1975), 51–60 (in Polish).

10. The method of minimal B-errors for large systems of linear equations with an arbitrary matrix. *Mat. Stos.* (3) 5 (1975), 5–27 (in Polish).

11. Generalized information and maximal order of iteration for operator equations. *SIAM J. Num. Anal.* 12 (1975), 121–135.

12. Maximal order of multipoint iterations using n evaluations. In *Analytic computational complexity*, ed. by J. F. Traub, Academic Press, New York, 1976, 75–107.

13. Numerical stability of iterations for solution of nonlinear equations and large linear systems. In *Analytic computational complexity*, ed. by J. F. Traub, Academic Press, New York, 1976, 177–190.

14. (with J. F. Traub) Strict lower and upper bounds on iterative computational complexity. In *Analytic computational complexity*, ed. by J. F. Traub, Academic Press, New York, 1976, 15–34.

15. (with J. F. Traub) Optimal linear information for the solution of nonlinear operator equations. In *Algorithms and complexity: new directions and recent results*, ed. by J. F. Traub, Academic Press, New York, 1976, 103–119.

16. Numerical stability for solving nonlinear equations. *Numer. Math.* 27 (1977), 373–390.

17. Numerical stability of the Chebyshev method for the solution of large linear systems. *Numer. Math.* 28 (1977), 191–209.

18. (with B. Kacewicz) A survey of recent problems and results in analytic computational complexity. In *Mathematical foundations of computer science*, ed. by J. Gruska, Lecture Notes in Comput. Sci. 53, Springer-Verlag, Berlin, 1977, 93–107.

19. (with M. Jankowski) Iterative refinement implies numerical stability. *Nordisk Tidskr. Informationsbehandling* (*BIT*) 17 (3) (1977), 303–311.

20. Round-off error analysis of iterations for large linear systems. *Numer. Math.* 30 (1978), 301–314.

21. (with J. F. Traub) Optimal radius of convergence of interpolatory iterations for operator equations. *Aequationes Math.* 21 (2–3) (1980), 159–172.

22. (with J. F. Traub) Convergence and complexity of interpolatory-Newton iteration in a Banach space. Comput. Math. Appl. 6 (4) (1980), 385–400.

23. Round-off error analysis of a new class of conjugate gradient algorithms. *Linear Algebra Appl.* 29 (1980), 507–529.

24. (with J. F. Traub) Complexity of linear programming, *Oper. Res. Lett.* 1 (2) (1981/82), 59–62.

25. (with M. Jankowski and A. Smoktunowicz) A note on floating point summation of very many terms. *Elektron. Informationsverarb. Kybernet.* 19 (9) (1983), 435–440.

26. (with J. F. Traub and G. W. Wasilkowski) Average case optimality for linear problems. *Theoret. Comput. Sci.* 29 (1984), 1–25.

27. (with J. F. Traub) Information and computation. In *Advances in computers* 23, ed. by M. C. Yovits, Academic Press, Orlando, FL, 1984, 35–92.

28. (with G. W. Wasilkowski) Can adaption help on the average? *Numer. Math.* 44 (1984), 169–190.

29. (with J. F. Traub) On the optimal solution of large linear systems. *J. Assoc. Comput. Mach.* 31 (1984), 545–559.

30. (with J. F. Traub and G. W. Wasilkowski) When is nonadaptive information as powerful as adaptive information? In *Proceedings of the 23rd IEEE Conference on Decision and Control*, 1984, 1536–1540.

31. (with J. F. Traub and Y. Yemini) Statistical security of data base. *Transactions on Data Base Systems* 9 (1984), 672–679.

32. (with M. Jankowski) The accurate solution of certain continuous problems using only single precision. *BIT* 25 (1985), 635–651.

33. (with M. A. Kowalski and A.G. Werschulz) Is Gauss quadrature optimal for analytic functions? *Numer. Math.* 47 (1985), 89–98.

34. A survey of information-based complexity. *J. of Complexity* 1 (1985), 11–44.

35. (with K. Sikorski) For which error criterion can we solve nonlinear equations? *J. of Complexity* 2 (1986), 163–178.

36. (with G. W. Wasilkowski) Average case optimal algorithms in Hilbert spaces. *J. Approx. Theory*, 47 (1986), 17–25.

37. (with M. Shub, N. Weiss and G. W. Wasilkowski) Average condition number for solving linear equations. *Linear Algebra Appl.* 83 (1986), 79–102.

38. Probabilistic case setting of information-based complexity. *J. of Complexity* 2 (1986), 255–269.

39. Information-based complexity. *Annual Review of Computer Science* 1 (1986), by Annual Reviews Inc., 319–380.

40. Complexity of integration in different settings (extended abstract). In *Optimal algorithms*, Proceedings of the international symposium in Blagoevgrad, (1986), Sofia, Bulgarian Academy of Sciences, 235–240.

41. (with A. G. Werschulz) Are linear algorithms always good for linear problems? *Aequationes Math.* 31 (1986), 202–211.

42. Average complexity for linear operators over bounded domains. *J. of Complexity* 3 (1987), 57–80.

43. (with E. W. Packel) Recent developments in information-based complexity. *Bull. Amer. Math. Soc.* (*N.S.*) 17 (1987), 9–36.

44. (with T. Jackowski) Complexity of approximation with relative error criterion in worst, average and probabilistic settings. *J. of Complexity* 3 (1987), 114–134.

45. (with K. Sikorski) Complexity of fixed points I. *J. of Complexity* 3 (1987), 388–405.

46. (with G. W. Wasilkowski) On optimal algorithms in an asymptotic model with gaussian measure. *SIAM J. Math. Anal* 3 (1988), 632–647.

47. (with J. B. Kadane and G. W. Wasilkowski) Can adaption help on the average for stochastic information? *J. of Complexity* 4 (1988), 257–276.

48. (with G. W. Wasilkowski) Mixed settings for linear problems. *J. of Complexity* 5 (1989), 457–465.

49. (with J. F. Traub) Information-based complexity: new questions for mathematicians. *Mathematical Intelligencer* 13 (1991), 34–43.

50. Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* (*N.S.*) 24 (1991), 185–194.

51. (with J. F. Traub) Theory and applications of information-based complexity. In *1990 Lectures in complex systems*, ed. by L. Nadel and D. Stein, Santa Fe Inst. Stud. Sci. Complexity Lectures III, Addison-Wesley, Redwood City, CA, 1991, 163–193.

52. (with E. W. Packel and J. F. Traub) Measures of uncertainty and information in computation. *Inform. Sci.* 65 (1992), 253–273.

53. (with J. Kuczyński) Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start. *SIAM J. Matrix Anal.* Appl. 13 (1992), 1094–1122.

54. (with J. F. Traub) The Monte Carlo algorithm with a pseudo-random generator. *Math. Comput* 58 (1992), 323–339.

55. (with J. F. Traub) Perspectives on information-based complexity. *Bull. Amer. Math. Soc.* (*N.S.*) 26 (1992), 29–52.

56. Complexity of verification and computation for IBC problems. *J. of Complexity* 8 (1992), 93–123.

57. (with E. Novak) Relaxed verification for continuous problems. *J. of Complexity* 8 (1992), 124–152.

58. Average case complexity of linear multivariate problems. I: Theory; II: Applications. *J. of Complexity* 8 (1992), 337–373; 374–392.

59. Average Case complexity of linear multivariate problems. *Bull. Amer. Math. Soc.* (*N.S.*) 29 (1993), 70–76.

60. (with K. Sikorski and C. W. Tsay) An ellipsoid algorithm for the computation of fixed points. *J. of Complexity* 9 (1993), 181–200.

61. (with G. W. Wasilkowski) There exists a linear problem with infinite combinatory complexity. *J. of Complexity* 9 (1993), 326–337.

62. (with K. Ritter and G. W. Wasilkowski) On multivariate integration for stochastic processes. In *Numerical integration IV*, ed. by H. Brass and G. Hämmerlin, Internat. Ser. Numer. Math. 112, Birkhäuser, Basel, 1993, 331–347.

63. (with J. F. Traub) Recent progress in information-based complexity. *Bulletin of EATCS* 51 (1993), 141–154.

64. (with J. F. Traub) Breaking intractability. *Scientific American* 270 (1994), 102–107.

65. Tractability and strong tractability of linear multivariate problems. *J. of Complexity* 10 (1994), 96–128.

66. On tractability of linear multivariate problems (extended abstract). In *Proceedings of the conference open problems in approximation theory*, Voneshta Voda, Bulgaria, 1994, 213–236.

67. (with J. Kuczyński) Probabilistic bounds on the extremal eigenvalues and condition number by the Lanczos algorithm. *SIAM J. Matrix Anal. Appl.* 15 (2) (1994), 672–691.

68. Information-based complexity. Abstract of three lectures, Special semester in approximation theory, Institute of Advanced Studies in Mathematics, Technion, Haifa, Israel, 1994, 25–31.

69. (with G. W. Wasilkowski) On strong tractability of multivariate problems. In *Proceedings of IFIP' 94*, Hamburg, Germany, 1994, 621–628.

70. Tractability and strong tractability of multivariate tensor product problems. *J. of Computing and Information* 4 (1994), 1–19.

71. (with K. Ritter and G. W. Wasilkowski) Multivariate integration and approximation for random fields satisfying Sacks–Ylvisaker conditions. *Ann. Appl. Probab.* 5 (1995), 518–541.

72. (with G. W. Wasilkowski) Explicit cost bounds of algorithms for multivariate tensor product problems. *J. of Complexity* 11 (1995), 1–56.

73. (with E. Novak and K. Ritter) Average case optimality of a hybrid secant-bisection method. *Math. Comput.* 64 (1995), 1517–1539.

74. (with D. Lee) Testing linear operators. *BIT* 35 (1995), 331–351.

75. (with D. Lee) Testing nonlinear operators. *Numer. Algorithms* 9 (1995), 319–342.

76. (with G. W. Wasilkowski) On tractability of path integration. *J. of Math. Physics*, 37(4), 2071–2088, 1996.

77. Overview of information-based complexity. In *The mathematics of numerical analysis*, ed. by J. Renegar, M. Shub and S. Smale, Lectures in Appl. Math. 32, Amer. Math. Soc., Providence, R.I., 1996, 915–927.

78. An introduction to information-based complexity. ICIAM/GAMM 95, Hamburg, July 3–7, 1995, vol. 3, special issue of *ZAMM* (1996), 128–130.

79. Complexity of multivariate problems with applications to path integrals. ICIAM/GAMM 95, Hamburg, July 3–7, 1995, vol. 3, special issue of *ZAMM* (1996), 131–134.

80. (with E. Novak) Topological complexity of zero finding. *J. of Complexity* 12 (1996), 380–400.

81. Computational complexity of continuous problems. In *Nonlinear dynamics, chaotic and complex systems*, ed. by E. Infeld, R. Zelazny and A. Gałkowski, Cambridge University Press, Cambridge, 1997, 283–295.

82. Strong tractability of weighted tensor products. Ann. Numer. Math. 4 (1997), 607–622.

83. (with D. Lee) Testing linear operators – an average case analysis. *BIT* 37 (1997), 355–376.

84. (with I. H. Sloan) An intractability result for multiple integration. *Math. Comput.* 66 (1997), 1119–1124.

85. (with G. W. Wasilkowski) The exponent of discrepancy is at most 1.4778.... *Math. Comput.* 66 (1997), 1125–1132.

86. (with E. Novak and I. H. Sloan) Tractability of tensor product linear operators. *J. of Complexity* 13 (1997), 387–418.

87. (with I. H. Sloan) When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. of Complexity* 14 (1998), 1–33.

88. (with V. Agrawal and D. Lee) Numerical computation of characteristic polynomials of Boolean functions and its applications. *Numer. Algorithms* 17 (1998), 261–278.

89. (with Z. Leyk) Estimating a largest eigenvector by polynomial algorithms with a random start. *Numer. Linear Algebra Appl.* 5 (1998), 147–164.

90. (with I. H. Sloan) Multiple integrals in many dimensions. In *Advances in computational mathematics*, Proceedings of the Guangzhou international symposium, ed. by Z. Chen, Y. Li, Ch. Micchelli, and Y. Xu, Lecture Notes in Pure Appl. and Math. 202, Marcell Dekker, New York, 1999, 507–516.

91. (with E. Novak) On the cost of uniform and nonuniform algorithms. *Theoret. Comput. Sci.* 219 (1999), 301–318.

92. Why does information-based complexity use the real number model? *Theoret. Comput. Sci.* 219 (1999), 451–466.

93. (with G. W. Wasilkowski) Weighted tensor product algorithms for linear multivariate problems. *J. of Complexity* 15 (1999), 402–447.

94. Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals. In *Monte Carlo and quasi-Monte Carlo methods 1998*, ed. by H. Niederreiter and J. Spanier, Springer-Verlag, Berlin, 1999, 114–136.

95. (with F. J. Hickernell) Integration and approximation in arbitrary dimensions. *Adv. Comput. Math.* 12 (2000), 25–58.

96. (with G. W. Wasilkowski) Complexity of weighted approximation over R. *J. Approx.* 103 (2000), 223–251.

97. (with E. Novak) Complexity of linear problems with a fixed basis for the output. *J. of Complexity* 16 (2000), 333–362.

98. (with L. Plaskota and G. W. Wasilkowski) A new algorithm and worst case complexity for Feynman–Kac path integration. *J. Comput. Phys.* 164 (2000), 335–353.

99. (with E. Novak) Intractability results for integration and discrepancy. *J. of Complexity* 17 (2001), 388–441.

100. (with S. Heinrich, E. Novak and G. W. Wasilkowski) The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.* 96 (3) (2001), 279–302.

101. (with G. W. Wasilkowski) On the complexity of stochastic integration. *Math. Comput.* 70 (2001), 685–698.

102. (with A. G. Werschulz) What is the complexity of surface integration? *J. of Complexity* 17 (2001), 442–466.

103. Tractability of tensor product linear operators in weighted Hilbert spaces. Dedicated to Professor Nicholas Vakhania on the occasion of his 70th birthday, *Georgian Math. J.* 8 (2001), 415–426.

104. (with D. Lee) A probabilistic analysis of linear operator testing. *J. of Complexity* 17 (2001), 516–540.

105. (with E. Novak) When are integration and discrepancy tractable? In *Foundations of computational mathematics*, ed. by R. A. DeVore, A. Iserlis and E. Süli, London Math. Soc. Lecture Note Ser. 284, Cambridge University Press, Cambridge, 2001, 211–266.

106. (with D. Lee) Evaluations of characteristic polynomials of Boolean functions. *Theoret. Comput. Sci.* 262 (2001), 37–68.

107. (with G. W. Wasilkowski) On the power of standard information for weighted approximation. *Found. Comput. Math.* 1 (2001), 417–434.

108. (with F. J. Hickernell) The price of pessimism for multidimensional quadratures. *Complexity of multivariate problems* (Kowloon, 1999), *J. of Complexity* 17 (2001), 625–659.

109. (with F. J. Hickernell) Tractability of multivariate integration for periodic functions. *Complexity of multivariate problems* (Kowloon, 1999), *J. of Complexity* 17 (2001), 660–682.

110. (with I. H. Sloan) Tractability of multivariate integration for weighted Korobov classes. *Complexity of multivariate problems* (Kowloon, 1999), *J. of Complexity* 17 (2001), 697–721.

111. (with G. W. Wasilkowski) Complexity of weighted approximation over R^d. *Complexity of multivariate problems* (Kowloon, 1999), *J. of Complexity* 17 (2001), 722–740.

112. (with J. Rust and J. F. Traub) Is there a curse of dimensionality for contraction fixed points in the worst case? *Econometrica* 70 (2002), 285–329.

113. (with G. W. Wasilkowski) Tractability of approximation and integration for weighted tensor product problems over unbounded domains. In *Monte Carlo and quasi-Monte Carlo methods 2000*, ed. by K.-T. Fang, F. J. Hickernell, and H. Niederreiter, Springer-Verlag, Berlin, 2002, 497–522.

114. (with E. Novak) IBC-problems related to Steve Smale. In *Foundations of computational mathematics,* (Proceedings of the Smalefest, 2000), ed. by F. Cucker and J. M. Rojas, World Scientific, River Edge, N.J., 2002, 295–318.

115. (with I. H. Sloan) Tractability of integration in non-periodic and periodic weighted tensor product Hilbert spaces. *J. of Complexity* 18 (2002), 479–499.

116. (with A. G. Werschulz) What is the complexity of volume calculation? *J. of Complexity* 18 (2002), 660–678.

117. (with J. F. Traub) Path integration on a quantum computer. *Quantum Inf. Process.* 1 (2002), 365–388.

118. Open problems for tractability of multivariate integration, *J. of Complexity* 19 (2003), 434–444.

119. (with A. G. Werschulz) Surface approximation is sometimes easier than surface integration. *Constr. Approx.* 20 (2004), 267–302.

120. (with I. H. Sloan) When does Monte Carlo depend polynomially on the number of variables? In *Monte Carlo and quasi-Monte Carlo methods 2002*, ed. by H. Niederreiter, Springer-Verlag, Berlin, 2004, 407–438.

121. (with M. Kwas) Sharp error bounds on quantum Boolean summation in various settings. *J. of Complexity* 20 (2004), 669–698.

122. (with S. Heinrich and M. Kwas) Quantum Boolean summation with repetitions in the worst-average case setting. In *Monte Carlo and quasi-Monte Carlo methods 2002*, ed. by H. Niederreiter, Springer-Verlag, Berlin, 2004, 243–258.

123. (with I. H. Sloan and X. Wang) Finite-order weights imply tractability of multivariate integration. *J. of Complexity* 20 (2004), 46–74.

124. (with E. Novak and I. H. Sloan) Tractability of approximation for weighted Korobov spaces on classical and quantum computers. *Found. Comput. Math.* 4 (2004), 121–156.

125. (with J. Dick, I. H. Sloan, and X. Wang) Liberating the weights. *J. of Complexity* 20 (2004), 593–623.

126. (with G. W. Wasilkowski) Finite-order weights imply tractability of linear multivariate problems. *J. Approx.* 130 (2004), 57–77.

127. On generalized tractability for multivariate problems. In *Approximation theory: a volume dedicated to Borislav Bojanov*, ed. by Dimitar K. Dimitrov, Geno Nikolov and Rumen Uluchev, Prof. M. Drinov Acad. Publ. House, Sofia, 2004, 312–324.

128. (with A. Papageorgiou) Classical and quantum complexity of the Sturm–Liouville eigenvalue problem. *Quantum Inf. Process.* 4 (2) (2005), 87–127.

129. (with G. W. Wasilkowski) Polynomial-time algorithms for multivariate problems with finite-order weights: worst case setting. *Found. Comput. Math.* 5 (2005), 451–491.

130. (with F. Y. Kuo and I. H. Sloan) Lattice rules for multivariate approximation in the worst case setting. In *Monte Carlo and quasi-Monte Carlo methods 2004*, ed. by H. Niederreiter and D. Talay, Springer-Verlag, Berlin, 2006, 289–330.

131. (with J. Dick, I. H. Sloan and X. Wang) Good lattice rules in weighted Korobov spaces with general weights. *Numer. Math.* 103 (2006), 63–97.

132. Tractability of multivariate linear problems for weighted spaces of functions. In *Approximation and Probability*, ed. by A. Kamont and T. Figiel, Banach Center Publ. 72, Polish Academy of Sciences, Institute of Mathematics, Warsaw, 2006, 407–427.

133. (with M. Griebel) On the optimal convergence rate of universal and non-universal algorithms for multivariate integration and approximation. *Math. Comput.* 75 (2006), 1259–1286.

134. (with G. W. Wasilkowski) The power of standard information for multivariate approximation in the randomized setting. *Math. Computation*, 76, 965–988, 2006.

135. The quantum setting with randomized queries for continuous problems, *Quantum Inf. Process.* 5 (2006), 83–130.

136. (with A. G. Werschulz) Tractability of quasilinear problems I: general results. *J. Approx.* 145 (2007), 266–285.

137. (with A. G. Werschulz) Tractability of quasilinear problems II: second order elliptic problems. *Math. Comp.* 76 (2007), 745–776.

138. (with M. Gnewuch) Generalized tractability for multivariate problems, Part I. Linear tensor product problems and linear information. *J. of Complexity* 23 (2007), 262–295.

139. (with A. Papageorgiou) The Sturm-Liouville eigenvalue problem and NP-complete problems in the quantum setting with queries. *Quantum Inf. Process.* 6 (2) (2007), 101–120.

140. (with F. Y. Kuo and I. H. Sloan) Periodization strategy may fail in high dimensions. *Numer. Algorithms* 46 (2007), 369–391.

141. (with F. Y. Kuo and I. H. Sloan) Lattice rules for multivariate approximation in the average case setting. *J. of Complexity* 24 (2008), 283–323.

142. (with F. J. Hickernell and G. W. Wasilkowski) Tractability of linear multivariate problems in the average case setting. In *Monte Carlo and quasi-Monte Carlo* 2006, ed. by A. Keller, S. Heinrich, H. Niederreiter, Springer-Verlag, Berlin, 2008, 461–493.

143. (with M. Gnewuch) Generalized tractability for linear functionals. In *Monte Carlo and quasi-Monte Carlo 2006*, ed. by A. Keller, S. Heinrich, H. Niederreiter, Springer-Verlag, Berlin, 2008, 359–381.

144. (with F. Y. Kuo and G. W. Wasilkowski) Multivariate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces. *J. Approx.* 152 (2008), 135–160.

145. Tractabilty of multivariate problems. In *Foundations of computational mathematics* (Hong Kong 2008), ed. by F. Cucker, A. Pinkus and M. Todd, London Math. Soc. Lecture Note Ser. 363, Cambridge University Press, Cambridge, 236–276.

146. (with F. Y. Kuo and G. W. Wasilkowski) On the power of standard information for multivariate approximation in the worst case setting. *J. Approx. Theory* 158 (2009), 97–125.

147. Complexity as a challenge for mathematicians, this booklet.

148. (with E. Novak) L_2 discrepancy and multivariate integration. In *Analytic number theory. Essays in honour of Klaus Roth*, ed. by W. W. L. Chen, W. T. Gowers, H. Halberstam, W. M. Schmidt and R. C. Vaughan, Cambridge University Press, Cambridge, 359–388.

149. (with G. W. Wasilkowski) Polynomial-time algorithms for multivariate linear problems with finite-order weights: average case setting. *Found. Comput. Math.* 9 (2009), 105–132.

150. (with G. W. Wasilkowski) A survey of average case complexity for linear multivariate problems. *Izv. VUZ Mat.* 4 (2009), 3–20 (in Russian).

151. (with M. Gnewuch) Generalized tractability for multivariate problems, Part II. Linear tensor product problems, linear information, and unrestricted tractability. To appear in *Found. Comput. Math.*, DOI 10.1007/s10208-009-9044-6.

152. (with A. G. Werschulz) Tractability of multivariate approximation for a weighted unanchored Sobolev space. To appear in *Constr. Approx.*

153. (with E. Novak) Approximation of infinitely differentiable multivariate functions is intractable. To appear in *J. of Complexity.*

154. (with E. Novak) Optimal order of convergence and (in-) tractability of multivariate approximations of smooth functions. To appear in *Constr. Approx.*

155. (with G. W. Wasilkowski) On the exponent of discrepancies. To appear in *Math. Comput.*

156. (with F. Y. Kuo and G. W. Wasilkowski) Lattice algorithms for multivariate L_∞ approximation in the worst case setting.To appear in *Constr. Approx.*

157. (with F. Y. Kuo and G. W. Wasilkowski) On the power of standard information for L_∞ approximation in the randomized setting. To appear in *BIT.*

Articles submitted for publication

1. (with F. Y. Kuo, I. H. Sloan and G. W. Wasilkowski) On decompositions of multivariate functions.
2. (with A. Papageorgiou) Tractability through increasing smoothness.
3. (with F. Y. Kuo, I. H. Sloan and G. W. Wasilkowski) Liberating the dimension.


Friedrich Schiller University Jena
Prof. Dr. Heinrich Wessenborn

Complexity as a new challenge for mathematicians

Henryk Woźniakowski

1 Introduction

This paper is based on my presentation when I received an honorary doctoral degree from the Friedrich Schiller University of Jena in June 2008. I wish to begin with a few personal thoughts regarding the University of Jena and Jena's mathematicians.

In 1964 I started my mathematical study at the University of Warsaw. My favorite subject was always analysis, and gradually I became more and more interested in theoretical numerical analysis. We had many classes covering classical and modern aspects of analysis. Many famous names of mathematicians were mentioned as we studied their theorems. When functional analysis and function spaces were covered, I learnt that the University of Jena is a famous place in mathematics and the names of Professors Albrecht Pietsch and Hans Triebel, two prominent mathematicians from University of Jena were often mentioned. I have not then expected that after many years I would have a privilege to meet them in Jena and listen to many of their talks.

In 1985, Joseph Traub and I organized a conference on approximately solved problems at Columbia University in New York. During this conference, I first met Erich Novak. It was obvious from the very beginning that we share many common research interests with Erich. As a partial proof of this claim, I would like to add that so far, I have published 13 papers with Erich, and that the first volume of our book *Tractability of Multivariate Problems* has been recently published by the European Mathematical Society. In 2001, Erich moved to the University of Jena as Professor of Theoretical Numerics. I visited him in Jena a couple of times before 2006. In 2006, I received a Humboldt Research Award and spent the 2006–2007 academic year at the University of Jena working with Erich on our book. I am happy that Erich was asked to present the Laudatio of my honorary doctoral degree.

I believe it was in 1987 when I met two young mathematicians, Stefan Heinrich and Peter Mathé, who both had strong ties with the University of Jena. Stefan

received his PhD degree in 1976 and his advisors were Professors Albrecht Pietsch and M. I. Kadec from the Kharkov University. In 1980, Stefan got his habilitation from the University of Jena, and in 1983–1985, he was a Visiting Professor at the University of Jena. Peter got his PhD degree in 1981 from the University of Jena and his advisor was Professor Werner Linde.

Today Erich, Stefan and Peter are my close friends, and we and our many colleagues are working in information-based complexity, which addresses computational complexity of continuous problems.

I wish to add a few words about the time I spent in Jena as a Humboldt fellow in 2006–2007. It was a happy time for me and my wife. I got sabbatical leaves from both Columbia University and the University of Warsaw, and I was without any teaching or administrative duties. My time was fully devoted to research, and more specifically to writing a book on tractability with Erich. This would have been enough for a successful stay, but pretty soon we realized that there is much more we would enjoy in Jena. We enjoyed the friendly atmosphere of the Mathematisches Institut. I still remember nice lunches at the local mensa, and in a Japanese restaurant close to the Institute, many dinners in excellent restaurants of Jena after colloquia and seminars of many Jena visitors. My wife often also took part in these evening events. We wish to thank all colleagues from Jena to make our stay in Jena so nice.

I am very honored and feel privileged to receive an honorary doctoral degree from the University of Jena. I wish to thank all my Jena colleagues who thought about me as a possible candidate for the honorary degree and who organized such a nice celebration of this event. I wish to add that Jena will be always a precious place for me.

As a token of my appreciation, I would like to talk now about my favorite mathematical subject, that is, about complexity. I want to address a general audience of mathematicians who are not necessarily familiar with computational complexity. My major message is not to burden my fellow mathematicians with many new complexity results but to convince them that computational complexity presents a new set of questions that are relatively new and whose answers probably require new proof techniques. From this point of view, computational complexity is a new challenge for mathematicians, and I would like to encourage as many mathematicians as possible to get familiar with this new field and maybe to stay in this field for a long time.

I wish to illustrate computational complexity by three examples. The first one is the easiest to explain. We have two $n \times n$ matrices and we want to multiply them. This looks like a trivial problem. Indeed, it is obvious that knowing all the

coefficients of two matrices we can compute their product using $\mathcal{O}(n^3)$ multiplications and additions. So, where is the problem? Well, the new and key question is to ask whether we can do it better. Or more generally, we can ask what is the computational complexity of matrix multiplication? That is, what is the *minimal* number of multiplications and additions, or arithmetic operations, that is necessary to multiply two arbitrary $n \times n$ matrices with real coefficients? Clearly, we must perform at least of order n^2 operations since we have $2n^2$ real coefficients and the resulting matrix depends on all of them. But are $\mathcal{O}(n^2)$ arithmetic operations enough? We do not know. We also do not know the smallest β for which $\mathcal{O}(n^\beta)$ arithmetic operations are enough to compute matrix multiplication. We only know that $\beta \in [2, 2.388]$, with the upper bound on β due to Coppersmith and Winograd [4]. Why is it so difficult to solve this problem that is so easy to define? A possible answer is that we do not have a proper proof technique. This complexity question was asked only about forty years ago, and apparently forty years is not enough to develop proper tools and concepts needed for the solution.

Obviously, the ease of formulation of a problem has nothing to do with the difficulty of its solving, and Fermat's last theorem is probably the best example. We needed to wait some 350 years for the solution of the Fermat problem. I hope we do not have to wait so long for the solution of matrix multiplication, since I would like to see the solution during my time.

The purpose of the second and third examples is to show that some multivariate problems may have an excellent rate of convergence but still suffer from the curse of dimensionality. More precisely, we will consider multivariate integration and approximation defined over certain classes of functions. For multivariate integration, as in [10], we choose a class of periodic and smooth functions f of d variables. We approximate the integral of f by an algorithm that uses n function values at optimally chosen sample points. We measure the error by the worst performance of an algorithm. It turns out that the worst case error as a function of n tends to zero like n^{-p}, where p can be arbitrarily close to the smoothness parameter α of the class. Hence, for large α, the exponent p is also large and we enjoy an excellent rate of convergence. If we want to guarantee that the worst case error is at most ε then

$$n = \mathcal{O}\left(\varepsilon^{-1/\alpha-\delta}\right) \quad \text{for all } \delta > 0.$$

So for large α, the exponent of ε^{-1} is small and the problem seems to be easy. Usually the standard analysis finishes at this point with the claim of a positive result that multivariate integration of periodic and smooth functions is easy.

As the third example, we consider the space of r-times continuously differentiable functions with respect to all d variables and equip the space with the r-folded Wiener measure. We now approximate functions f by an algorithm that uses n arbitrary linear continuous functionals and measure the error by an average performance of an algorithm. Papageorgiou and Wasilkowski [8] proved that the average case error of an optimal algorithm tends to zero like n^{-p}, where p can be arbitrarily close to $r + 1/2$. Hence, p is at least roughly $1/2$ and p is large for large r. So we enjoy an excellent rate of convergence for large r. Again, if we want to guarantee that the average case error is at most ε then

$$n = \mathcal{O}\left(\varepsilon^{-1/(r+1/2)-\delta}\right) \quad \text{for all } \delta > 0.$$

In this case, the exponent of ε^{-1} is roughly at most 2 for $r = 0$ which corresponds to the classical Wiener sheet measure. For large r, the exponent of ε^{-1} is much smaller. Again, the problem seems to be easy and such a message is usually sent by the standard analysis.

What is possibly missing in this line of thought? Well, all the good news is based on the big $\mathcal{O}$ notation. This means that we do *not* control the factors in this notation. These factors, although independent of ε^{-1}, may depend on d. As long as d is relatively small this is probably irrelevant. But what happens if d is large? Problems with large d occur quite often in computational practice and there is a fast growing interest in solving problems with large d, that is, in solving high dimensional problems.

Let us then use the complexity approach and try to find the complexity of these two problems as a function of two parameters ε^{-1} and d. It turns out that, modulo some details, the complexity is equal to $n(\varepsilon, d)$, which is the *minimal* number of information operations needed to solve the d-variate problem to within ε. Here, by one information operation we mean computing one function value for multivariate integration, and one linear continuous functional for multivariate approximation. For multivariate integration it is shown in [10] that

$$n(\varepsilon, d) \geq 2^d \quad \text{for all } \varepsilon < 1.$$

It should be added that $n(\varepsilon, d) = 0$ for all $\varepsilon \geq 1$. Hence, we see that $n(\cdot, d)$ is a discontinuous function at 1.

For multivariate approximation it is shown in [6] that for every non-negative integer r there exists a number $C_r > 1$ such that

$$n(\varepsilon, d) \geq [C_r]^d (1 - \varepsilon^2) \quad \text{for all } \varepsilon < 1.$$

The essence of these two lower bounds is that complexity depends exponentially on d. We have to wait exponentially long in d before we can make use of the excellent convergence rate. For large d, this is really bad news. For example, take $d = 100$, which is quite small by today's standards. Then for multivariate integration we have to compute at least 2^{100} function values even if we want to guarantee the error close to 1, say, 0.9999.

The exponential dependence on d is called the *curse of dimensionality*, after Bellman who coined this phrase back in 1957. Hence our two problems suffer from the curse of dimensionality, this holds independently of the smoothness parameters α and r. This curse holds even though asymptotically we have excellent rates of convergence for large α and r.

This means that many standard positive results do not necessarily imply good complexity bounds. We need a more thorough analysis if we want to establish tight complexity bounds; this usually requires a new proof technique. Furthermore, even for problems studied thoroughly in the past, we need to revisit them if we want to find sharp bounds on their complexity.

It is natural to ask what to do if we establish that complexity is huge, or if the multivariate problem suffers from the curse of dimensionality. Since this is a complexity result it is impossible to find a better algorithm. The problem is simply too difficult. The only way to vanquish the curse of dimensionality is either to change the problem or to switch to a more lenient way of defining the error of an algorithm. For instance, we may shrink the class of functions by assuming additional properties of functions. In particular, we may have an application for which functions of d variables may be well approximated by sums of functions of only a few variables, or functions for which the dependence on successive variables or groups of variables vary significantly. This leads to the study of *weighted* spaces. The main point is to find necessary and sufficient conditions on the weights to vanquish the curse of dimensionality and to guarantee that complexity is, for instance, a polynomial function of ε^{-1} and d. This subject is beyond the scope of this paper and we only illustrate one of the many aspects of weighted spaces in the last section. The reader who is interested in how the curse of dimensionality can be indeed vanquished is referred to our book *Tractability of Multivariate Problems* [6].

2 Matrix multiplication

Let $A = (a_{i,j})_{i,j=1}^n$ and $B = (b_{i,j})_{i,j=1}^n$ be two arbitrary $n \times n$ matrices with real coefficients $a_{i,j}$ and $b_{i,j}$. We want to compute

$$C = AB, \quad \text{so that} \quad c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j} \text{ for } i, j, = 1, 2, \ldots, n.$$

It is probably fair to say that matrix multiplication is not practically important per se. However, it is known that matrix multiplication is equivalent to the problem of solving linear equations $Ax = b$ which, in turn, is a very important and practical problem. So if we know how to quickly compute the matrix C then we also know how to quickly compute $x = A^{-1}b$.

Let comp(n) denote the *minimal* number of arithmetic operations needed to compute C for arbitrary $n \times n$ real matrices A and B. By arithmetic operations we mean additions, subtractions, multiplications, division and comparisons of real numbers. This point is important, since we now assume the *real number model* which is usually used for solving continuous problems, and this is indeed the case for practically all scientific and numerical computations. It is worth to stress that in *discrete* complexity, the Turing model of computation is usually used, in which we count operations on bits instead of operations on real numbers. So, multiplication of two real numbers has unit cost in the real number model, but is usually infeasible in the Turing machine model.

We also stress that complexity is defined as the minimal cost of a solution, and not as a cost of a specific algorithm. This distinction is crucial since in many cases, "complexity" is used by many people as a synonym of the word "cost". For us, complexity is a property of a problem and we are seeking an algorithm whose cost is equal to the complexity or, more likely, whose cost is as close to the complexity as possible.

From a mathematical point of view, to know the complexity of a problem means that we can prove two bounds. First, we can prove that the cost of *any* possible algorithm that solves the problem is bounded from *below* by the complexity. Second, we know a specific algorithm that solves the problem whose cost is bounded from *above* by the complexity. Not surprisingly, in most cases, we only know lower and upper bounds on the complexity. If they are equal, we are done; if they are sharp we are almost done. If the bounds are not sharp (the typical case), we only have some ideas about complexity.

Matrix multiplication is one such problem. As already mentioned in the introduction, comp(n) must be at least equal to n^2 since the matrix C depends on $2n^2$ real inputs, and all of them are important. Since we are using operations involving two arguments, if we perform k such operations we may use at most $2k$ inputs. Therefore $2k \geq 2n^2$ and $k \geq n^2$. So we have a trivial lower bound,

$$\text{comp}(n) \geq n^2.$$

Interestingly enough, no one so far was able to prove that complexity depends on a higher power of n. Hence the case that complexity is of order n^2 has not been ruled out.

What do we know about upper bounds on comp(n)? The first bound is trivial. We just compute $c_{i,j}$ by n multiplications and $n-1$ additions, and obtain the whole matrix C at cost of $(2n-1)n^2 \leq 2n^3$. Hence

$$\text{comp}(n) \leq 2n^3.$$

Up to 1969, many people believed that comp(n) is indeed proportional to n^3. It was a big surprise when Strassen [11] proved in 1969 that

$$\text{comp}(n) = \mathcal{O}\left(n^{\log_2 7}\right) \quad \text{with} \quad \log_2 7 \approx 2.81.$$

His algorithm is based on recursion, and the basic reduction comes from the fact that he showed how to multiply two 2×2 matrices using only 7 multiplications. The usual way requires 8 multiplications and leads to $\mathcal{O}(n^{\log_2 8}) = \mathcal{O}(n^3)$ cost.

The title of Strassen's paper is "Gaussian elimination is not optimal" and refers to the fact that Gaussian elimination is the most often used algorithm for solving $Ax = b$ at cost $\mathcal{O}(n^3)$. However, using Strassen's algorithm, $Ax = b$ can be solved at cost $\mathcal{O}(n^{\log_2 7})$.

Strassen's paper is today regarded as the beginning of a new subfield of complexity which is called *algebraic complexity*. Many people tried to improve Strassen's result. The first approach was to reduce 7 multiplications needed for two 2×2 matrices. This failed, but other approaches came along. There is a long trail of successive improvements of Strassen's algorithm; the last step was done some twenty years ago by Coppersmith and Winograd [4], who proved that

$$\text{comp}(n) = \mathcal{O}\left(n^{2.388}\right).$$

We wish to add that the factor in the last big $\mathcal{O}$ notation is enormous. This means that the last bound is better than, say, the bound $2n^3$ only if n is very large. Such large n *do* occur in computational practice but then the matrices A and B have extra properties. For instance, they are sparse or generated by a

few parameters; this additional structure allows us to usually find much better algorithms. This is especially the case when we switch to the equivalent problem of solving $Ax = b$. This is why, despite of the decrease of the exponent of n from 3 to 2.388, Gaussian elimination is still used for solving $Ax = b$ if n is moderate, say, $n \le 100$.

We stress that complexity is defined as the minimal cost, not as the minimal exponent of n. Hence, if one day an algorithm with cost equal to the complexity is found, then it cannot lose against Gaussian elimination for *all* n. Does this mean that Gaussian elimination will be then abandoned? Not necessarily, since we also need to match the stability property of Gaussian elimination. This is because all algorithms run in the floating point arithmetic and there are subtle differences between the real number model and floating point arithmetic. The new algorithm must be resistant in a certain sense to rounding errors before it can be recommended for computational use. So there seems to be still a long way to go before we can forget about Gaussian elimination for moderate n.

We stop here and recommend the readers interested in learning more on this subject to turn to the books of Bürgisser, Clausen and Shokrollahi [3] on algebraic complexity, and Pan [7], and Bini and Pan [2] on fast matrix multiplication.

3 Multivariate integration of smooth periodic functions

Consider periodic complex valued functions f defined on the d dimensional unit cube $[0,1]^d$. For a given real parameter $\alpha > 1$, consider the class of functions defined as

$$F_{d,\alpha} = \left\{ f : [0,1]^d \to \mathbb{C} \mid |\hat{f}(h)| \le \prod_{j=1}^d \max(1, |h_j|)^{-\alpha} \text{ for all } h \in \mathbb{Z}^d \right\}.$$

Here $h = [h_1, h_2, \ldots, h_d]$ with integers h_j, and $\hat{f}(h)$ denotes the Fourier coefficient,

$$\hat{f}(h) = \int_{[0,1]^d} f(x)\, \exp(-2\pi\, \mathrm{i}\, h \cdot x)\, \mathrm{d}x,$$

with $\mathrm{i} = \sqrt{-1}$, and $x = [x_1, x_2, \ldots, x_d]$ for $x_j \in [0,1]$, and the inner product $h \cdot x = h_1 x_1 + h_2 x_2 + \cdots + h_d x_d$.

For functions in $F_{d,\alpha}$ we control the decay of the Fourier coefficients. Since $\alpha > 1$, functions from $F_{d,\alpha}$ are continuous. For large α, the class $F_{d,\alpha}$ consists of

smooth functions. Indeed, it is easy to check that if $f \in F_{d,\alpha}$ then f is k-times differentiable with respect to all variables for all $k < \alpha - 1$.

Multivariate integration is defined as approximation of

$$I_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x \quad \text{for all } f \in F_{d,\alpha}.$$

We assume that we can compute function values at any point from $[0,1]^d$, and approximate $I_d(f)$ by algorithms $A_{n,d}$ that use at most n function values,

$$A_{n,d}(f) = \phi_n(f(t_1), f(t_2), \ldots, f(t_n))$$

for some points t_j and a scalar mapping ϕ_n. The points t_j can be chosen adaptively, that is, t_j can depend in an arbitrary way on the previously computed values $f(x_1), f(x_2), \ldots, f(x_{j-1})$; the mapping ϕ_n can also be arbitrary.

The error of the algorithm $A_{n,d}$ is defined in the worst case setting as

$$e(A_{n,d}) = \sup_{f \in F_{d,\alpha}} |I_d(f) - A_{n,d}(f)|.$$

Let $e_{n,d}$ denote the minimal worst case error that can be achieved by using n function values, i.e.,

$$e_{n,d} = \inf_{A_{n,d}} e(A_{n,d}).$$

It turns out that the last infimum is attained for linear algorithms and for nonadaptive choice of points t_j. That is, it is enough to consider algorithms $A_{n,d}$ of the form

$$A_{n,d}(f) = \sum_{j=1}^{n} a_j f(t_j)$$

for some complex numbers a_j and some $t_1, t_2, \ldots, t_n$ chosen independently of f. Then

$$e_{n,d} = \inf_{a_j, t_j} \sup_{f \in F_{d,\alpha}} \left| I_d(f) - \sum_{j=1}^{n} a_j f(t_j) \right|.$$

Optimality of linear algorithms was proved by Smolyak in his PhD thesis in 1965, whereas optimality of non-adaption was proved by Bakhvalov in 1971. Both results may be found in [1]. These results hold for more general spaces and more general linear functionals. They were also generalized in many aspects. The reader is referred to Chapter 4 of [6] for more information.

From Niederreiter [5] and Sloan and Joe [9], we know that

$$e_{n,d} = \mathcal{O}(n^{-p}) \quad \text{for all } p < \alpha.$$

Hence, for large α the rate of convergence is excellent.

It is natural to ask "How long do we have to wait to enjoy this excellent rate of convergence?" A surprising negative result was proved in [10] using an elementary argument. Namely, we have

$$e_{n,d} = 1 \quad \text{for all } n = 0, 1, \ldots, 2^d - 1. \tag{1}$$

The proof that $e_{n,d} = 1$ is short; we present it here to show the reader how such a complexity result can be obtained. First of all, take the most trivial algorithm, $A_{n,d} = 0$. Its worst case error is just

$$\sup_{f \in F_{d,\alpha}} |I_d(f)| = \sup_{f \in F_{d,\alpha}} |\hat{f}(0)| = 1.$$

Hence, $e_{n,d} \le 1$ and we need to prove that as long as $n < 2^d$, all algorithms behave as badly as the zero algorithm. Take an arbitrary algorithm $A_{n,d}(f) = \phi_n(f(t_1), f(t_2), \ldots, f(t_n))$ for some (perhaps non-linear) mapping ϕ_n and some (perhaps adaptively) chosen points t_j. We will construct a function $f \in F_{d,\alpha}$ for which

$$f(t_j) = 0 \quad \text{for } j = 1, 2, \ldots, n \quad \text{but} \quad |I_d(f)| = 1.$$

More precisely, if points t_j are given adaptively, we take the first point t_1 that is independent of functions from $F_{d,\alpha}$, and construct f such that $f(t_1) = 0$. This implies a choice of the second point t_2 for which we construct f for which $f(t_1) = f(t_2) = 0$, and so on. That is, knowing that $f(t_1) = f(t_2) = \cdots = f(t_k) = 0$, the $(k+1)$st point t_{k+1} is chosen and we again construct f for which additionally $f(t_{k+1}) = 0$. This is done for $k = 1, 2, \ldots, n-1$.

We first choose a trigonometric polynomial of the form

$$\theta(x) \sum_{h \in \{0,1\}^d} a_h \, e^{2\pi \mathrm{i}\, h \cdot x}$$

with another trigonometric polynomial θ to be specified later, and complex coefficients a_h that are a non-trivial solution of the homogeneous linear system

$$\sum_{h \in \{0,1\}^d} a_h \, e^{2\pi \mathrm{i}\, h \cdot t_j} = 0.$$

Here, we need the assumption that $n < 2^d$. Indeed, we have 2^d unknowns a_h and n homogeneous linear equations; hence for $n < 2^d$, a non-zero solution exists. The non-zero solution a_h can be normalized and we choose the normalization such that

$$\max_{h\in\{0,1\}^d} |a_h| = a_{h^*} = 1,$$

for some $h^* \in \{0,1\}^d$. We now define $\theta(x) = e^{-2\pi \mathrm{i} h^*\cdot x}$. Our function f is given as

$$f(x) = c \sum_{h\in\{0,1\}^d} a_h \, e^{2\pi \mathrm{i}\,(h-h^*)\cdot x},$$

where $c = 1$ if the real part of $\phi(0,0,\ldots,0)$ is non-positive, and $c = -1$ if the real part of $\phi(0,0,\ldots,0)$ is positive.

We now show that f belongs to $F_{d,\alpha}$. Indeed, observe that f is a trigonometric polynomial with

$$h_j - h_j^* \in \{-1,0,1\} \quad \text{for all } j = 1,2,\ldots,d \ \text{ and } \ h \in \{0,1\}^d.$$

This implies that $\max(1, |h_j - h_j^*|) = 1$ and

$$\prod_{j=1}^{d} \max(1, |h_j - h_j^*|) = 1 \quad \text{for all } h \in \{0,1\}^d.$$

We have $|\hat{f}(h-h^*)| = |a_h| \le 1$ for all $h \in \{0,1\}^d$, and $\hat{f}(h-h^*) = 0$ for all $h \notin \{0,1\}^d$. Hence, $|\hat{f}(h)| \le \prod_{j=1}^d \max(1,|h_j|)^{-\alpha}$ for all $h \in \mathbb{Z}^d$. This means that $f \in F_{d,\alpha}$, as claimed.

Clearly, $f(t_j) = 0$ for all $j = 1,2,\ldots,n$ and therefore

$$A_{n,d}(f) = \phi(0,0,\ldots,0).$$

Furthermore, $I_d(f) = \hat{f}(0) = c\, a_{h^*} = c$, and

$$|I_d(f) - A_{n,d}(f)| = |c - \phi(0,0,\ldots,0)| \ge |c - \Re\phi(0,0,\ldots,0)| \ge |c| = 1.$$

Hence, the worst case error of $A_{n,d}$ is at least 1, which completes the proof.

The essence of this result is that the zero algorithm is best if we use less than 2^d function values. Hence, we have to wait exponentially long in d to see the rate of convergence in action. Furthermore, for $n = 2^d$ we have

$$e_{2^d,d} \le \frac{d}{2^{\alpha-1}}(1 + o(1)) \quad \text{as } \alpha \to \infty.$$

This means that for large α, nothing happens as long as $n < 2^d$ and the error stays as 1, whereas for $n = 2^d$ the worst case error is exponentially small in α and is suddenly almost zero. This is indeed very irregular behavior of the minimal worst case errors.

We now turn to complexity, i.e., the minimal cost of solving multivariate integration to within ε. Obviously, for $\varepsilon \geq 1$ the problem is trivial, since it is solved by the zero algorithm with cost zero. Without loss of generality we thus assume that $\varepsilon \in (0, 1)$.

Since linear algorithms are optimal, it is clear that the minimal cost will be achieved by taking the minimal n for which the worst case error is at most ε and by computing n function values, and performing at most n multiplications and $n - 1$ additions. Since the cost of computing one function value is usually much larger then the cost of one multiplication or addition, we conclude that complexity is roughly equal to the minimal number of function values needed to solve the problem,

$$n(\varepsilon, d) = \min\{\, n \mid e_{n,d} \leq \varepsilon \,\}.$$

From the bound (1) on $e_{n,d}$ we immediately conclude that

$$n(\varepsilon, d) \geq 2^d \quad \text{for all } \varepsilon \in (0, 1).$$

So we have exponential dependence on d, and multivariate integration of periodic and smooth functions suffers from the curse of dimensionality. Furthermore this holds for all α, that is, even if α is arbitrarily large and when the rate of convergence is excellent.

We finish this section by saying that there are more such examples where the excellent rate of convergence is present only after an exponential delay in d. That is, there are sufficiently smooth problems that are asymptotically trivial although their worst case complexity suffers from the curse of dimensionality. The reader is referred to [6] for more such examples.

4 Multivariate approximation

Consider now the class

$$F_{d,r} = C_0^{r,r,\ldots,r}([0, 1]^d)$$

of real functions defined on $[0,1]^d$ that satisfy the boundary conditions and are r times continuously differentiable with respect to all variables. The boundary conditions mean that all derivatives up to order r are zero if the argument of a function has at least one zero component. For example, if $r = 0$ then $f(x) = 0$ if some component of x is zero. The space $F_{d,r}$ is equipped with the norm

$$\|f\|_{F_{d,r}} = \sup_{x\in[0,1]^d} \left|f^{(r,r,\dots,r)}(x)\right|,$$

where $f^{(r,r,\dots,r)}$ denotes the r times differentiation of f with respect to all variables.

Clearly, $F_{d,r}$ is a separable Banach space. We equip $F_{d,r}$ with the r-folded Wiener sheet measure $w_{d,r}$, which is the classical Wiener sheet measure placed on partial derivatives of order r. The measure $w_{d,r}$ is a zero-mean Gaussian measure whose covariance kernel $K_{d,r}$ is given by

$$K_{d,r}(x,y) = \prod_{j=1}^{d} \int_0^1 \frac{(x_j-u)_+^r}{r!}\,\frac{(y_j-u)_+^r}{r!}\,\mathrm{d}u \quad \text{for all } x,y\in[0,1]^d.$$

For $r = 0$, we have $K_{d,0}(x,y) = \prod_{j=1}^{d} \min(x_j, y_j)$, which is the usual Wiener sheet measure.

Multivariate approximation is defined as approximation of an embedding $\mathrm{APP}_d : F_{d,r} \to L_2 = L_2([0,1]^d)$ given by

$$\mathrm{APP}_d(f) = f \quad \text{for all } f \in F_{d,r}.$$

We assume that we can compute arbitrary linear continuous functionals $L(f)$ for $L \in F_{d,r}^*$. We approximate $\mathrm{APP}_d(f)$ by algorithms $A_{n,d} : F_{d,r} \to L_2$ of the form

$$A_{n,d}(f) = \phi_n(L_1(f), L_2(f), \dots, L_n(f))$$

for some $L_j \in F_{d,r}^*$ and some mapping ϕ_n. As in the previous section we may use adaption.

The error of the algorithm $A_{n,d}$ is defined in the average case setting as

$$e(A_{n,d}) = \left[\int_{F_{d,r}} \|\mathrm{APP}_d(f) - A_{n,d}(f)\|_{L_2}^2\, w_{d,r}(\mathrm{d}f)\right]^{1/2}.$$

For simplicity we assume that $A_{n,d}$ is measurable but this assumption can be removed, see e.g., Chapter 4 of [6].

Let $e_{n,d}$ denote the minimal average case error that can be achieved by using n linear continuous functionals, i.e.,

$$e_{n,d} = \inf_{A_{n,d}} e(A_{n,d}).$$

It turns out, see the results of Wasilkowski [12] and Papageorgiou and Wasilkowski [8], that

$$e_{n,d} = \Big[\sum_{j=n+1}^{\infty} \lambda_{d,j} \Big]^{1/2},$$

where $\lambda_{d,j}$ are the ordered eigenvalues of the compact operator $W_d : F_{d,r} \to F_{d,r}$ given by

$$W_{d,r} f = \int_{[0,1]^d} K_{d,r}(\cdot, x)\, f(x)\, \mathrm{d}x \quad \text{for all } f \in F_{d,r}. \tag{2}$$

The eigenvalues $\lambda_{d,j}$ are asymptotically known and it allowed Papageorgiou and Wasilkowski to show that

$$e_{n,d} = \mathcal{O}(n^{-p}) \quad \text{for all } p < r + 1/2.$$

So the order of convergence is excellent, again for large r.

We now address average case complexity, working with the normalized error criterion, see [6] pp. 58-62. More precisely, for $n = 0$ we know that the minimal error is achieved by the zero algorithm and is equal to

$$e_{0,d} = \Big[\sum_{j=1}^{\infty} \lambda_j \Big]^{1/2} = \Big[\int_{F_{d,r}} \|f\|_{L_2}^2 \, w_{d,r}(\mathrm{d}f) \Big]^{1/2}$$
$$= \frac{1}{[(r!)^2(2r+1)(2r+2)]^{d/2}}.$$

The quantity $e_{0,d}$ is called the initial error since it is determined by the definition of the problem without sampling f. For the normalized error criterion we want to reduce the initial error by a factor of ε. Let

$$n(\varepsilon, d) = \min \big\{ n \mid e_{n,d} \le \varepsilon\, e_{0,d} \big\}$$

be the minimal number of linear continuous functionals needed for solving the problem. From general results for the average case setting, $n(\varepsilon, d)$ is practically the same as the average case complexity. We obviously have

$$n(\varepsilon, d) = \min \big\{ n \mid \textstyle\sum_{j=n+1}^{\infty} \lambda_{d,j} \le \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j} \big\},$$

and

$$n(\varepsilon, d) = \mathcal{O}\Big(\varepsilon^{-1/(r+1/2)-\delta}\Big) \quad \text{for all } \delta > 0.$$

Hence, the exponent of ε^{-1} is roughly at most 2 for $r = 0$, whereas for large r, it is small. Again, it is tempting to say that the problem is easy, especially for large r.

However, note that $e_{n(\varepsilon,d)} \le \varepsilon\, e_{0,d}$ and

$$e^2_{n(\varepsilon,d)} = \sum_{j=1}^{\infty} \lambda_{d,j} - \sum_{j=1}^{n(\varepsilon,d)} \lambda_{d,j} \ge e^2_{0,d} - n(\varepsilon, d)\lambda_{1,d}.$$

This yields

$$n(\varepsilon, d) \ge (1 - \varepsilon^2)\frac{e^2_{0,d}}{\lambda_{d,1}}.$$

Since $K_{d,r}$ is given as a product of univariate kernels, the eigenvalues of the operator W_d are products of univariate eigenvalues for the operator W_1. Therefore we have

$$e^2_{0,d} = \Big(\sum_{j=1}^{\infty} \lambda_{1,j}\Big)^d \quad \text{and} \quad \lambda_{d,1} = \lambda^d_{1,1}.$$

Hence,

$$n(\varepsilon, d) \ge (1 - \varepsilon^2)\Big(1 + \sum_{j=2}^{\infty} \frac{\lambda_{1,j}}{\lambda_{1,1}}\Big)^d.$$

Since all eigenvalues $\lambda_{1,j}$ are positive, we have an exponential lower bound on $n(\varepsilon, d)$, and multivariate approximation suffers from the curse of dimensionality independently on how large is r. This shows that also in the average case setting we may have an excellent rate of convergence and the curse of dimensionality.

5 Weighted multivariate approximation

We briefly discuss how the curse of dimensionality of multivariate approximation studied in the previous section can be vanquished by introducing *weights*.

Define the measure $\nu_{d,r} = w_{d,r}\mathrm{APP}_d^{-1}$, where $w_{d,r}$ and APP_d are from the previous section. Then $\nu_{d,r}$ is a zero-mean Gaussian measure on the space $L_2 = L_2([0,1]^d)$ whose covariance operator is the operator $W_{d,r}$ given by (2).

As before, let $\{\lambda_{d,j}\}_{j\in\mathbb{N}^d}$ be the sequence of the eigenvalues of the operator $W_{d,r}$. For $j = [j_1, j_2, \ldots, j_d]$ we have

$$\lambda_{d,j} = \prod_{k=1}^{d} \lambda_{j_k}$$

for the univariate eigenvalues $\lambda_{j_k} = \lambda_{1,j_k}$.

Let $\{\eta_{d,j}\}_{j\in\mathbb{N}^d}$ denote the sequence of the corresponding orthonormal eigenfunctions of $W_{d,r}$, so that

$$W_{d,r}\eta_{d,j} = \lambda_{d,j}\eta_{d,j} \quad \text{for all } j \in \mathbb{N}^d.$$

We now consider a weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ for all $d \in \mathbb{N}$ and subsets $\mathfrak{u}$ of the index set $[d] := \{1, 2, \ldots, d\}$. We assume that $\gamma_{d,\mathfrak{u}} \geq 0$, and that for each d at least one $\gamma_{d,\mathfrak{u}}$ is positive for a non-empty $\mathfrak{u}$.

For $j \in \mathbb{N}^d$, define $\mathfrak{u}(j) = \{k \mid j_k \geq 2\}$ and consider the *weighted* eigenvalues

$$\lambda_{d,\gamma,j} = \gamma_{d,\mathfrak{u}(j)}\lambda_{d,j} = \gamma_{d,\mathfrak{u}(j)}\lambda_1^{d-|\mathfrak{u}(j)|} \prod_{k\in\mathfrak{u}(j)} \lambda_{j_k}.$$

Weighted multivariate approximation is defined as before, the only change being that the correlation operator $W_{d,\gamma,r}$ of the measure $\nu_{d,r}$ has now the eigenpairs $(\lambda_{d,\gamma,j}, \eta_{d,j})$, i.e.,

$$W_{d,\gamma,j}\eta_{d,j} = \lambda_{d,\gamma,j}\eta_{d,j} \quad \text{for all } j \in \mathbb{N}^d.$$

By considering different weights $\gamma_{d,\mathfrak{u}}$, we can model different a priori knowledge about the distribution of

$$\mathrm{APP}_d(f) = f = \sum_{j\in\mathbb{N}^d} L_j(f)\,\eta_{d,j},$$

where $L_j(f) = \langle f, \eta_{d,j}\rangle_{L_2}$. Note that

$$\int_{F_{d,r}} L_i(f)L_j(f)\,w_{d,r}(\mathrm{d}f) = \lambda_{d,\gamma,j}\delta_{i,j} \quad \text{for all } i, j \in \mathbb{N}^d.$$

For instance, for *finite-order weights*, when $\gamma_{d,\mathfrak{u}} = 0$ for all $|\mathfrak{u}| > \omega^*$ for some integer ω^* independent of d, we know a priori that f is orthogonal (with probability 1) to all $\eta_{d,j}$ whenever $|\mathfrak{u}(j)| > \omega^*$. This means that f is (again with probability one) a linear combination of functions $\eta_{d,j}$ that depend on at most ω^* variables. Furthermore, for *product weights*, when $\gamma_{d,\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_{d,j}$ for some

$\gamma_{d,1} \geq \gamma_{d,2} \geq \cdots \geq \gamma_{d,d}$, the importance of successive variables is decreasing. Obviously, if all $\gamma_{d,\mathfrak{u}} = 1$ then we have the unweighted case studied before.

It is now natural to ask what are necessary and sufficient conditions on the weight sequence γ for which the curse of dimensionality is not present. The curse is certainly not present if we can prove that the average case complexity of weighted multivariate approximation is polynomial in ε^{-1} and d. For simplicity we restrict ourselves to this case, referring the reader to [6] for a more general situation.

It is easy to check that the initial error is now

$$e_{0,\gamma,d} = \Big[\sum_{j \in \mathbb{N}^d} \lambda_{d,\gamma,j} \Big]^{1/2} = \Big[\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \lambda_1^{d-|\mathfrak{u}|} \Big(\sum_{j=2}^{\infty} \lambda_j \Big)^{|\mathfrak{u}|} \Big]^{1/2}.$$

As always, the nth minimal average case error $e_{n,\gamma,d}$ is given as the square root of the sum of the eigenvalues $\lambda_{d,\gamma,j}$, omitting the n largest. This means that we have to order the eigenvalues $\{\lambda_{d,\gamma,j}\}_{j \in \mathbb{N}^d}$.

Formally let $\{\lambda_{d,\gamma,k}\}_{k \in \mathbb{N}} = \{\lambda_{d,\gamma,j}\}_{j \in \mathbb{N}^d}$ be the ordered sequence

$$\lambda_{d,\gamma,1} \geq \lambda_{d,\gamma,2} \geq \cdots \geq \lambda_{d,\gamma,k} \geq \cdots .$$

Then

$$e_{n,\gamma,d} = \Big[\sum_{k=n+1}^{\infty} \lambda_{d,\gamma,k} \Big]^{1/2}.$$

To estimate $e_{n,\gamma,d}$ we need to know how the ordered eigenvalues $\lambda_{d,\gamma,k}$ behave. One useful technical trick is to consider powers of the eigenvalues. More precisely, for $\tau > 0$ we have

$$M_\tau := \sum_{k=1}^{\infty} \lambda_{d,\gamma,k}^{\tau} = \sum_{j \in \mathbb{N}^d} \lambda_{d,\gamma,j}^{\tau} = \sum_{u \subseteq [d]} \gamma_{d,\mathfrak{u}}^{\tau} \lambda_1^{(d-|\mathfrak{u}|)\tau} \Big(\sum_{j=2}^{\infty} \lambda_j^{\tau} \Big)^{|\mathfrak{u}|}.$$

It is known that $\lambda_j = \Theta(j^{-2(r+1)})$. This implies that for $\tau > 1/(2(r+1))$ we have $M_\tau < \infty$. For finite M_τ, we use $k \lambda_{d,\gamma,k}^{\tau} \leq M_\tau$, which implies that

$$\lambda_{d,\gamma,k} \leq \frac{M_\tau^{1/\tau}}{k^{1/\tau}} \quad \text{for all } k \in \mathbb{N}.$$

Since we have to sum up the eigenvalues $\lambda_{d,\gamma,k}$ and their sum should be finite, we must assume that $\tau < 1$. Hence, we can use τ from the interval $(1/(2(r+1)), 1)$. It turns out that this leads to sharp estimates on the $e_{n,\gamma,d}$.

More precisely, let $n(\varepsilon, \gamma, d)$ be the minimal number of linear continuous functionals needed to reduce the initial error by a factor ε, i.e.,

$$n(\varepsilon, \gamma, d) = \min\{\, n \mid e_{n,\gamma,d} \le \varepsilon\, e_{0,\gamma,d} \,\}.$$

We say that weighted multivariate approximation is *polynomially tractable* iff there are non-negative numbers C, p and q such that

$$n(\varepsilon, \gamma, d) \le C\, \varepsilon^{-p}\, d^{q} \quad \text{for all } \varepsilon \in (0,1),\ d \in \mathbb{N}.$$

If $q = 0$ in the formula above, then we say that weighted multivariate approximation is *strongly polynomially tractable* and the infimum of p satisfying the formula above with $q = 0$ is called the *exponent of strong polynomial tractability*.

We are ready to present part of Theorem 6.8 from [6], specialized to our case. In what follows we let

$$\alpha_\tau = \sum_{j=2}^{\infty} \left(\frac{\lambda_j}{\lambda_1}\right)^{\tau} \quad \text{and} \quad p_\lambda = \frac{1}{2(r+1)},$$

and we use *WMA* as an abbreviation of weighted multivariate approximation. Then

- WMA is polynomially tractable iff there exist $q_1 \ge 0$ and $\tau \in (p_\lambda, 1)$ such that

$$C_1 := \sup_d \frac{\left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{\tau}\, \alpha_\tau^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}\, \alpha_1^{|\mathfrak{u}|}}\, d^{-q_1} < \infty. \tag{3}$$

- If (3) holds then

$$n(\varepsilon, \gamma, d) \le \left(\left(\frac{\tau\, C_1}{1-\tau}\right)^{\tau/(1-\tau)} + 1\right) d^{\,q_1\, \tau/(1-\tau)}\, \varepsilon^{-2\tau/(1-\tau)}$$

  for all $\varepsilon \in (0,1]$ and $d \in \mathbb{N}$.

- If WMA is polynomially tractable, so that $n(\varepsilon, d) \le C\, d^{\,q}\, \varepsilon^{-p}$ for some positive C and p with $q \ge 0$, then $p > 2/(2r+1)$ and (3) holds for $\tau \in ((1+2/p)^{-1}, 1)$ with $q_1 = q \max(1/\tau, 2/p)$ and

$$C_1 \le \left(2C + 1 + \left(2(4C)^{2/p}\, \zeta(\tau(1+2/p))^{1/\tau}\right)^{\tau}\right)^{1/\tau}.$$

- WMA is strongly polynomially tractable iff (3) holds with $q_1 = 0$. The exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \inf\left\{ \tfrac{2\tau}{1-\tau} \mid \tau \in (p_\lambda, 1) \text{ and satisfies (3) with } q_1 = 0 \right\}.$$

- For product weights $\gamma_{d,\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_{d,j}$ with $\gamma_{d,j+1} \le \gamma_{d,j}$ for $j \in [d-1]$ and $\sup_d \gamma_{d,1} < \infty$, we have the following results:
  - WMA is polynomially tractable iff there exists $\tau \in (p_\lambda, 1)$ such that

$$\limsup_{d\to\infty} \frac{\sum_{j=1}^d \gamma_{d,j}^\tau}{\ln d} < \infty. \tag{4}$$

    If (4) holds then for any

$$q > \frac{1}{\tau}\left(\sum_{j=2}^\infty \left(\frac{\lambda_j}{\lambda_2}\right)^\tau\right) \limsup_{d\to\infty} \frac{\sum_{j=1}^d \gamma_{d,j}^\tau}{\ln d}$$

    there exists a positive C such that

$$n(\varepsilon, d) \le C d^{q\,\tau/(1-\tau)}\, \varepsilon^{-2\tau/(1-\tau)} \quad \text{for all } \varepsilon \in (0,1] \text{ and } d \ge 1.$$

  - WMA is strongly polynomially tractable iff $p_\gamma < 1$, where

$$p_\gamma = \inf\left\{\tau \ge 0 \mid \limsup_{d\to\infty} \textstyle\sum_{j=1}^d \gamma_{d,j}^\tau < \infty\right\}.$$

    The exponent of strong tractability is

$$p^{\text{str-avg}} = \frac{2\max\left(p_\lambda, p_\gamma\right)}{1-\max\left(p_\lambda, p_\gamma\right)}.$$

- For finite-order weights $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > \omega^*$, we have
  - WMA is polynomially tractable. Then for any $\tau \in (p_\lambda, 1)$, we have

$$n(\varepsilon, d) \le C \left|\left\{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \ne 0\right\}\right| \varepsilon^{-2\tau/(1-\tau)},$$

    where

$$C = \left(\left(\frac{\alpha_\tau^{1/\tau}}{\alpha_1}\right)^{\omega^*} \frac{\tau}{1-\tau}\right)^{\tau/(1-\tau)} + 1.$$

    Hence, for arbitrary finite-order weights we have

$$n(\varepsilon, d) \le 2\, C\, d^{\,\omega^*}\, \varepsilon^{-2\tau/(1-\tau)}.$$

– WMA is strongly polynomially tractable iff there exists $\tau \in (p_\lambda, 1)$ such that

$$\sup_d \frac{\left(\sum_{\mathfrak{u}\subseteq[d],\,|\mathfrak{u}|\le\omega^*} \gamma_{d,\mathfrak{u}}^{\tau}\alpha_\tau^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u}\subseteq[d],\,|\mathfrak{u}|\le\omega^*} \gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}} < \infty. \tag{5}$$

The exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \inf\left\{\tfrac{2\tau}{1-\tau} \mid \tau \in (p_\lambda, 1) \text{ and satisfies (5)}\right\}.$$

The essence of this theorem is that we can vanquish the curse of dimensionality and even obtain polynomial or strongly polynomial tractability for sufficiently decaying weights. In particular, for product weights we obtain polynomial tractability if the sum of some power of the weights grows no faster than $\ln d$, and obtain strong polynomial tractability if this sum is uniformly bounded in d. For arbitrary finite-order weights we always have polynomial tractability.

Such conditions on weights are typical for tractability of other multivariate problems; they also hold in different settings, including the worst case setting, see again [6]. In this way, the curse of dimensionality, which is often present for the unweighted case, can be vanquished.

References

[1] N. S. Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions. *USSR Comput. Maths. Math. Phys.* **11** (1971), 244–249. 47

[2] D. Bini and V. Pan, *Polynomial and matrix computations*. Progr. Theoret. Comput. Sci., Birkhäuser Verlag, Boston, MA, 1994. 46

[3] P. Bürgisser, M. Clausen and M. A. Shokrollahi, *Algebraic complexity theory*. Grundlehren Math. Wiss. 315, Springer-Verlag, Berlin, 1997. 46

[4] D. Coppersmith and S. Winograd, Matrix multiplication via arithmetic progression. *J. Symbolic Comput.* **9** (1990), 251–280. 41, 45

[5] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*. SIAM, Philadelphia, 1992. 48

[6] E. Novak and H. Woźniakowski, *Tractability of multivariate problems.* Volume I: Linear information, EMS Tracts Math. 6, European Mathematical Society Publishing House, Zürich, 2008. 42, 43, 47, 50, 51, 52, 55, 56, 58

[7] V. Ya. Pan, *How to multiply matrices faster.* Lecture Notes in Comput. Sci. 179, Springer-Verlag, Berlin, 1984. 46

[8] A. Papageorgiou and G. W. Wasilkowski, On the average complexity of multivariate problems. *J. of Complexity* **6** (1990), 1–23. 42, 52

[9] I. H. Sloan and S. Joe, *Lattice methods for multiple integration*. Clarendon Press, Oxford, 1994. 48

[10] I. H. Sloan and H. Woźniakowski, An intractability result for multiple integration. *Math. Comp.* **66** (1997), 1119–1124. 41, 42, 48

[11] V. Strassen, Gaussian elimination is not optimal. *Numer. Math.* **13** (1969), 354–356. 45

[12] G. W. Wasilkowski, Information of varying cardinality. *J. of Complexity* **2** (1986), 204–228. 52

A brief history of information-based complexity

Joseph F. Traub

This talk is titled a brief history of information-based complexity. Like all memoirs it reflects my personal view of the history of the field. Furthermore, it has a Carnegie Mellon and Columbia University slant. If one of my colleagues were to write a history I am sure it would differ from this one. I do hope it captures the essence of the field and I apologize to other researchers for any omissions.

I shall begin my history in 1972 when I was Head of the Computer Science Department at Carnegie Mellon University. I received a registered package containing a paper and a letter from someone named Henryk Woźniakowski in Warsaw. I do not recall the date because I did not realize it was to be the beginning of a life transforming relation which has already lasted 36 years.

The paper, which was titled "Maximal Stationary Iterative Methods for the Solution of Operator Equations" proved conjectures I had framed in the early 60s with a very important difference. My conjectures had been for scalar problems; Woźniakowski proved them for finite-dimensional and infinite-dimensional operator equations.

In a flashback to 1959 I shall tell you why I was so excited by this paper. I had just received my PhD and was working at the Mathematics Research Center at Bell Laboratories. One day a colleague named Joseph Kruskal asked me for advice on how to numerically approximate a zero of a function involving an integral. Since the integral had to be approximated at each iterative step the function was expensive to compute. I could think of a number of ways to solve this problem. What was the optimal algorithm, that is, the method which would minimize the required computational resources? To my surprise there was no theory of optimal algorithms.

Indeed, the phrase computational complexity, which is the study of the minimal computational resources required to solve problems, was not introduced until 1965 by Hartmanis and Stearns [1].

I became fascinated with creating what might be called optimal iteration theory. The initial problem was to solve the scalar nonlinear equation $f(x) = 0$. Assume for simplicity that the zero is simple.

The key insight was that the information used by an iteration determines the maximal order and the most effective methods are iterations of maximal order. The focus was on the information and not the particular algorithm that used the information. Let me give you some concrete examples. You are all familiar with Newton iteration. As you know, it requires the evaluation of f and its first derivative at each step and its order of convergence is two. The order of convergence is a measure of how fast the iteration converges. We can obtain an iteration of order 3 which uses evaluations of f and its first and second derivatives. Generally, there are known methods due to Euler and Chebyshev which use the first $s-1$ derivatives and are of order s [2, p. 81]. These are one-point iterations; that is, all the evaluations are at one point. The question that interested me was whether we could do better. Was it possible for there to exist a one-point iteration of order s that did not require the evaluation of f and its first $s-1$ derivatives? The answer is no. I called this the fundamental theorem of one-point iterations [2, p. 97]. Any one-point iteration of order s must evaluate f and its first $s-1$ derivatives.

I do not have to look at the structure of the iteration, only at the information it uses. To put it another way, the maximal order of any one-point iteration using $s-1$ derivatives is s. For later research on maximal order, see for example [3].

This is a huge simplification. The maximal order of convergence is determined completely by the information available to the iteration, not by its particular form. The significance of iterations of maximal order is that if the cost of information, that is the evaluation of f and its derivatives, is sufficiently large one can neglect the cost of combining the information and the best methods are iterations of maximal order.

You are all familiar with the secant method. Where does it fit in? From an information point of view the secant method evaluates f and reuses one previous evaluation of f. Its order is the golden mean which, as you know, is about 1.62. A method that uses a new evaluation of f and no previous evaluations is of order 1. So the previous evaluation of the secant method adds .62 to the order. There is a method which reuses 2 previous evaluations of f and is of order about 1.84. There is a method which reuses three previous values whose order is about 1.92. These iterations are examples of one point iterations with memory. This and other data suggest that with any finite number of previous values the order will be less than two. That is, the previous evaluations add less than one to the order.

Iterations that use new values of f are a special case. It is natural to consider iterations which use new values of f and its first $s-1$ derivatives at a point

and reuse any number of previous values. I defined the class of interpolatory iterations and proved that all the old information adds less than one to the order. This is a theorem for interpolatory iteration. I conjectured that this was true for any one-point iteration with memory. This was one of the topics covered in a 1964 monograph called "Iterative Methods for the Solution of Equations" [2]. I am pleased that it is been reissued by the American Mathematical Society and is still in print.

That is the end of the flashback and I want to return to 1972 when I received the paper from Henryk Woźniakowski. As I mentioned earlier, he attacked the problem of maximal order for finite-dimensional and infinite-dimensional operator equations [4]. He proved the maximal order of interpolatory iteration in the scalar case thus settling the conjecture about one-point iterations with memory. He also proved that in the operator case any one-point iteration of order s requires the evaluation of the first $s-1$ derivatives.

I invited Woźniakowski to give a talk at a May 1973 Carnegie Mellon University Symposium. He could not obtain a passport in time to participate. He finally arrived on October 16, 1973. He told me later what led to his paper. He was attending a summer school in Gdansk. Stefan Paszkowki of the University of Wroclaw asked if he had read my 1964 monograph. He had not but he obtained a copy – the rest is history. Henryk continued to visit me and together with my former PhD student, H. T. Kung, now a chaired professor at Harvard, we continued to work on optimal iteration theory. Then in 1976 there came an event that changed the course of our research.

A PhD student named Arthur Werschulz, now a professor at Fordham University and part of our research group at Columbia, gave a seminar where he used some of the techniques from nonlinear equations to attack the complexity of integration. Our reaction was that integration is inherently different from solving nonlinear equations; one does not solve integration iteratively. Because these problems are so different there must be a very general structure which underlies this and many other problems. Henryk and I always maintained long lists of research ideas. But we were so interested in this issue that we called it the S problem which stood for Special problem.

Our search for the general structure led to the 1980 monograph "A General Theory of Optimal Algorithms" [5]. We developed the theory over normed linear spaces with applications to problems such as approximation and linear partial differential equations. We confined ourselves to the worst case setting. That is we guaranteed an approximation for all inputs in a class.

We called the field analytic complexity. This was to differentiate it from algebraic complexity which was a very active research area in the late 60s and 70s. Algebraic complexity deals with problems such as the complexity of matrix multiplication which can be solved exactly while analytic complexity deals primarily with problems from analysis which cannot be solved exactly.

Part B of our 1980 book deals with an iterative information model. It turns out that this material is conceptually and technically more difficult. It was a historical accident which I have told you about earlier that we started with the study of nonlinear equations.

We also gave a brief history of the precursors to the general theory. I would like to mention a few of the earlier results. These all dealt with specific problems and did not attempt a general theory. The earliest paper which we discovered only recently is by Richard von Mises [**?**] which was published in 1933 in the Zeitschrift für Angewandte Mathematik und Mechanik. He considered univariate integration with fixed nodes and found the best weights. Arthur Sard authored a series of papers starting in 1949 and a monograph [7] in which he studied optimal algorithms for univariate quadrature with fixed nodes. He discussed extending his results to the approximation of linear functionals. Sard was apparently not familiar with the paper of von Mises. In 1950, Sergei Nikolskij [8] independently studied univariate integration but permitted the evaluation points to be optimally chosen. Another 1950 paper on univariate integration was written by Hans Bückner [9]. In a series of remarkable papers starting in 1959 Nikolaj Bakhvalov [10] studied optimal methods for multivariate integrals and obtained lower bounds on the error.

All these authors assume linear algorithms; that is, algorithms that are a linear combination of the information. Then in 1965 Sergei Smolyak [11] proved that for convex and balanced sets the optimal algorithm for the approximation of linear functionals is linear. Therefore, the assumption of linear algorithm is often not needed.

As you know the optimal strategy for approximating a zero of a continuous scalar function with a sign change is bisection. What about approximating a maximum of a unimodal function, that is a function which has only one maximum. In a 1953 publication Jack Kiefer [12] proved that if function evaluations are used then Fibonacci search is optimal. This was his 1948 MIT Master's thesis which was only published later with the encouragement of Jacob Wolfowitz. The previous work on optimal algorithms was for linear problems such as integration and approximation. To the best of my knowledge this was the first result for a nonlinear problem.

In 1983 Grzegorz Wasilkowski joined Henryk and me to write the monograph "Complexity and Information" [13]. We showed that uncertainty could be measured without a norm or metric. We decided to rename the field ε-complexity.

One day my wife, Pamela McCorduck, asked me why ε-complexity. I explained that ε denotes a small quantity. She did not seem impressed. Since Pamela is the author of numerous books I took her lack of enthusiasm seriously and started thinking about a new name. One day I was chatting with my friend, Richard Karp, who as you know was a pioneer in the study of NP-completeness. I mentioned to Richard that key concepts were information and complexity and he suggested information-based complexity which we adopted as the name of the field. For brevity I will often refer to the field as IBC.

The Journal of Complexity was born in 1985. To the best of my knowledge it was the first journal with complexity as its title. In preparing for writing this talk I took a look at Volume 1. There were thirteen people on the Editorial Board. They included three Nobel Laureates (Kenneth Arrow, Gerard Debreu, and Leonid Hurwicz, who is the most recent winner of the Economics Prize), one Fields medalist (Steven Smale), two Turing Prize winners (Michael Rabin and Richard Karp), the founder of Mathematica (Stephen Wolfram), one of the pioneers of algebraic complexity (Shmuel Winograd), the current President of Tel Aviv University (Zvi Galil), a chaired Harvard professor (H. T. Kung) and the recipient of an honorary doctorate from the Friedrich Schiller University Jena (Henryk Woźniakowski).

The first volume consisted of two issues containing 285 pages. All the papers were from the Symposium on the Complexity of Approximately Solved Problems held at Columbia in April 1985. Jumping forward to the present the Journal of Complexity now publishes some 1000 pages annually in six issues.

In 1988 Erich Novak published "Deterministic and Stochastic Error Bounds in Numerical Analysis" [14] based on his Habilitation thesis. He studies worst case error bounds which he connects with Kolmogorov n-widths. He also studies error bounds in the randomized and average case settings. The theory is applied to problems such as approximation, optimization, and integration.

In 1988 Wasilkowski, Woźniakowski and I published "Information-Based Complexity" [15]. It integrates the work of numerous researchers and reports many new results. The theory is developed over abstract linear spaces, usually Hilbert or Banach spaces. The worst, average, probabilistic, and asymptotic settings are analyzed. Numerous applications are also presented; these are developed over function spaces. Applications include function approximation, linear partial

differential equations, integral equations, ordinary differential equations, large linear systems, and ill-posed problems.

Information-based complexity is defined as the branch of computational complexity that deals with the intrinsic difficulty of the approximate solution of problems for which information is partial, contaminated, and priced. To motivate this characterization consider the numerical solution of a partial differential equation. The coefficients and the initial or boundary values are specified by functions. Since functions cannot be input to a digital computer we have to discretize them by, for example, evaluating them at a finite number of points. Thus a function is represented by a vector of numbers in the computer. There are usually an infinite number of functions which are all represented by the same vector; the mapping is many to one. We say the information about the mathematical input is partial. In addition, there will be round-off errors in evaluating the function and so the information is contaminated. If information is partial and contaminated the problem can only be approximately solved. Finally we will be charged for evaluating the functions. So the information is priced. Indeed for many problems the cost of the information dominates the cost of combining the information to get an answer.

The next decade was one of rapid progress in IBC which I shall indicate by briefly summarizing five monographs published during that period.

In 1991 Arthur Werschulz published "The Computational Complexity of Differential and Integral Equations: An Information-Based Approach" [16]. Werschulz studies algorithms and complexity of elliptic partial differential equations in the worst case setting. He also studies Fredholm integral equations of the second kind as well as ill-posed problems. In addition, there is a chapter on the average case setting.

In 1996 Leszek Plaskota published "Noisy Information and Computational Complexity" [17]. Plaskota studies both bounded and stochastic noise. Before his work the study of noisy information had lagged due, at least in part, to the technical difficulties.

1998 saw the publication of "Complexity and Information" [18] by Traub and Werschulz. This monograph is a greatly expanded and updated version of a series of lectures I gave in 1993 in Pisa at the invitation of the Accademia Nazionale dei Lincee. It starts with an introduction to IBC and then moves to a variety of topics including very high-dimensional integration and mathematical finance, complexity of path integration, and assigning values to mathematical hypotheses. It concludes with a bibliography of over 400 papers and books published since 1987.

Klaus Ritter's monograph on "Average-Case Analysis of Numerical Problems" [19], which was based on his Habilitation thesis, appeared in 2000. The book provides a survey of results that were mainly obtained in the last ten years as well as many new results. Background material on reproducing kernel Hilbert spaces, random fields, and measures on function spaces is included.

2001 saw the publication of "Optimal Solution of Nonlinear Equations" [20] by Kris Sikorski. The monograph studies algorithms and complexity in the worst case setting. Topics include nonlinear equations, fixed points of contractive and noncontractive mappings, and topological degree.

I shall now return to 1991 when we held the first Schloss Dagstuhl Seminar on Algorithms and Complexity for Continuous Problems. The Schloss Dagstuhl Seminars are the computer science equivalent to the Oberwolfach meetings in mathematics. In 2009 we will hold our ninth Seminar which may be a record. The Seminars are limited to 40 participants. As the field has grown it has become increasingly difficult for the Organizing Committees to issue only enough invitations to have some 40 participants.

In 1991 Woźniakowski published a paper "Average Case Complexity of Multivariate Integration" [21] in the Bulletin of the American Mathematical Society which was to be a rich source of new directions for IBC. He showed that the optimal points were related to the low-discrepancy points which had been extensively studied by number theorists including Fields Medalist Klaus Roth.

Quasi-Monte Carlo methods are deterministic methods based on low discrepancy points. A Columbia student, Spassimir Paskov, found empirically that Quasi-Monte Carlo was far superior to Monte Carlo for real-world problems in computational finance [22]. In trying to understand why, Ian Sloan and Henryk Woźniakowski introduced the idea of weighted spaces [23]. Discrepancy theory itself has also been a rich source of IBC problems.

1994 saw the first MCQMC Conference organized by Harald Niederreiter. This biennial conference is devoted to Monte Carlo and Quasi-Monte Carlo methods. Many IBC topics and researchers are represented in MCQMC. The 2008 conference will be in Montreal with Stefan Heinrich as Chair of the Steering Committee. The 2010 conference will be in Warsaw with Henryk as the Chair.

1995 saw the first Conference on the Foundations of Computational Mathematics. These triennial meetings always feature an IBC Workshop and an IBC plenary speaker. The 2008 conference was in Hong Kong and the IBC plenary speaker was Henryk.

1996 saw the creation of the Best Paper Award of the Journal of Complexity which carries a $3000 prize and a plaque. Roughly half the winning papers have been in IBC. Incidentally, Erich Novak won in 2001 for a paper whose significance I shall describe later. The only double winner is Stefan Heinrich, University of Kaiserslautern. Heinrich was a co-winner in 1998 and won again in 2004.

Since I am on this topic I shall mention two more IBC Prizes. The Prize for Achievement in Information-Based Complexity was created in 1999. The winners of this annual prize have been Erich Novak; Sergei Pereverzev, Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Science; Grzegorz Wasilkowski, University of Kentucky; Stefan Heinrich, University of Kaiserslautern; Arthur Werschulz, Fordham University; Peter Mathé, Weierstrass Institute for Applied Analysis and Stochastics, Berlin; Ian Sloan, University of New South Wales; Leszek Plaskota, University of Warsaw; Klaus Ritter, TU Darmstadt; and Anargyros Papageorgiou, Columbia University.

In 2003 we created a third annual prize, the Information-Based Complexity Young Researcher Award for researchers who have not yet reached their 35th birthday [24]. The recipients to date have been Frances Kuo, University of New South Wales; Christiane Lemieux, University of Calgary; Josef Dick, University of New SouthWales; Friedrich Pillichshammer, University of Linz; Jakob Creutzig, TU Darmstadt; Dirk Nuyens, Catholic University, Leuven; and Andreas Neuenkirch, University of Frankfurt.

Another new direction for IBC was initiated by Erich Novak [25] with the publication in 2001 of "Quantum Complexity of Integration". Until this seminal paper, only the complexity of discrete problems on quantum computers had been studied. Novak studied multivariate integration over Hölder classes. He proved exponential quantum speedups over the classical worst case and quadratic speedups over the classical randomized case.

Since then there has been much progress on IBC problems in the quantum setting with many surprising results mostly due to Stefan Heinrich (see for example [26]) and also due to Papageorgiou and Woźniakowski (see for example [27]).

This talk is titled "A Brief History of Information-Based Complexity". I want to end with the present and the future. The European Mathematical Society recently published Volume I of "Tractability of Multivariate Problems" [28] by Erich Novak and Henryk Woźniakowski. This volume is some 400 pages in length. Volume II and III, of similar length, are in progress. There is a huge literature on the complexity of problems in d variables. The complexity bounds are usually sharp with respect to ε^{-1} where ε is the error threshold but have,

unfortunately, unknown dependence on d. To determine if a problem is tractable the dependence on both ε^{-1} and d must be determined. Tractability requires new proof techniques to obtain sharp bounds on d. There are many surprising results.

Volume I lists 30 open problems which continues an IBC tradition. Dozens of open questions have been listed in many IBC papers and books. Almost all are still open. Many years ago I gave a talk at MIT which I concluded with half a dozen open questions. Marvin Minsky was in the audience and told me he always saves open questions for his students. I told him there were many more open questions where these came from. Why are there so many open questions? IBC is a relatively young field that covers a huge area of optimal algorithms and complexity for continuous mathematics.

What are some of the future directions? I believe the three volume monograph by Novak and Woźniakowski opens up a whole new area of investigation.

Another huge area for research is that of problems specified by nonlinear operators. To date, much of IBC deals with linear operators and their applications such as integration, approximation, integral equations and linear partial differential equations. Attacking problems defined by nonlinear operators will present us with entirely new challenges.

We have come a long way starting with specific problems such as univariate integration and the solution of scalar nonlinear equations and progressing to a general abstract theory with applications ranging from discrepancy theory and computational finance to quantum computing. I believe the next 50 years will see even greater progress.

References

[1] J. Hartmanis and R.E. Stearns, On the computational complexity of algorithms. *Trans. Amer. Math. Soc.* **117** (1965), 285–306. 61

[2] F. Traub, *Iterative methods for the solution of equations*. Prentice-Hall, Englewood Cliffs, N. J., 1964; now available from the American Mathematical Society. 62, 63

[3] H. Woźniakowski, Maximal order of multipoint iterations using n evalutions. In *Analytic computational complexity* (Proc. Symp., Carnegie-Mellon Univ., Pittsburgh, 1975), ed. by J. F. Traub, Academic Press, New York, 1976, 75–107. 62

[4] H. Woźniakowski, Maximal stationary iterative methods for the solution of operator equations. *SIAM J. Numer. Anal.* **11** (1974), 934–949. 63

[5] J. F. Traub and H. Woźniakowski, *A general theory of optimal algorithms*. Academic Press, New York, NY, 1980. 63

[6] R. von Mises, Zur mechanischen Quadratur. *Z. Angew. Math. Mech.* **13** (1933), 53–56.

[7] A. Sard, *Linear approximation*. Amer. Math. Soc., Providence, RI, 1963. 64

[8] S. Nikolskij, On the problem of approximation estimate by quadrature formulas. *Uspekhi Mat. Nauk* **5** (1950), 165–177 (in Russian). 64

[9] H. Bückner, Bemerkungen zur numerischen Quadratur, I; II. *Math. Nachr.* **3** (1950), 142–145; 146–151. 64

[10] N. S. Bakhvalov, On the approximate calculation of multiple integrals. *Vestnik Moskov. Univ. Ser. Mat. Meh. Astr. Fiz. Him.* **4** (1959), 3–18 (in Russian). 64

[11] S. Smolyak, On optimal restoration of functions and functionals of them. Candidate Dissertation, Moscow State University, 1965 (in Russian). 64

[12] J. Kiefer, Sequential minimax search for maximum. *Proc. Amer. Math. Soc.* **4** (1953), 503–505. 64

[13] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, *Complexity and information*. Addison-Wessley, Reading, MA, 1983. 65

[14] E. Novak, *Deterministic and stochastic error bounds in numerical analysis*. Lecture Notes in Math. 1349, Springer-Verlag, Berlin, 1988 65

[15] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, *Information-based complexity*. Academic Press, San Diego, CA, 1988. 65

[16] A. G. Werschulz, *The computational complexity of differential and integral equations: an information-based approach*. Oxford University Press, Oxford, 1991. 66

[17] L. Plaskota, *Noisy information and computational complexity*. Cambridge University Press, Cambridge, 1996. 66

[18] J. F. Traub and A. G. Werschulz, *Complexity and information*. Cambridge University Press, Cambridge, 1998. 66

[19] K. Ritter, *Average-case analysis of numerical problems*. Lecture Notes in Math. 1733, Springer-Verlag, Berlin, 2000. 67

[20] K. Sikorski, *Optimal solution of nonlinear equations*. Oxford University Press, Oxford, 2001. 67

[21] H. Woźniakowski, Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* **24** (1991), 185–194. 67

[22] S. Paskov and J. F. Traub, Faster evaluation of financial derivatives. *J. Portfolio Management* **22** (1995), 113–120. 67

[23] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. of Complexity* **14** (1998), 1–33. 67

[24] See www.cs.columbia.edu/ traub for further information on all three prizes. 68

[25] E. Novak, Quantum complexity of integration. *J. of Complexity* **17** (2001), 2–16. 68

[26] S. Heinrich, Quantum integration in Sobolev spaces. *J. of Complexity* **19** (2003), 19–24. 68

[27] A. Papageorgiou and H. Woźniakowski, The Sturm–Liouville eigenvalue problem and NP-complete problems in the quantum setting with queries. *Quantum Inf. Process.* **6** (2007), 101–120. 68

[28] E. Novak and H. Woźniakowski, *Tractability of multivariate problems*. Volume I: Linear information. EMS Tracts Math. 6, European Mathematical Society Publishing House, Zürich, 2008. 68

How high is high-dimensional?

Ian H. Sloan

1 Introduction

It is a pleasure for me to be able to join in the celebration of the achievements of my good friend and colleague, Henryk Woźniakowski. Henryk's achievements are wide-ranging, but in this lecture I want to concentrate on just the problem of high-dimensional integration.

Consider the problem of evaluating the d-dimensional integral

$$I_d(f) := \int_0^1 \cdots \int_0^1 f(x^1, \ldots, x^d) \mathrm{d}x^1 \ldots \mathrm{d}x^d = \int_{[0,1]^d} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x},$$

where f is a continuous function over the d-dimensional unit cube.

To take up immediately the question in the title, let us ask: how large can we reasonably take d to be? Of course the answer depends critically on the nature of f. But there was a time in the pre-computer era when, except for special cases, even two-dimensional integrals seemed taxing. Then with the coming of computers, as I well remember as a young PhD student solving electron-atom scattering problems, one could contemplate numerical evaluation of integrals with $d = 4$ or even $d = 5$, but certainly not $d = 10$. The entry of number theorists such as Hlawka, Korobov, Sobol′ into the subject in the 1950s and 1960s (an intrusion not generally noticed at that time by the physics world) made it thinkable to go perhaps as high as $d = 15$ or possibly even $d = 20$, but certainly not to d in the hundreds. Yet now it is possible to contemplate numerical integration with d in the hundreds or thousands or even more, if the problem is right.

Part of the change in attitude came from the experimental courage of Joe Traub and Spassimir Paskov [13], who in 1995 surprised everybody by successfully using low discrepancy sequences to value a parcel of mortgage backed securities, with the problem expressed as a 360-dimensional integral.

At that time there seemed to be a fundamental barrier to understanding the success of low-discrepancy sequences for d in the hundreds. That difficulty was the Koksma–Hlawka inequality. This is an error bound for any quadrature

approximation of the equal-weight (or quasi-Monte Carlo, or QMC) form

$$Q_{n,d}(f) = \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{t}_k), \tag{1}$$

with $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^d$. The Koksma–Hlawka inequality states (see [11])

$$|I_d(f) - Q_{n,d}(f)| \le D^*(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n) V(f),$$

where $D^*(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$ is the "star discrepancy", which is a positive number that depends on the points but is independent of f, while $V(f)$ is a measure of the variation of f (technically, it is the variation in the sense of Hardy and Krause), which is independent of the choice of the points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^d$. For the particular case of a Sobol′ sequence, or any other "low-discrepancy" sequence, the star discrepancy is bounded by

$$D^*(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n) \le C_d \frac{1}{n} (\log n)^{d-1},$$

with C_d independent of n. For small values of d the logarithmic factor in the Koksma–Hlawka inequality causes no concern, but it is a different matter when d is in the hundreds: for the Koksma–Hlawka error bound has the disconcerting property that the bound becomes worse as n increases with d fixed (that is, the error bound gets progressively larger), until $n \approx e^{d-1}$; a number that is truly astronomical when d is in the hundreds.

But the Koksma–Hlawka inequality is an upper bound, not the error itself, and by now we have other error bounds that tell a different story. In this presentation I want to explain how Henryk and I came to be involved together in this story. It will become clear that if over the past decade our understanding of high-dimensional integration has advanced rather dramatically, then Henryk has played a very large part in that advance.

I first met Henryk in 1992, at an Oberwolfach conference devoted to Numerical Integration. Henryk talked about "Numerical integration in various settings", touching among other things on his important work on numerical integration in the average case setting. I talked about "lattice rules" for high-dimensional integration, a topic I shall come back to later, and which by that time had already been of interest to me for about a decade.

We found many points of common interest, so it was not long before I invited Henryk to visit me in Australia, thereby starting a collaboration which for me has been one of the finest in a long professional life. To me a good collaboration is

one in which each person brings to the table something unique, something that the other party could not provide. In my case the contribution was no doubt some knowledge of lattice methods. Henryk brought many things, including his vast knowledge of information-based complexity. But for me the truly special thing brought by Henryk came in the form of a question: he kept asking "What happens to the error as d goes to infinity?".

It is a striking historical aspect of the work of both numerical analysts and number theorists working in the area of multiple integration that the question of what happens as $d \to \infty$ seems almost never to arise. Until very recently d was considered to be fixed, and the interest lay in what happens as n (the number of quadrature points) goes to infinity. Even when asked, Henryk's question of what happens when $d \to \infty$ cannot be easily answered, because the dependence on d tends to be hidden in the "constant". And of course the question has no meaning until we specify the setting, since the function being integrated must change as $d \to \infty$. But the value of a good question, such as Henryk's, is incalculable.

During the first of Henryk's many visits to Australia, in 1994, we talked very often about the classical work of the number theorists on the use of "lattice" rules for the numerical integration of periodic functions. Lattice rules are a particular kind of quasi-Monte Carlo method which in the simplest "rank-1" case are given by the formula

$$L_{n,d,\boldsymbol{z}}(f) := \frac{1}{n} \sum_{k=1}^{n} f\left(\left\{k \frac{\boldsymbol{z}}{n}\right\}\right). \tag{2}$$

Here $\boldsymbol{z}$ is an integer vector of length d, and the braces around $k\boldsymbol{z}/n$ mean that each component of that d-vector is to be replaced by its fractional part in $[0, 1)$. A typical theorem of Korobov and Hlawka from around 1960 is the following one. (For a proof and generalisation see [15], Section 4.5.) Here E_d is the set of continuous 1-periodic functions on $[0, 1]^d$ whose Fourier coefficients

$$\hat{f}(\boldsymbol{h}) := \int_{[0,1]^d} f(\boldsymbol{x}) \exp(-2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

decay at a prescribed rate: specifically, $f \in E_d$ if

$$|\hat{f}(\boldsymbol{h})| = |\hat{f}(h_1, \ldots, h_d)| \le \frac{1}{(\bar{h}_1 \ldots \bar{h}_d)^2}, \tag{3}$$

where

$$\bar{h} = 1 \text{ if } h = 0, \quad \bar{h} = |h| \text{ if } h \ne 0.$$

Theorem 1. *Assume n is prime. Then there exists $\boldsymbol{z} \in \mathbb{Z}^d$ and $C_d > 0$ such that*

$$\left| I_d(f) - \frac{1}{n}\sum_{k=1}^{n} f\left(\left\{\frac{k\boldsymbol{z}}{n}\right\}\right)\right| \le \frac{C_d(\log n)^{2d}}{n^2} \quad \text{for all } f \in E_d,$$

where C_d is independent of n.

You can see that the theorem is interesting and significant, but that in the form stated it contains no information about what happens as $d \to \infty$.

In Henryk's own contribution to this booklet he has explained the result we obtained in our first paper, namely that for $n < 2^d$ the error bound in the theorem must be at least 1, since for all $n < 2^d$ and every choice of $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ we were able to construct explicitly an $f \in E_d$ such that $I_d(f) = 1$ and $Q_{n,d}(f) = 0$. Thus while the asymptotic rate of convergence might be very good, the asymptotic rate is not observable until n is at least 2^d. (In that paper we also showed that the number 2^d is sharp, in the sense that a faster rate of decay of the Fourier coefficients than we assumed in (3) really helps for $n \ge 2^d$, whereas for $n < 2^d$ extra smoothness of f does not help at all.)

In technical language, the integration problem in that classical setting is intractable. Nontechnically, since the difficulty of the problem increases exponentially with d, it is clearly hopeless to seek an explanation for the success of the Paskov and Traub experiments within that setting.

With a growing belief that the integration problem is intractable in all of the classical settings, Henryk and I began to wonder if we could find a non-classical setting in which the Paskov and Traub success could be explained. In the mortgage backed security problem the early years of the mortgage are the most important, because the product has longer to run. Perhaps, we pondered, it is always the case for finance problems that some variables are more important than others? That thought led us to introduce, in our 1998 paper [20], the concept of "weighted spaces". The notion of weighted spaces, which I shall explain in the next section, turned out to be a fruitful one, with many generalisations and extensions. In time, as I shall explain, it led to the development of a constructive (or semi-constructive, depending on your point of view) algorithm for constructing an integration rule that is "good" in a well defined sense.

In the remainder of this paper I want to give a simple account of the weighted space work that Henryk and I carried out, inspired by his question of what happens as $d \to \infty$, and then give a brief account of the constructions that developed from it. In the next section I introduce weighted spaces, informally but in essentially

the way we did in that first paper, and indicate how we established (to the surprise of both of us) a necessary and sufficient condition for "tractability". I then show in Section 3 how our existence results led to constructions. Finally, in the last section I outline, for the sake of completeness, a number of generalisations and further developments.

2 Weighted Sobolev spaces, tractability and existence

As already mentioned, the guiding idea in the 1998 paper with Henryk [20] was that perhaps the reason why some problems can be successfully tackled by Sobol′ or other QMC integration rules is that the d variables are not equally important. Of itself this was not a new idea: the suggestion that the "effective dimension" in some finance applications might be small was promoted by Caflisch, Morokoff and Owen in [1]. But our thought was that if some variables are more important than others, then this should in some way be built into the function spaces in which we work. This seems to have been a new idea.

Specifically, we decided to try to quantify the declining importance of the variables (assuming that we have already labeled the variables in order of declining importance) by introducing a non-increasing infinite sequence of positive "weights" $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots)$. The first d of these weights we built into the norm $\|f\|_{d,\boldsymbol{\gamma}}$ in our (Sobolev) space $H_{d,\boldsymbol{\gamma}}$, a Hilbert space of functions with square-integrable mixed first derivatives. Precisely, for $d = 1$ we defined

$$\|f\|_{1,\boldsymbol{\gamma}}^2 = |f(1)|^2 + \frac{1}{\gamma_1}\int_0^1 \left|\frac{\mathrm{d}f}{\mathrm{d}x}(x)\right|^2 \mathrm{d}x;$$

while for $d = 2$ we took for the square of the norm

$$\|f\|_{2,\boldsymbol{\gamma}}^2 = |f(1,1)|^2 + \frac{1}{\gamma_1}\int_0^1 \left|\frac{\partial f}{\partial x^1}(x^1,1)\right|^2 \mathrm{d}x^1 + \frac{1}{\gamma_2}\int_0^1 \left|\frac{\partial f}{\partial x^2}(1,x^2)\right|^2 \mathrm{d}x^2 + \frac{1}{\gamma_1\gamma_2}\int_0^1\int_0^1 \left|\frac{\partial^2 f}{\partial x^1 \partial x^2}(x^1,x^2)\right|^2 \mathrm{d}x^1\mathrm{d}x^2;$$

and so on. The expression for the square of the norm for general d can be written concisely as

$$\|f\|_{d,\boldsymbol{\gamma}}^2 = \sum_{\mathfrak{u}} \frac{1}{\left(\prod_{j\in\mathfrak{u}}\gamma_j\right)} \int_{[0,1]^{|\mathfrak{u}|}} \left|\frac{\partial^{|\mathfrak{u}|} f}{\partial \boldsymbol{x}_{\mathfrak{u}}}(\boldsymbol{x}_{\mathfrak{u}}, \mathbf{1})\right|^2 \mathrm{d}\boldsymbol{x}_{\mathfrak{u}}, \tag{4}$$

where the sum is over all subsets $\mathbf{u}$ of $\{1,\dots,d\}$. Here $|\mathbf{u}|$ is the cardinality of $\mathbf{u}$, $\boldsymbol{x}_{\mathbf{u}}$ denotes the vector of length $|\mathbf{u}|$ consisting of the components x_j of $\boldsymbol{x}$ with $j \in \mathbf{u}$, and $(\boldsymbol{x}_{\mathbf{u}}, \mathbf{1})$ is the vector of length d in which the components of $\boldsymbol{x}$ with $j \notin \mathbf{u}$ are replaced by the particular number 1. Observe where the weights occur: the term labeled by $\mathbf{u}$ has in the denominator the product of the weights γ_j for $j \in \mathbf{u}$.

The next step is to define the "worst case error": the worst case error of the quadrature rule (1) in the space $H_{d,\boldsymbol{\gamma}}$ is

$$\begin{aligned} e_{n,d,\boldsymbol{\gamma}}(Q_{n,d}) &:= e_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1,\dots,\boldsymbol{t}_n) \\ &:= \sup\left\{\left|I_d(f) - \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{t}_k)\right| : \|f\|_{d,\boldsymbol{\gamma}} \le 1\right\}. \end{aligned}$$

In words, the worst case error is the largest error of the rule (1) for functions f whose $H_{d,\boldsymbol{\gamma}}$ norm is no greater than 1. Of course we want to choose QMC rules that make the worst case error small.

Note the role of the weights in the terms of (4): if the weights γ_j for $j \in \mathbf{u}$ are small then their product will be very small, and f can belong to the unit ball in $H_{d,\boldsymbol{\gamma}}$ only if the corresponding mixed partial derivative of f in (4) is correspondingly small. In simple terms, small weights imply small partial derivatives for functions f in the unit ball of $H_{d,\boldsymbol{\gamma}}$.

The central result of that 1998 paper is that the behavior of the integration problem in the space $H_{d,\boldsymbol{\gamma}}$ as $d \to \infty$ depends crucially on whether or not the infinite sum

$$\sum_{j=1}^{\infty} \gamma_j \tag{5}$$

converges or diverges.

Theorem 2. *If the sum* (5) *converges, then for each* $n \ge 1$ *there exists a choice of QMC points* $\boldsymbol{t}_1,\dots,\boldsymbol{t}_n$ *such that*

$$e_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1,\dots,\boldsymbol{t}_n) \le \frac{D_{\boldsymbol{\gamma}}}{\sqrt{n}} \quad \textit{for all } n,\, d, \tag{6}$$

with $D_{\boldsymbol{\gamma}}$ *independent of* d *and* n.

If the sum (5) *does not converge then for every choice of QMC points* $\boldsymbol{t}_1,\dots,\boldsymbol{t}_n$ *and every* $\epsilon \in (0,1)$ *the number* n *of points required to ensure*

$$e_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1,\dots,\boldsymbol{t}_n) \le \epsilon e_{0,d,\boldsymbol{\gamma}}$$

increases with d *faster than any polynomial.*

Here $e_{0,d,\boldsymbol{\gamma}}$ is the worst case error with no quadrature points, that is

$$e_{0,d,\boldsymbol{\gamma}} := \sup\{|I_d(f)| : \|f\|_{d,\boldsymbol{\gamma}} \le 1\} = \|I_d\|_{d,\boldsymbol{\gamma}}.$$

Technically, the integration problem in $H_{d,\boldsymbol{\gamma}}$ is "strongly QMC-tractable" if and only if the sum in (5) is finite. I think we were both pleasantly surprised that we were able to obtain a necessary and sufficient condition in this way. Note that the sum in (5) is finite for $\gamma_j = j^{-2}$, but is certainly not finite for the classical case $\gamma_j = 1$ for all j.

A word about the proof of the theorem. The second part of the theorem (that the integration problem is not strongly QMC tractable if (5) diverges) was proved by obtaining an explicit lower bound for the worst case error in $H_{d,\boldsymbol{\gamma}}$, a lower bound that grows unboundedly as $d \to \infty$. For the classical unweighted case (that is, with $\gamma_j = 1$ for all j) the lower bound obtained in [20] is

$$e^2_{n,d,\boldsymbol{\gamma}}(Q_{n,d}) \ge e^2_{0,d,\boldsymbol{\gamma}}\big(1 - e^{-0.05478d} n\big).$$

It follows immediately that to ensure $e_{n,d,\boldsymbol{\gamma}} \le \epsilon e_{0,d,\boldsymbol{\gamma}}$ we must have $n \ge (1-\epsilon^2)e^{0.05478d}$, which increases exponentially fast in d. While the rate of exponential growth is quite slow, for the unweighted case we certainly have a problem by the time d is in the hundreds.

The proof of the first part of the theorem, that there exists a set of QMC points $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n$ such that (6) holds, has a quite different character. Here the need is to prove the existence of a suitable set of QMC points, even when we are not clever enough to know how to construct such a set. The proof (which is one we borrowed from the number theory literature, and indeed similar to the technique used to prove Theorem 1) is based on an averaging argument. In brief, we showed that the average (i.e. integral) of $e^2_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1, \dots, \boldsymbol{t}_n)$ over all choices of $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n$ is bounded above by the square of the right-hand side of (6). From that it follows (using the wonderful principle that there is always one choice that is as good as average) that there must be at least one choice of $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n$ for which

$$e^2_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1, \dots, \boldsymbol{t}_n) \le \frac{D^2_{\boldsymbol{\gamma}}}{n} \quad \text{for all } n, d.$$

But that argument leaves us no closer to finding a good set of QMC points, if we define "good" to mean one that achieves the error bound in (6). In the next section I turn towards construction. Before that, however, I want first to note an important result of Hickernell and Woźniakowski [8], one that improves upon the upper bound (6), in that the rate of convergence is increased to almost $O(1/n)$ under a stronger condition on the weights.

Theorem 3. *Assume that the weights $\gamma_1, \gamma_2, \ldots$ satisfy*

$$\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty. \tag{7}$$

Then for all $d \geq 1$ and all $n \geq 1$ there exist $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^d$ such that for all $\delta > 0$ we have

$$e_{n,d,\boldsymbol{\gamma}}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n) \leq \frac{C_{\boldsymbol{\gamma},\delta}}{n^{1-\delta}}, \tag{8}$$

with $C_{\boldsymbol{\gamma},\delta}$ depending on δ but independent of d and n.

The condition in the theorem is satisfied, for example, by $\gamma_j = j^{-3}$. Whereas in Theorem 2 the rate of convergence is no better than the well-known Monte Carlo rate $O(n^{-1/2})$, the convergence rate in Theorem 3 is arbitrarily close to $O(n^{-1})$.

3 From existence to construction

In a 2001 paper [21] Henryk and I took what I now see as a key step towards construction, when we proved the next theorem. It asserts, under the same condition (7) as in Theorem 3, that there exists a "shifted lattice rule" whose worst case error achieves the error bound (8). A shifted lattice rule is a QMC rule whose points are obtained by subjecting the points of a lattice rule to a collective displacement, or shift, with the points that fall outside the half-open unit cube being wrapped around to the other side. More precisely, the shifted lattice rule corresponding to the lattice rule (2) is

$$S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}}(f) := \frac{1}{n} \sum_{k=1}^{n} f\left(\left\{k\frac{\boldsymbol{z}}{n} + \boldsymbol{\Delta}\right\}\right), \tag{9}$$

with $\boldsymbol{\Delta} \in [0,1]^d$ being the "shift".

Theorem 4. *Assume that n is prime, and that the weights $\gamma_1, \gamma_2, \ldots$ satisfy*

$$\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty.$$

Then for all $d \geq 1$ *there exists a shifted lattice rule* $S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}}$ *such that for all* $\delta > 0$ *we have*

$$e_{n,d,\boldsymbol{\gamma}}(S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}}) \leq \frac{C_{\boldsymbol{\gamma},\delta}}{n^{1-\delta}},$$

with $C_{\boldsymbol{\gamma},\delta}$ *depending on* δ *but independent of* d *and* n.

Essentially this is the same result as in Theorem 3 (though restricted to prime values of n), but now the existence occurs within the much smaller class of "shifted lattice rules".

I claim (I am not sure if Henryk agrees) that we discovered Theorem 4 by accident. It came about this way. We had decided to study lattice rules not in the space $H_{d,\boldsymbol{\gamma}}$, but rather in a related Hilbert space of *periodic* functions, called, whether appropriately or not, "Korobov spaces". Periodic functions are easy to analyse because all the tools of Fourier analysis are available. The main purpose of our 2001 paper [21] was to show, under the same conditions as in Theorem 4, that there exists a lattice rule for which the worst case error in the weighted *Korobov* space satisfies the error bound in Theorem 4. Only when the paper was essentially finished did we realise that we could infer a result for the non-periodic space $H_{d,\boldsymbol{\gamma}}$.

The argument goes like this: Hickernell and Woźniakowski had shown, in [8], that the average of the squared worst case error $e^2_{n,d,\boldsymbol{\gamma}}(S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}})$ for the shifted lattice rule $S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}}$, taken over all shifts $\boldsymbol{\Delta}$, is just the worst case error for the lattice rule $L_{n,d,\boldsymbol{z}}$ in an appropriate weighted Korobov space. For the latter we had already shown that the desired bound holds. That allowed us to use that averaging argument yet again: in this case to say that *there exists a shift* $\boldsymbol{\Delta}$ *for which the bound on the worst case error holds*, if we keep the already selected (better than average) value of the integer vector $\boldsymbol{z}$.

Since the proof of that theorem in our 2001 paper relied once more on averaging arguments (averaging first over $\boldsymbol{z}$ and then, as indicated, over $\boldsymbol{\Delta}$), it again does not tell us how to find good choices of the parameters, in this case $\boldsymbol{z}$ and $\boldsymbol{\Delta}$. Nevertheless, because the parameters that describe a shifted lattice rule are few (and if we can leave aside the question of choosing $\boldsymbol{\Delta}$, because there are only $(n-1)^d$ choices of $\boldsymbol{z}$ to worry about, a number that is finite even if exponentially large), it did encourage us to think about the possibility of construction.

For me the idea for construction started in 1999, at the Hong Kong Workshop for the Complexity of Multivariate Problems. By the time of that Workshop it was already accepted that weighted spaces gave a good setting for the non-constructive proof of the existence of good QMC rules. Also, Henryk and I were

about to submit the paper on the existence of good lattice rules that later appeared as [21]. At that time nothing was known about construction, but I remember that there was some vigorous discussion at the Workshop on the desirability of constructive proofs. During a coffee break I sat down together with Stephen Joe of the University of Waikato, and we said to each other something along these lines: "Now that we know that a good lattice exists, is it thinkable that we can construct such a thing one component at a time?" Fortunately, we had at that moment forgotten the conventional wisdom for classical lattice rules, that it is folly to attempt to construct a good lattice rule in d dimensions from one in $d-1$ dimensions. I say fortunately, because in the context of weighted spaces it turns out to be perfectly possible to build up good lattices one coordinate at a time.

The work that Stephen Joe and I did in that coffee break led in due course to what is now called the "component-by-component" or CBC construction of good lattice rules, see [16], [17], [18]. Here I shall describe not the original version, but rather the version for a "randomly shifted lattice rule", because this is at the same time the simplest to describe, the easiest to implement, and the one that yields the strongest theoretical results. A randomly shifted lattice rule is again given by (9) but now with $\boldsymbol{\Delta}$, far from being deterministic, being chosen randomly from a uniform distribution on $[0,1]^d$. In practice we use (9) with 10 or 20 independent samples of $\boldsymbol{\Delta}$, take their mean as the estimate of the integral, and use the spread of results as a probabilistic estimate of the error, just as is done in the Monte Carlo method. For more precise information see [17].

One advantage of the randomly shifted version of the CBC algorithm is that the only quantity we need to specify is the integer vector $\boldsymbol{z}$ in (9). Here is the entire algorithm for the particular case of our weighted space $H_{d,\boldsymbol{\gamma}}$.

Algorithm

Set $z^1 = 1$.
For $d = 2, 3, \ldots, d_{\max}$, choose z^d from $\{1, 2, \ldots, n-1\}$ so as to minimise

$$m_{n,d,\boldsymbol{\gamma}}(z^1, z^2, \ldots, z^d) := \frac{1}{n} \sum_{k=1}^{n} \prod_{j=1}^{d} \left[1 + \gamma_j \left(B_2 \left(\left\{ \frac{k z^j}{n} \right\} \right) + \frac{1}{3} \right) \right].$$

End

In the algorithm $B_2(x) := x^2 - x + \frac{1}{6}$ is the Bernoulli polynomial of degree 2, and $m_{n,d,\boldsymbol{\gamma}}(\boldsymbol{z})$ is related in a simple way to the root-mean-square average of $e_{n,d,\boldsymbol{\gamma}}(S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}})$ over the shifts, which happens to have a very simple explicit

expression:

$$e^{\mathrm{rms}}_{n,d,\boldsymbol{\gamma}}(\boldsymbol{z}) := \left(\int_{[0,1]^d} e^2_{n,d,\boldsymbol{\gamma}}(S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}})\mathrm{d}\boldsymbol{\Delta} \right)^{\frac{1}{2}}$$
$$= \left(-\prod_{j=1}^{d}\left(1+\frac{\gamma_j}{3}\right) + \frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{d}\left[1+\gamma_j\left(B_2\left(\left\{\frac{kz^j}{n}\right\}\right)+\frac{1}{3}\right)\right]\right)^{\frac{1}{2}}$$
$$= \left(-\prod_{j=1}^{d}\left(1+\frac{\gamma_j}{3}\right) + m_{n,d,\boldsymbol{\gamma}}(\boldsymbol{z})\right)^{\frac{1}{2}}.$$

It follows that minimising $m_{n,d,\boldsymbol{\gamma}}(z^1,\ldots,z^d)$ is equivalent to minimising $e^{\mathrm{rms}}_{n,d,\boldsymbol{\gamma}}(z^1,\ldots,z^d)$.

The cost of a naive implementation is $O(n^2 d^2_{\max})$, but a clever implementation by Nuyens and Cools (see [12]) has since reduced the total cost to $O(n\log(n)d_{\max})$, making it now feasible to run the algorithm with $d_{\max}$ in the thousands and n in the millions.

Of course a construction is of no value unless we know that the result is of good quality. The first result on the quality of the CBC construction, presented along with the algorithm in [17], was the following:

Theorem 5. *Let n be prime, and assume $\sum_{j=1}^{\infty}\gamma_j < \infty$. If $z^1,\ldots,z^d$ are given by the algorithm above, then*

$$e^{\mathrm{rms}}_{n,d,\boldsymbol{\gamma}}(z^1,\ldots,z^d) \le \frac{D_{\boldsymbol{\gamma}}}{\sqrt{n}},$$

where $D_{\boldsymbol{\gamma}}$ is exactly as in Theorem 2.

The proof, naturally enough, is by induction, with that familiar averaging argument being used in the inductive step to establish existence of a "good" choice z^d (good in the sense of having a worst case error no larger than the bound in (6)). For if at the dth step the average of the squared worst case error over all choices of z^d satisfies the error bound, then certainly there exists at least one good choice of z^d. Once we know that a good value of z^d exists, we know with certainty that the value that minimises the worst-case error is a good one.

It is clear that we have already come a long way: we have now proved in a constructive way the same result Henryk and I had established non-constructively in our 1998 paper.

On the other hand, we have seen in Theorem 4 an existence result with a better rate of convergence, if the sequence of weights $\boldsymbol{\gamma}$ satisfies the stronger condition (7). The high point in the story of weighted spaces was reached when Frances Kuo (at that time a PhD student of Stephen Joe at the University of Waikato, and now my colleague at the Unversity of New South Wales) proved, in [9], that the above algorithm, without any change, achieves the same essentially optimal rate of convergence as in Theorem 4. Here is her wonderful result. (See also [3].)

Theorem 6. *Let n be prime, and let $z^1, z^2, \ldots, z^d$ be chosen by the CBC algorithm above. Assume that the weights $\gamma_1, \gamma_2, \ldots$ satisfy*

$$\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty.$$

Then for all $d \geq 1$ and all $\delta > 0$ we have

$$e^{\rm rms}_{n,d,\boldsymbol{\gamma}}(z^1, \ldots, z^d) \leq \frac{C_{\boldsymbol{\gamma},\delta}}{n^{1-\delta}},$$

with $C_{\boldsymbol{\gamma},\delta}$ depending on δ but independent of d and n.

This result implies (by yet another appeal to the averaging argument) that there exists a shift $\boldsymbol{\Delta}$ such that

$$e_{n,d,\boldsymbol{\gamma}}(S_{n,d,\boldsymbol{z},\boldsymbol{\Delta}}) \leq \frac{C_{\boldsymbol{\gamma},\delta}}{n^{1-\delta}},$$

with $\boldsymbol{z} = (z^1, \ldots, z^d)$. It does not give us the shift $\boldsymbol{\Delta}$, but that seems to be of no practical consequence since in practice we think it better to choose shifts randomly, as described above, rather than to try to find the "best" shift. The last result thus completes the circle, with all of the non-constructive existence proofs now supplemented by concrete constructions that achieve the same results.

The CBC construction also opens the way to applications, and in doing so opens new questions, of which much the most urgent is: how, for a particular application, should one choose the weights? Somehow the question of how to choose the weights did not seem so pressing when we were merely proving academic results on existence. It becomes very pressing when a practitioner wants to know the best choice of parameters to feed into the CBC algorithm. Guiding the choice of weights has become one of the most urgent tasks facing researchers in this area. I personally believe that there is still much to be done in that direction.

4 Consequences and conclusions

The simple idea of weighted spaces that Henryk and I began playing with a dozen years ago has since extended in many different directions. Weights γ_j have been replaced by weights $\gamma_{d,j}$, see [23], allowing the weights to have explicit dependence on the number of variables as well as on the particular coordinate, and the weights have also been liberated (see [5]) from the condition arbitrarily imposed in the original paper [20] that $\gamma_1 = 1$. The product of the weights $\prod_{j\in\mathbf{u}} \gamma_j$ used for the term labeled by $\mathbf{u}$ in (4) has been replaced (see [6]) by the more general weight $\gamma_{\mathbf{u}}$, which may be prescribed independently for each subset $\mathbf{u}$. There is now much interest (and some controversy) surrounding the special case of "finite-order" weights (see [6], [14], [19], [24]), in which $\gamma_{\mathbf{u}} = 0$ for $|\mathbf{u}|$ bigger than some number q, with q a small number, say 2 or 3. The choice of function space has been widened: nowadays the particular space $H_{d,\boldsymbol{\gamma}}$ defined in this paper is called the "anchored" Sobolev space (of smoothness parameter 1), with anchor at 1, but there are many other choices, including general choices of anchor, see [22]. Earlier Hickernell [7] had proposed anchoring at the mid-point $\frac{1}{2}$. Others see advantages in using "unanchored" Sobolev spaces [10], [22]. Finally, but importantly, weighted Walsh spaces have been defined, and results analogous to those above have been obtained in these spaces, with shifted lattice rules replaced by digitally shifted digital nets, see [4].

Progress has also been made in the direction of having lattice rules available for non-prime n, and for a range of values of n. In particular, and impressively, in the paper [2] good quality lattice rules obtained by a component-by-component technique (and hence arbitrarily extensible in the dimension d) are presented for all powers of 2 between 2^{10} and 2^{20}, and all integers in between (that is, for all numbers n between roughly a thousand and a million).

So how high is high dimensional? If the problem is right, then there is now surely *no* limit.

Acknowledgments. I am grateful to the Australian Research Council for support under its Centres of Excellence Program. And I am especially grateful to Henryk Woźniakowski for his excellent question, and for many years of inspiration and friendship.

References

[1] R. E. Caflisch, W. Morokoff and A. Owen, Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance* **1** (1997), 27–46. 77

[2] R. Cools, F. Y. Kuo and D. Nuyens, Constructing embedded lattice rules for multivariate integration. *SIAM J. Sci. Computing* **28** (2006), 2162–2188. 85

[3] J. Dick, On the convergence rate of the component-by-component construction of good lattice rules. *J. of Complexity* **20** (2004), 493–522. 84

[4] J. Dick and F. Pillichshammer, Multivariate integration in weighted Hilbert spaces based on Walsh functions and weighted Sobolev spaces. *J. of Complexity* **21** (2005), 149–195. 85

[5] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Liberating the weights. *J. of Complexity* **20** (2004), 593–623. 85

[6] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights. *Numer. Math.* **103** (2006), 63–97. 85

[7] F. J. Hickernell, A generalized discrepancy and quadrature error bound. *Math. Comp.* **67** (1998), 299–322. 85

[8] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimensions. *Adv. Comput. Math.* **12** (2000), 25–58. 79, 81

[9] F. Y. Kuo, Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *J. of Complexity* **19** (2003), 301–320. 84

[10] F. Y. Kuo and I. H. Sloan, Lifting the curse of dimensionality. *Notices Amer. Math. Soc.* **52** (2005), 1320–1329. 85

[11] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF Regional Conf. Ser. in Appl. Math. 63, SIAM, Philadelphia, PA, 1992. 74

[12] D. Nuyens and R. Cools, Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces. *Math. Comp.* **75** (2006), 903–920. 83

[13] S. H. Paskov and J. Traub, Faster valuation of financial derivatives. *J. Portfolio Management* **22** (1995), 113–120. 73

[14] I. H. Sloan, Finite order integration weights can be dangerous. *Comput. Methods Appl. Math.* **7** (2007), 239–254. 85

[15] I. H. Sloan and S. Joe, *Lattice methods for multiple integration*. Oxford Science Publications, Oxford, 1994. 75

[16] I. H. Sloan, F. Y. Kuo and S. Joe, On the step-by-step construction of quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces. *Math. Comp.* **71** (2002), 1609–1640. 82

[17] I. H. Sloan, F. Y. Kuo and S. Joe, Constructing randomly shifted lattice rules in weighted Sobolev spaces. *SIAM J. Numer. Anal.* **40** (2002), 1650–1665. 82, 83

[18] I. H. Sloan and A. Reztsov, Component-by-component construction of good lattice rules. *Math. Comp.* **71** (2001), 263–273. 82

[19] I. H. Sloan, X. Wang, and H. Woźniakowski, Finite-order weights imply tractability of multivariate integration. *J. of Complexity* **20** (2004), 46–74. 85

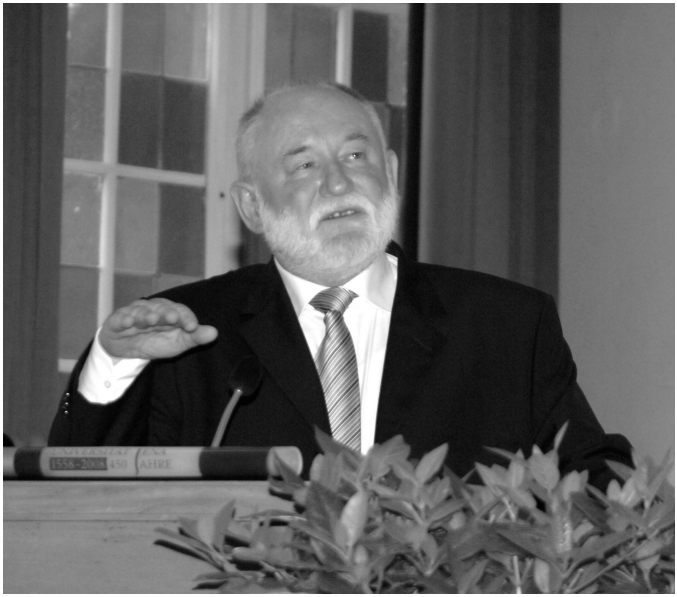
[20] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals? *J. of Complexity* **14** (1998), 1–33. 76, 77, 79, 85

[21] I. H. Sloan and H. Woźniakowski, Tractability of multivariate integration for weighted Korobov classes. *J. of Complexity* **17** (2001), 697–721. 80, 81, 82

[22] I. H. Sloan and H. Woźniakowski, Tractability of integration in non-periodic and periodic weighted tensor product Hilbert spaces. *J. of Complexity* **18** (2002), 479–499. 85

[23] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor-product algorithms for linear multivariate problems. *J. of Complexity* **15** (1999), 402–447. 85

[24] G. W. Wasilkowski and H. Woźniakowski, Finite-order weights imply tractability of linear multivariate problems. *J. Approx. Theory* **130** (2004), 57–77. 85


1555–2005 450 JAHRE

What is information-based complexity?

Henryk Woźniakowski

The purpose of this short note is to informally introduce information-based complexity (IBC). We describe the basic notions of IBC for the approximate solution of continuous mathematically posed problems.

IBC is the branch of computational complexity that studies continuous mathematical problems. Typically, such problems are defined on spaces of functions of d variables and often d is huge. Since one cannot enter a function of real or complex variables into a digital computer, the available information is usually given by finitely many function values at some prescribed sample points.[1] The sample points can be chosen adaptively, that is, the choice of the jth point may be a function of the already computed function values at the $j-1$ previously used points.

Such information is

partial, *contaminated* and *priced*.

It is *partial* since knowing finitely many function values, we cannot in general recover the function exactly and we are unable to find the exact solution of a continuous problem. It is *contaminated* since the function values are computed with model, experimental, and/or rounding errors. It is also *priced* since we are charged for each experiment leading to a function value or for each computation needed to obtain a function value. Often, it is expensive to obtain a function value. For example, some functions occurring in computational practice require thousands or millions of arithmetic operations to compute one function value.

Continuous problems for which partial, contaminated and priced information is available arise in many areas including numerical analysis, statistics, physics, chemistry and many computational sciences. Such problems can only be solved approximately to within some error threshold ε.

The goal of IBC is to create a theory of computational complexity for such problems. Intuitively, complexity is defined as the minimal cost of all possible algorithms that compute the solution of a continuous problem to within error

[1] Sometimes we may use more general information consisting of finitely many linear functionals but, for simplicity, we restrict ourselves in this note only to function values as available information.

at most ε. For many problems, the minimal cost is determined by the minimal number of function values needed for computing the solution to within ε.

Depending on precisely how the error and the cost are defined we have various settings. In the *worst case* setting, the error and cost of algorithms are defined by their worst case performance. That is, the error is the supremum of the distance between the exact solution and the approximation computed by the algorithm for all functions from a given set. The distance may be defined by a norm, or by a metric, and by a specific error criterion. For example, we may have the absolute, relative or normalized error criterion. Similarly, the cost is defined as the supremum of the costs of the algorithm for all functions from the same set of functions. The cost for a single function is equal to the sum of *information* and *combinatory* costs. The information cost is the number of function values times the cost of computing one function value. The combinatory cost is the number of all arithmetic operations needed to combine the already computed function values. Here we assume for simplicity that the cost of one arithmetic operation is taken as unity, and by arithmetic operations we mean such operations as addition, multiplication, subtraction, division and comparison of numbers. The set of permissible arithmetic operations can be extended by permitting the computation of special functions such as logarithms, exponential and trigonometric functions. The sets of functions studied in IBC are usually unit balls or whole spaces, depending on the error criterion. Typical spaces are Hilbert or Banach spaces of infinite dimension.

In the *average case* setting, the error and the cost of algorithms are defined by their average performance. That is, we assume a probability measure on a given set of functions and we take the expectation of the errors and costs of the algorithm for all functions from the given set with respect to this probability measure. The probability measures studied in IBC in the average case setting are usually Gaussian or truncated Gaussian measures defined on infinitely dimensional spaces.

In the *probabilistic* setting, the error and the cost of algorithms are defined as in the worst case setting by taking the supremum over a given set of functions modulo a subset of small measure. That is, we agree that the algorithms behave properly on a set of measure at least, say, $1 - \delta$, and we do not control their behavior on a set of measure at most δ.

In the worst case, average case and probabilistic settings discussed so far, we consider only *deterministic* algorithms. In the *randomized* setting, we also permit randomization. That is, we use *randomized* algorithms that can compute function values at randomized points, and can combine these values using also randomiza-

tion. The randomized error and the randomized cost are defined analogously as before by taking the expectation with respect to a random element of the algorithm and then the worst case, average case or probabilistic performance with respect to functions from a given set.

Quite recently, one more setting of IBC has been added. This is the *quantum setting* where computations are performed on a (still) hypothetical quantum computer. This leads to different definitions of the error and the cost of algorithms, and the quantum complexity is defined as the minimal cost needed to compute the solution of a continuous problem to within error at most ε with a given probability.

The model of computation used in IBC is the *real number model* which is consistent with the fact that most continuous problems are solved today using floating point arithmetic, usually with a fixed precision. The cost of operations in floating point arithmetic does not depend on the size of the input and modulo rounding errors is equivalent to the real number model. We simplify the IBC analysis by not considering rounding errors. Surprisingly enough, for most algorithms whose cost in the real number model is close to the complexity of a continuous problem we can find numerically stable implementation of such algorithms. Then we obtain essentially the same results in floating point arithmetic as in the real number model when we assume that the problem is not too ill-conditioned and that ε is related to the relative precision of floating point arithmetic.

Today we know rather tight bounds on the complexity of many continuous problems, and this holds in all settings we mentioned above. There are many multivariate problems whose complexity in the worst case, average case and randomized settings is exponential in d. These IBC results may be contrasted with discrete problems, such as factorization, where the exponential complexity is conjectured and not yet proved. Of course, for discrete problems we have complete information and we cannot use powerful proof techniques for partial information that essentially allow us to obtain lower bounds and prove intractability.

We now briefly explain how lower and upper bounds are obtained in IBC. For simplicity, we assume that function values can be computed exactly. We start with lower bounds since they are usually harder to obtain.

Lower bounds are possible to obtain by using so-called *adversary* arguments. That is, we want to identify two functions that are indistinguishable with respect to finitely many function values used by the algorithm, and with the most widely separated solutions. That is, they have the same function values used by the algorithm but with the maximal distance between solutions we are trying to approximate. Clearly, the algorithm cannot distinguish between these two functions

and the best it can do is to take the mean of their two solutions. So no matter how the algorithm is defined, there is no way to beat half of the distance between these two solutions. Hence, the maximal distance between the solutions for indistinguishable functions gives us a lower bound on the error. This is usually expressed as a function of n, where n is the number of function values used by algorithms. We stress that in many cases, it is quite hard to find this function of n, although we may use a whole arsenal of mathematical tools to help us to find this function. Today, there are many techniques for finding lower bounds. Just to name a few, they are based on n-widths, especially Gelfand and Kolmogorov n-widths, on decomposable reproducing kernels, on optimality of linear algorithms for linear problems, on adaption versus non-adaption techniques etc.

Upper bounds can be obtained, for example, by using so-called *interpolatory* algorithms. Namely, when we have already computed, say n function values $f(x_j)$ for $j = 1, 2, \ldots, n$, we want to find a function g belonging to the same set of functions as f, and sharing the same function values as f. That is $g(x_j) = f(x_j)$ for all $j = 1, 2, \ldots, n$, so we interpolate the data. Then the interpolatory algorithm takes the exact solution for g as the approximate solution for f. For given information by n function values at x_j, the error of the interpolatory algorithm is almost minimal since it can differ from the lower bound only by a factor of at most 2. Obviously, we still need to find optimal information, i.e., points x_j for which the error is minimal, and it is usually a hard nonlinear problem.

The cost of the interpolatory algorithm is, in general, more tricky. For some spaces and solutions, it turns out that *splines* are interpolatory and we can use vast knowledge about splines to compute them efficiently. For some spaces or solutions, it may, however, happen that the cost of an interpolatory algorithm is large. For some cases, we can use different algorithms. For example, for many linear problems, it has been proved that the error is minimized by linear algorithms. This is a vast simplification that helps enormously for the search of easy to implement algorithms with almost minimal error. The first such result for general linear functionals defined over balanced and convex sets of functions was proved by S. Smolyak in 1965. There are many generalizations of this result for linear operators but this is beyond this note.

We want to mention one more technique of obtaining upper bounds which uses a *randomization* argument, although the original problem is studied, say, in the worst case setting. It turns out that for Hilbert spaces and linear functionals, we can explicitly compute the worst case error of any linear algorithm. This worst case error obviously depends on the sample points used by the algorithm. So assume

for a moment that the sample points are independent randomly chosen points. We then compute the expected worst case error with respect to some distribution of these points. It turns out that for many cases, the expectation is small. By the mean value theorem, we know that there must be sample points for which the worst case error is at least as small as the expectation. Furthermore, using Chebyshev's inequality, we know that the measure of sample points with error exceeding the expectation by a factor larger than one is large. This proves non-constructively that good linear algorithms exist. Obviously, we now face the problem of how to construct them. There are a number of options but we want to mention only one of them for multivariate integration of d-variate functions with d in the hundreds or thousands. There is a beautiful algorithm, called the CBC algorithm, that permits the construction of n sample points component by component (so the name CBC) by using the fast Fourier transform (FFT) in time of order $n\ \ln(n)\, d$. The CBC algorithm was designed by the Australian school of I. Sloan, S. Joe, F. Kuo and J. Dick starting from 2001, and its fast implementation was proposed by R. Cools and D. Nuyens in 2006. In this way, today we approximate integrals of functions with even 9125 variables.

For many multivariate problems defined on spaces of d-variate functions, we know that the worst case complexity of computing the solution to within ε is $\Theta(\varepsilon^{-p_d})$. That is, lower and upper bounds are proportional to ε^{-p_d} with factors in the Θ notation independent of ε^{-1} but possibly dependent on d. Sometimes such estimates hold modulo a power of $\ln\ \varepsilon^{-1}$ which we omit for simplicity. The exponent p_d usually depends on the smoothness of the set of functions. If they are r times differentiable then usually

$$p_d = \frac{d}{r}.$$

Note that for fixed r and varying d, we have an arbitrarily large power of ε^{-1}. In general, we cannot, however, claim that the complexity is exponential in d since it also depends on how the factor in the upper bound in the Θ notation depends on d. For Lipschitz functions we have $r = 1$, and for multivariate integration it is known due to A. Sukharev from 1979, who used the proof technique and results of S. Bakhvalov mostly from 1959, that the complexity in the worst case setting is roughly

$$\frac{1}{e}\left(\frac{1}{2\varepsilon}\right)^d \quad \text{for } \varepsilon < \tfrac{1}{2},$$

with $e = \exp(1)$. Hence, it depends exponentially on d. If the complexity of a multivariate problem is exponential in d, we say that the problem is *intractable*, or suffers from the *curse of dimensionality* following Bellman who coined this phrase in 1957. One central issue of IBC research today is to determine for which multivariate problems and for which settings we have tractability, that is, when the complexity is *not* exponential in ε^{-1} and d. Depending on how we measure the lack of exponential dependence, we have various notions of tractability such as polynomial, strong polynomial and weak tractability. There is a huge literature on the complexity of multivariate problems. However, most of these papers and books have results which are sharp with respect to ε^{-1} but have unfortunately unknown dependence on d. To prove tractability we must establish also sharp dependence on d. Therefore tractability requires new proof techniques to obtain sharp bounds also on d. The second book in the list below is devoted to tractability of multivariate problems.

We end this note with a list of IBC books, where the reader can find more information, results and proofs on the complexity of continuous problems.

IBC Books

- E. Novak, *Deterministic and stochastic error bounds in numerical analysis.* Lecture Notes in Math. 1349, Springer-Verlag, Berlin, 1988.
- E. Novak and H. Woźniakowski, *Tractability of multivariate problems.* Volume I: Linear information, EMS Tracts Math. 6, European Mathematical Society Publishing House, Zürich, 2008.
- L. Plaskota, *Noisy information and computational complexity.* Cambridge University Press, Cambridge, UK, 1996.
- K. Ritter, *Average-case analysis of numerical problems.* Lecture Notes in Math. 1733, Springer-Verlag, Berlin, 2000.
- K. Sikorski, *Optimal solution of nonlinear equations.* Oxford University Press, 2000.
- J. F. Traub and H. Woźniakowski, *A general theory of optimal algorithms.* Academic Press, New York, 1980.
- J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information, uncertainty, complexity.* Addison-Wesley, Reading, MA, 1983.

- J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity.* Comput. Sci. Sci. Comput., Academic Press, New York, 1988.
- J. F. Traub and A. G. Werschulz, *Complexity and information.* Lezioni Lincee, Cambridge University Press, Cambridge, UK, 1998.
- A. G. Werschulz, *The computational complexity of differential and integral equations: An information-based approach.* Oxford Math. Monogr., Oxford University Press, Oxford, 1991.