EMS Tracts in Mathematics 6

Erich Novak
Henryk Woźniakowski

Tractability of Multivariate Problems

Volume I: Linear Information

European Mathematical Society

Authors:

Erich Novak
Mathematisches Institut
Universität Jena
Ernst-Abbe-Platz 2
07743 Jena
Germany

E-Mail: novak@math.uni-jena.de

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

and

Institute of Applied Mathematics
University of Warsaw
Banacha 2, 02-097
Warsaw
Poland

E-Mail: henryk@cs.columbia.edu

For (für) Gabi.

*Most likely, she will not solve
any of the open problems
but you never know.*

*For (dla) my grandson Filip Woźniakowski
and his parents Agnieszka and Artur.*

*Today, Filip is only two years old.
But we hope that one day he will
read the book and solve some of
the remaining open problems.*

Preface

This book is a comprehensive study of the tractability of multivariate problems. We present various existing results obtained by many researchers in the last years, as well as numerous new results. We hope that this book will encourage many people to study tractability and further advance our understanding of tractability. In the course of the book we present a lot of open problems and we would be delighted to see the solution of these problems in the near future.

This book consists of several volumes. The first volume begins with a general introduction to tractability, which is provided by twelve examples of particular multivariate problems, along with a survey of results in information-based complexity that are especially relevant to tractability. The rest of Volume I is devoted to tractability for the case of algorithms using linear information (arbitrary continuous linear functionals). Volumes II and III deal with algorithms using mostly function values. In Volume II we approximate functionals, whereas in Volume III we approximate general operators.

<div align="right">

Erich Novak

Henryk Woźniakowski

</div>

Contents

Chapter 1
Overview

Multivariate continuous problems are defined on spaces of functions of d variables, where d may be in the hundreds or even in the thousands. They occur in numerous applications including physics, chemistry, finance, economics, and the computational sciences. For path integration, which lies at the foundation of quantum mechanics, statistical mechanics and mathematical finance, the number of variables is infinite; approximations to path integrals result in arbitrarily large d.

Such problems can almost never be solved analytically. Since they must be solved numerically they can only be solved approximately to within a threshold ε. Algorithms for solving multivariate problems use n information operations given typically by either function values or linear functionals. Computational complexity is the study of the *minimal* resources needed to solve a problem. It is defined as the minimal number of information operations and combinatory operations needed to combine computed information operations in order to obtain the solution to within ε. The *information complexity* is defined as the *minimal* number $n(\varepsilon, d)$ of information operations needed to solve the d-variate problem to within ε. For all problems, the information complexity is a *lower* bound on the computational complexity. Surprisingly, for many computational problems, including in particular many linear multivariate problems, the information complexity is *proportional* to the computational complexity. For this reason, we concentrate on the information complexity in this book. It is studied in various settings including the worst case, average case, randomized and probabilistic settings for the absolute, normalized and relative error criteria. Depending on the setting and on the error criterion, ε has different meanings, but it always represents the error tolerance.

A central issue is the study of how the information complexity depends on ε^{-1} and d. If it depends *exponentially* on ε^{-1} or d, we say the problem is *intractable*. For many multivariate problems, we have exponential dependence on d, which is called after Bellman the *curse of dimensionality*. If the information complexity depends polynomially on ε^{-1} and d then we have *polynomial tractability*, if it depends only on a polynomial in ε^{-1} we have *strong polynomial tractability*. If the information complexity is *not* exponential in both ε^{-1} and d, then we have *weak tractability*. There are more types of tractability depending on how we measure the lack of exponential behavior.

There is a huge literature on the computational complexity of d-variate problems. Most of these papers and books study error bounds that lead to bounds on the information complexity. These bounds are usually sharp with respect to ε^{-1} but have, unfortunately, unknown dependence on d. But to determine if a problem is *tractable* we need to know the dependence on both ε^{-1} and d. Tractability requires new proof techniques to obtain sharp bounds on d. There is therefore a need to revisit even multivariate problems thoroughly studied in the past if we want to investigate their tractability.

Research on tractability of multivariate continuous problems started in 1994 and there are many surprising results. Today this subject is thoroughly studied by many people. This is the first book on tractability of multivariate continuous problems. We summarize the known results and present many new results. So far only *polynomial tractability* has been thoroughly studied in many papers. The study of more general tractability and, in particular, weak tractability has just begun. Therefore most of results on weak tractability are new. Weak tractability means that we allow a more general dependence on ε^{-1} and d, as long as it is non-exponential. This obviously enlarges the class of tractable problems.

Many multivariate problems suffer from the curse of dimensionality when they are defined over standard (unweighted) spaces. In this case, all variables and groups of variables play the same role and this causes the information complexity to be exponential in d.

But many practically important problems, such as problems in financial mathematics, are solved today for huge d in a reasonable time. One of the most intriguing challenges of the theory is to understand why this is possible. We believe the reason is that many practically important multivariate problems belong to *weighted* spaces. For weighted spaces, the dependence on the successive variables or groups of variables can be moderated by weights. We consider various weights such as product weights, order-dependent weights, finite-order and finite-diameter weights. For example, for *finite-order* weights, functions of d variables can be represented as sums of functions of ω variables, where ω is fixed and moderate, and d varies and can be arbitrarily large. For finite-order weights, most multivariate problems are polynomially tractable.

Multivariate problems may become weakly tractable, polynomially tractable or even strongly polynomially tractable if they are defined over *weighted* spaces with properly decaying weights. One of the main purposes of this book is to study weighted spaces and obtain necessary and sufficient conditions on weights for various notions of tractability.

The tractability results are illustrated for many specific multivariate problems. We consider general linear problems including multivariate integration, approximation, as well as a number of specific non-linear problems such as partial differential or integral equations, including the Schrödinger equation. Some of these applications will be presented in Volumes II and III.

The book contains a number of open problems, including the 15 open problems in Chapter 3, and the other 15 open problems in the remaining chapters. They should be of interest to a general audience of mathematicians. Volume I of the book contains a bibliography of over 290 papers and books, whereas Volume II will have additionally about 150 papers and books. We hope that the book will further intensify research on tractability of multivariate continuous problems.

Chapter 2
Motivation for Tractability Studies

High dimensional multivariate continuous problems occur in many applications:

- *High dimensional integrals* or *path integrals* with respect to the *Wiener measure* have many important applications, especially in physics and in mathematical finance. High dimensional integrals also occur when we want to compute certain parameters of stochastic processes, see Müller-Gronbach and Ritter [152]. Moreover, path integrals arise as solutions of partial differential equations given, for example, by the *Feynman–Kac formula*, see e.g., Gerstner and Griebel [62], Kwas [127], Kwas and Li [128], Morokoff [151], Tezuka [234], the book of Traub and Werschulz [243], as well as [197], [270]. We will study high dimensional integration in many chapters of this book.

- *Global optimization*, where we wish to compute the (global) minimum of a function of d variables. This occurs in many applications, for example, in *pattern recognition* and in *image processing*, see the book of Winkler [283], or in the modeling and prediction of the geometry of proteins, see Neumayer [157]. Simulated annealing strategies and genetic algorithms are often used, as well as smoothing techniques and other stochastic algorithms, see Boender and Romeijn [15] and Schäffler [208]. Some error bounds for deterministic and stochastic algorithms can be found in Nemirovsky [154], and in the books of Nemirovsky and Yudin [155], and Nesterov and Nemirovsky [156], as well in the book [160] and in [170]. We return to global optimization in the next volumes.

- The *Schrödinger equation* for m particles in $\mathbb{R}^3$ is an example of a $d = 3m$-dimensional problem. Since m is often large, d is even larger, see the books of Atkins and Friedman [1], Levine [132], and Messiah [144], the computational survey of Le Bris [21], as well as Flad, Hackbusch and Schneider [59] and Yserentant [295]. We illustrate general tractability results for the Schrödinger equation, in Chapters 5 and 6 we consider the linear case, and in the next volumes the nonlinear case.

These problems are all defined on spaces of d-variate functions and d can be huge – in the hundreds or even in the thousands! For path integration, the number of variables is infinite, and approximations of path integrals yield arbitrarily large d.

Some high dimensional problems, such as convex optimization and systems of ordinary differential equations, can be solved efficiently, i.e., their cost increases polynomially in d, see e.g., Nemirovsky and Yudin [155], and Nesterov and Nemirovsky [156] for convex optimization, and Kacewicz [105] for systems of ordinary differential equations. However, there are many other problems (including the ones mentioned

above) for which it appears that the cost increases exponentially in d. This exponential dependence on d is called *intractability* or the *curse of dimensionality*; the latter notion goes back to Bellman [9].

For example, the minimal error for integration of C^r-functions in one variable is of order n^{-r}, where n is the number of function values used by the algorithm. Using product rules, it is easy to obtain an error of order $n^{-r/d}$ for functions of d variables. It is known that the exponent r/d is best possible. Hence, if we want to guarantee an error ε we must take n of order $\varepsilon^{-d/r}$. For fixed regularity r, this is an exponential function of d. For large d, we cannot afford to sample the function so many times.

A central problem is to investigate which multivariate continuous problems are *tractable* and which are *intractable*. There are different types of tractability such as weak, polynomial and strong polynomial tractability. The essence of all these types of tractability is that the cost of the computational problem that we want to solve does *not* grow exponentially in terms of its input parameters.

In contrast, tractability of discrete decision problems means that the cost is polynomial in terms of the number of input bits, and the Turing machine is used as a model of computation. It is conjectured that numerous such problems are intractable. This is the essence of the famous conjecture P $\neq$ NP. According to Smale [221], the question of whether P $\neq$ NP is one of the three most important problems in mathematics[1].

We deal with multivariate continuous problems and we use the *real number model with oracles* as the model of computation[2]. In this model we assume that we can compute arithmetic operations over the reals exactly and compare real numbers. We can also compute function values or linear functionals as information operations. Information operations are sometimes called *oracles* or black box computations. This model is typically used for numerical and scientific computations, since it is an abstraction of fixed precision floating point arithmetic. The real number model is used for algebraic problems, see e.g., Coppersmith and Winograd [32], Pan [186], Strassen [226] for the famous matrix multiplication problem and the books of Bini and Pan [12], Bürgisser, Claussen and Shokrollahi [23] for more general algebraic problems, for computational geometry problems, see e.g., the book of Preparata and Shamos [199], and for information-based complexity problems, see e.g, [164] and the books [242], [244]. The real number model without oracles was formalized by Blum, Shub and Smale [13] and the famous P $\neq$ NP conjecture was, in particular, extended there for real numbers. Oracles were introduced in [164]. Pros and cons of the real number

---

[1] Smale wrote in 1998: "In fact, included are what I believe to be the three greatest open problems of mathematics: the Riemann Hypothesis, the Poincaré Conjecture, and 'Does P = NP?' ". The Poincaré Conjecture was recently established by G. Perelman.

[2] Some problems can be analyzed in both the Turing machine model and in the real number model. Which model is more appropriate depends on the application. A problem can be tractable in one model and intractable in the other, because the assumptions concerning cost and size of a problem are quite different in both models. For example, it is well known that the problem of linear optimization is tractable with respect to the Turing model. It is still unknown, and it is an important open problem, whether the same is true with respect to the real number model, see [245].

We stress that a P $\neq$ NP conjecture has been defined also over the real (or the complex) numbers, see Blum, Shub and Smale [13], and the book of Blum, Cucker, Shub and Smale [14].

model versus the Turing number model can be found in Traub [239]. More about the real number model can be found in Meer and Michaux [143], Weihrauch [274] and in [174], [175], [176], [290], as well as in the book of Blum, Cucker, Shub and Smale [14]; see also Chapter 4.

The subject of tractability for multivariate continuous problems was introduced in [287] and [288]. Today this topic is being thoroughly studied by many researchers in information-based complexity. Numerous tractability results have been recently obtained, and they will be reported in this book.

Multivariate continuous problems are defined on *classes (spaces) of d-variate functions*. We want to compute the solution to within ε. These two parameters ε and d are used in tractability studies. Following the approach used in the study of discrete decision problems it would be tempting to define tractability of a continuous problem when the cost is of order

$$\left(1 + \lceil \log_2 \varepsilon^{-1} \rceil\right)^p + (1 + \lceil \log_2 d \rceil)^q \quad \text{or} \quad \left(1 + \lceil \log_2 \varepsilon^{-1} \rceil\right)^p \left(1 + \lceil \log_2 d \rceil\right)^q$$

for some non-negative p and q. This is so because d requires $1 + \lceil \log_2 d \rceil$ bits for its representation, and assuming without loss of generality that ε^{-1} is a positive integer, ε requires $1 + \lceil \log_2 \varepsilon^{-1} \rceil$ bits.

More generally, we can argue that for d-variate functions we need to work with vectors of d components and the number of bits needed for d numbers to be represented to within ε is proportional to $d(1 + \lceil \log_2 \varepsilon^{-1} \rceil)$. Hence, tractability of a continuous problem should require that the cost is of order

$$\left(d \left(1 + \lceil \log_2 \varepsilon^{-1} \rceil\right)\right)^p$$

for some non-negative p.

Indeed, there are continuous problems for which the cost is of this form. Examples include convex optimization, fixed point problems, various zero finding problems for which the bisection or Newton-type algorithms work, as well as large linear systems with well conditioned matrices.

Unfortunately these definitions would imply that almost all *linear* continuous problems arising in computational practice are intractable, since even for the univariate case ($d = 1$) the cost is proportional to some power of ε^{-1}. To include this typical behavior of the cost of linear continuous problems, we usually say that a problem is tractable when the cost can be bounded by a multiple of a power of $T(\varepsilon^{-1}, d)$ for some function T.

More precisely, we take a function $T : [1, \infty) \times [1, \infty) \to [1, \infty)$ which is non-decreasing in both arguments and which is *not* exponential, i.e.,

$$\lim_{x+y \to \infty} \frac{\ln T(x, y)}{x + y} = 0.$$

Then we have T-tractability if the information complexity $n(\varepsilon, d)$ can be bounded by a polynomial of $T(\varepsilon^{-1}, d)$, see [69], [70], [71].

For $T(x, y) = (1 + \ln x)(1 + \ln y)$ we have the case often studied in theoretical computer science in which tractability means polynomial dependence on the number of input bits.

For $T(x, y) = xy$ we obtain *polynomial tractability*, and for

$$T(x, y) = \exp((1 + \ln x)(1 + \ln y))$$

we permit super-polynomial growth of the information complexity.

Furthermore, for T-tractability, also called *generalized tractability*, we can restrict the domain of ε^{-1} and d, and consider the case when only one of them goes to infinity. This seems especially well-suited for problems appearing in computational finance, where d is huge but it is enough to solve the problem with a relatively large error tolerance ε. This case corresponds to *restricted tractability*.

Multivariate continuous problems are given as operators defined on classes of d-variate functions that enjoy some degree of smoothness. We approximate these problems with an error at most ε. We will consider different types of errors in this book. We concentrate on the absolute, normalized and relative errors in the worst case, average case, probabilistic and randomized settings. Some of these notions will be studied in Volume II. We shall see that the tractability results depend crucially on the type of error and on the setting.

There is a huge literature on the complexity of multivariate problems. The typical approach is to fix d and consider the best possible rate of convergence of algorithms that use n information operations given usually by function values or linear functionals. Assume for simplicity that the optimal rate is p_d and we have an algorithm whose worst case error[3] is $e(n, d) = \Theta(n^{-p_d})$ for all n. More precisely, this means that for $d = 1, 2, \ldots$, there exist two positive numbers $C_{d,1}$ and $C_{d,2}$, such that

$$C_{d,1} n^{-p_d} \leq e(n, d) \leq C_{d,2} n^{-p_d} \quad \text{for all } n = 1, 2, \ldots. \tag{2.1}$$

The optimal *order* (or *rate*, and we will be using these two words interchangeably) of convergence p_d depends on d and is often of the form $p_d = r/d$, where r is a measure of the smoothness of the class of functions. The factors $C_{d,1}$ and $C_{d,2}$ in (2.1) depend on d, but this dependence has usually *not* been previously studied.

The essence of different types of tractability is that the *minimal* number of information operations needed to solve the problem to within ε must *not* be exponential in ε^{-1} and d. This minimal number, which we denote as $n(\varepsilon, d)$, is the *information complexity* of the problem. Since the total cost of computing the solution to within ε is often proportional to $n(\varepsilon, d)$, the information complexity $n(\varepsilon, d)$ is then proportional to the total complexity of the problem, and it is enough to study $n(\varepsilon, d)$. We do so in this book; see Section 4.1.5 of Chapter 4 for more details.

---

[3]For many multivariate problems the situation is a little more complicated since the worst case error of an optimal algorithm is of order $n^{-p_d} (\ln n)^{q_d}$ with q_d usually depending linearly on d. This means that (2.1) holds modulo some powers of logarithms. Then for any positive δ, one can replace the exponent p_d by $p_d + \delta$ in the left hand side of (2.1) and by $p_d - \delta$ in the right hand side of (2.1) and appropriately change $C_{d,1}$ and $C_{d,2}$. Since the presence of logarithms is not important for our discussion, we simplify the situation by assuming that $q_d = 0$.

Hence, we want to guarantee that $n(\varepsilon, d)$ is asymptotically much smaller than $a^{\varepsilon^{-1}+d}$ for any $a > 1$. This means that a *necessary* condition on tractability is

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0. \tag{2.2}$$

If condition (2.2) holds then we say that the problem is *weakly tractable*, whereas if (2.2) is *not* satisfied then we say that the problem is *not weakly tractable* or simply that it is *intractable*.

We stress that the concept of weak tractability is new and has been so far only studied in [70], [71] at the same time as this book was written. Most results on weak tractability that will be presented in this book are new.

There are many ways to measure the lack of exponential dependence, and so there are many specific types of tractability. The most commonly studied case is *polynomial* tractability, for which we want to guarantee that $n(\varepsilon, d)$ is polynomially bounded in ε^{-1} and d, i.e., that there exist three non-negative numbers C, p and q such that

$$n(\varepsilon, d) \le C \varepsilon^{-p} d^q \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d = 1, 2, \ldots. \tag{2.3}$$

If the condition (2.3) holds then we say that the problem is *polynomially tractable*, whereas if (2.3) is *not* satisfied we say that the problem is *polynomially intractable*.

We shall see in Chapter 5 that any function $n \colon \mathbb{R}^+ \times \mathbb{N} \to \mathbb{N}_0$ that is non-increasing in the first variable is the information complexity of a suitable problem. An example of such a function is

$$n(\varepsilon, d) = \left\lceil \inf_{\alpha \in [0,100]} \varepsilon^{\alpha - 100} \cdot d^\alpha \right\rceil.$$

This problem is polynomially tractable but of course the exponents p and q are not uniquely defined since we may take $q = \alpha$ and $p = 100 - \alpha$ for any $\alpha \in [0, 100]$. Hence, there is a trade-off between the exponents p and q for polynomial tractability. Another possible example would be

$$n(\varepsilon, d) = \left\lceil 2^d \cdot \ln(1 + \varepsilon^{-1}) \right\rceil.$$

Here we have an excellent order of convergence for any fixed dimension d, but the problem is intractable.

There are obviously functions that go faster to infinity than any polynomial but slower than any exponential function. For example, if

$$n(\varepsilon, d) = \left\lceil \exp\left(\ln(1 + \varepsilon^{-1}) \, \ln(1 + d) \right) \right\rceil \quad \text{or} \quad n(\varepsilon, d) = \left\lceil \exp\left(\sqrt{\varepsilon^{-1}} + \sqrt{d} \right) \right\rceil,$$

then the problem is polynomially intractable but it is weakly tractable. We will study tractability in full generality. That is, we estimate $n(\varepsilon, d)$ by a multiple of a power of $T(\varepsilon^{-1}, d)$ for a general T which is not exponential as explained before. However, our emphasis will be on polynomial tractability, and we postpone the study of generalized tractability to Chapter 8.

There is also the notion of *strong* tractability when $n(\varepsilon, d)$ can be bounded uniformly in d by a non-exponential function depending only on ε^{-1}. For example, *strong polynomial* tractability means that (2.3) holds with $q = 0$. The notion of strong tractability seems very demanding and one might suspect that only trivial problems are strongly tractable. Indeed, many problems defined over classical function spaces are intractable, much less strongly tractable. Sometimes, such problems become strongly polynomially tractable or polynomially tractable over function spaces with suitable weights that moderate the behavior with respect to successive variables or groups of variables. Such spaces are called *weighted* function spaces.

We stress that weighted spaces arise naturally in many applications. In computational finance, weights appear due to the discounted value of money, and d-variate functions depend on successive variables in a diminishing way, see the book of Traub and Werschulz [243]. In computational physics, weights appear due to the fact that usually only the influence of neighboring particles is significant, and functions can be represented as sums of functions of a few variables, see Coulomb pair potentials discussed a little later. In computational economics, weights appear when the Cobb-Douglas condition is used for the Bellman fixed point problem. This condition guarantees equal partitioning of goods, see [204]. In computational chemistry there is the need to construct poly-atomic potential energy surfaces which underlie molecular dynamics and spectroscopies. It is observed that functions depending on many variables often can be well approximated by a sum of functions that depend on only few variables, see Ho and Rabitz [99] and Rabitz and Alis [200].

We believe that the reason so many high dimensional problems are solved efficiently in computational practice is that these problems belong to weighted spaces for which tractability holds.

We now briefly discuss two approaches how multivariate problems are typically solved and indicate why these approaches are not good for tractability.

Suppose we deal with a class of d-variate functions with huge d, say $d = 1000$. If we assume that the class of functions is a classical (unweighted) space then we only know the global smoothness of functions, which is usually not enough to prevent intractability. Indeed, one of the typical approaches of solving multivariate problems is to use discretization of the domain of functions. If we use at least two points in each direction than we have at least $2^d = 2^{1000}$ subproblems and if each subproblem requires at least one information operation, we need at least 2^{1000} information operations, which really shows intractability in action! Only algorithms that use information operations not based on grids may have a computing time that does not increase exponentially with the dimension. As we shall see in Volumes II and III, we may indeed achieve even strong polynomial tractability for many multivariate problems if information operations are from *sparse* grids and we use the *weighted Smolyak* algorithm. The reader is referred to the recent survey of Griebel [72] where many papers on sparse grids and Smolyak's algorithm can be found.

Another popular approach is to guarantee that multivariate problems are exactly solved for some special functions, for instance, for multivariate polynomials of degree

at most k. The dimension of the latter space is $\binom{d+k}{k}$, which for fixed k behaves as $d^k/k!$, so at least this many information operations would usually be needed. Even for a relatively small k, say $k = 5$ and $d = 1000$, we would need to perform at least $\binom{1005}{5} \approx 8.46 \cdot 10^{12}$ information operations, which would be quite a job. In general, the second approach would contradict strong polynomial tractability but not tractability as long as k is fixed and does not depend on d.

In any case, it is clear that tractability requires a different approach for choosing proper classes of functions, as well as a different approach for choosing proper algorithms.

To put it simply, the classical spaces seem to be too large for large d, which explains why we have so many intractability results. We believe that multivariate problems of practical importance with large d are defined on spaces of functions enjoying *additional* properties. For instance, functions may depend on successive variables or some groups of variables in a diminishing way. So we may have functions of d variables approximately equal to a sum of functions, each depending on ω variables with ω independent of d. Or we might have functions for which the dependence on the first variable is more important than the dependence on the second variable, and the dependence on the jth variable is more important than on the $j + 1$st variable, and so on, assuming that the variables are properly labeled.

For the reader who is familiar with the ANOVA decomposition of functions, we may add that an arbitrary function f of d variables from the space L_2 can be decomposed as the sum of 2^d mutually orthogonal functions $f_{\mathfrak{u}}$, with $\mathfrak{u}$ being a subset of $\{1, 2, \ldots, d\}$ and $f_{\mathfrak{u}}$ depending only on variables from $\mathfrak{u}$; see Section 3.1.6 of Chapter 3, where the ANOVA decomposition is discussed. Then some $f_{\mathfrak{u}}$ may be negligible or even 0; alternatively only terms $f_{\mathfrak{u}}$ for which the set $\mathfrak{u}$ has a small cardinality are important.

Weighted function spaces play a major role in the study of tractability of multivariate problems. Weighted function spaces allow us to model problems for which some variables and groups of variables are more important than others, see [216] where this concept was first introduced for the study of tractability. The weighted space of d-variate functions depends on 2^d weights $\gamma_{d,\mathfrak{u}}$, with $\mathfrak{u}$ being a subset of $\{1, 2, \ldots, d\}$. Each weight $\gamma_{d,\mathfrak{u}}$ measures the importance of the set of variables x_j with $j \in \mathfrak{u}$. If we take the unweighted case $\gamma_{d,\mathfrak{u}} \equiv 1$, then we are back to the classical spaces, in which all variables and groups of variables play the same role. It is now natural to seek conditions on the weights $\gamma_{d,\mathfrak{u}}$ that are necessary and sufficient for weak, polynomial or strongly polynomial tractability. This will be one of the major subjects of our book. Not surprisingly, the case $\gamma_{d,\mathfrak{u}} \equiv 1$ often leads to polynomial intractability or even intractability, whereas for weights that decay sufficiently fast, we obtain polynomial tractability or even strong polynomial tractability.

In particular, we want to mention *finite-order weights*, which are defined by assuming that $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u}$ of cardinality greater than ω, with ω independent of d, see [45] where this concept was first introduced. That is, finite-order weights guarantee that functions can be decomposed as sums of functions depending on at most ω variables. An example is a polynomial of d variables whose order is at most ω with ω indepen-

dent of d. Another example of such a function is a sum of Coulomb pair potentials of $d = 3m$ variables, $f(x) = \sum_{1 \le i < j \le m} \|x_i - x_j\|^{-1}$ for vector $x = [x_1, x_2, \ldots, x_m]$ with $x_j \in \mathbb{R}^3$, and $\| \cdot \|$ denoting the Euclidean norm, see e.g., the book of Glimm and Jaffe [64]. Since f is not defined for $x_i = x_j$, we typically modify it by taking a small positive η and considering $f_\eta(x) = \sum_{1 \le i < j \le m} (\|x_i - x_j\|^2 + \eta)^{-1/2}$. Hence, f and f_η only depend on groups of two variables, each of which is a 3-dimensional vector. This corresponds to finite-order weights with $\omega = 6$. Such functions are called "partially separable" in optimization, see e.g., Griewank and Toint [73].

Finite-order weights have been analyzed in a number of papers. Typical results are positive, and we find that multivariate problems defined over weighted spaces with finite-order weights are polynomially tractable. Furthermore, we can achieve strong polynomial tractability if the finite-order weights satisfy some additional conditions. For finite-order weights, the minimal number $n(\varepsilon, d)$ of information operations depends, in particular, on $C^\omega d^{c\omega}$ for some $C > 1$ and a positive c. Hence, we have exponential dependence on the order ω of finite-order weights. Since ω is the same for all d, this does not matter, and C^ω is just a number. The order ω also effects the degree of the polynomial in d. Such estimates are especially useful if ω is relatively small.

We will also thoroughly study *product weights*. For these weights, we have $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ for non-increasing non-negative weights $\gamma_{d,j}$ that may sometimes even be independent of d, i.e., $\gamma_{d,j} = \gamma_j$. That is, $\gamma_{d,j}$ controls the importance of the variable x_j, and the groups of variables are controlled by the product of the weights of variables from the group. The smaller the weight $\gamma_{d,j}$ the less important the variable x_j. In the limiting case, if one sets $\gamma_{d,j} = 0$ for $j > d_0$, the functions depend only on at most the first d_0 variables. Typical results for product weights are that for some positive τ,

$$\limsup_{d \to \infty} \sum_{j=1}^{d} \gamma_{d,j}^\tau < \infty$$

is needed for strong polynomial tractability, whereas

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

is needed for polynomial tractability. The last condition is significantly relaxed for weak tractability, which usually holds if

$$\lim_{d \to \infty} \frac{\sum_{j=1}^{d} \gamma_{d,j}^\tau}{d} = 0.$$

The value of τ depends on the particular multivariate problem. For linear functionals such as multivariate integration studied in Volume II, typically $\tau = 1$.

Note that the unweighted case $\gamma_{d,j} \equiv 1$, with all variables playing the same role, is intractable. For $\tau = 1$ and decaying weights of the form $\gamma_{d,j} = j^{-k}$ we have strong

polynomial tractability if $k > 1$, polynomial tractability if $k = 1$, and polynomial intractability if $k < 1$. The subject of weighted spaces will be studied in detail, paying special attention to finite-order and product weights.

We explain relations between the optimal order of convergence and tractability. Suppose that (2.1) holds. Then the optimal rate of convergence is p_d, and $n(\varepsilon, d) = \Theta(\varepsilon^{-1/p_d})$ for all ε, or more precisely

$$C_{d,1}^{1/p_d} \varepsilon^{-1/p_d} \le n(\varepsilon, d) \le C_{d,2}^{1/p_d} \varepsilon^{-1/p_d} \quad \text{for all } \varepsilon \in (0, 1).$$

Note that we often know the optimal (smallest) exponent of ε^{-1}. However, since $C_{d,1}$ and $C_{d,2}$ are not known, we do not know the dependence on d. For example, suppose that $p_d = r/d$. Then $n(\varepsilon, d) = \Theta(\varepsilon^{-d/r})$. Hence for a fixed r, the exponent d/r is arbitrarily large for large d. This means that we have *polynomial intractability*. We stress that since the optimal exponent is $p_d = r/d$, we are able to conclude polynomial intractability without knowing how $C_{d,1}$ and $C_{d,2}$ depend on d. From this point of view, many classical works on multivariate problems are useful for polynomial tractability studies.

Suppose now that the optimal order of convergence is independent of d, i.e., $p_d = p$. Can we now claim polynomial or weak tractability? No, since the error depends on the factors $C_{d,1}$ and $C_{d,2}$ which, in turn, are often unknown functions of d. Indeed, if $C_{d,2} \le Cd^s$ then $n(\varepsilon, d) \le \lceil C^{1/p} d^{s/p} \varepsilon^{-1/p} \rceil$ and polynomial tractability holds. On the other hand, if $C_{d,1} \ge Cc^d$ with $c > 1$, then $n(\varepsilon, d) \ge C^{1/p} c^{d/p} \varepsilon^{-1/p}$ and we have intractability. Hence, even if p_d is independent of d, these results are too weak for tractability studies.

We now turn to weak tractability, assuming that we know that the optimal order of convergence is $p_d = r/d$ and polynomial intractability holds. Can we also claim that the problem is intractable? The answer again depends on how the factors $C_{d,1}$ and $C_{d,2}$ in (2.1) depend on d. Indeed, assume that $C_{d,1} \ge C$ for all d for some positive C. Then $n(\varepsilon, d) \ge C^{d/r} \varepsilon^{-d/r}$. Since (2.2) does not hold, the problem is intractable. On the other hand, assume that $C_{d,2} \le Cc^d$ for all d for some positive C and $c \in (0, 1)$. Then $n(\varepsilon, d) = \lceil C^{d/r} c^{d^2/r} \varepsilon^{-d/r} \rceil$. Since (2.2) holds, the problem is weakly tractable. This means that knowing the optimal order of convergence does *not* tell us whether the problem is weakly tractable or intractable. We again see that the known results are too weak for tractability studies.

This short discussion tells us that tractability studies require new analysis. Even for multivariate problems thoroughly studied in the literature, we can rarely conclude whether they are weakly or polynomial tractable. Therefore we need to revisit many of such problems and seek sharp error estimates in terms of *both ε and d*. We will use several specific multivariate problems to illustrate this point in Sections 3.1 and 3.2 of the introductory Chapter 3.

The tractability of a multivariate problem depends on the class of functions. In this book, we will present classes of functions with *four types of results*:

- Classes of functions for which a multivariate problem is intractable and suffers from the curse of dimensionality, or is polynomially intractable. This will be

done by establishing exponential lower bounds in d or arbitrarily large degree polynomial lower bounds in d on the number of function values or linear functionals.

- Classes of functions for which a multivariate problem is weakly, polynomially or strongly polynomially tractable by proving the *existence* of suitable algorithms by *non-constructive* arguments.

- Classes of functions for which a multivariate problem is weakly, polynomially or strongly polynomially tractable by *semi-construction* of suitable algorithms.

- Classes of functions for which a multivariate problem is weakly, polynomially or strongly polynomially tractable by *construction* of suitable algorithms.

For problems belonging to the first type we *prove* intractability. This is in contrast to discrete complexity theory, where it is often only *believed*, not proved, that certain problems are intractable.

For problems belonging to the second type, we only know that suitable algorithms exist, but we do not know how to construct them. This existence is usually established by using a probabilistic argument for some class of algorithms enjoying tractability error bounds. By a tractability error bound we mean, for instance, that the error of an algorithm that uses n information operations is bounded by $cd^a n^{-b}$ for some a, b and c independent of d and n and with a positive b. Then indeed, its error is at most ε if $n = \lceil Cd^q \varepsilon^{-p} \rceil$, with $C = c^{1/b}$, $q = a/b$ and $p = 1/b$, which means that we have polynomial tractability.

For problems belonging to the third type we know *semi-construction* of an algorithm with tractability error bounds. By semi-construction, we have two possibilities in mind:

- We know how to construct such an algorithm only if d and ε^{-1} are *not* too large. This is related to the cost of *precomputing* which is prohibitively expensive if either d or ε^{-1} are too large. An illustration of this situation can be found in [197] for the approximation of Feynman–Kac path integrals, where *many* complicated multivariate integrals need to be precomputed. These integrals are computed by the use of the Monte Carlo algorithm, which is feasible only for relatively small d and ε^{-1}.

- We know how to construct such an algorithm probabilistically. More specifically, we explain this construction for the worst case setting. Usually, we need to construct a linear algorithm that uses function values at some points. The major problem is to find these points. We first randomly select n points with the distribution depending on a given multivariate problem. Then we compute the worst case error of the linear algorithm that uses these points and check if a tractability error bound holds. If so, we are done; if not, we repeat this process. Usually we can prove that after k trials, the probability of failure is exponentially small in k, and therefore after a few trials we have a high probability of finding suitable points. This approach requires that we know

– how to select randomized points with a given distribution. This can usually be achieved only approximately by using a pseudo-random generator for the uniform distribution and by the corresponding transformation of uniformly distributed points, see e.g., Box and Muller [18].

– how to compute the worst case error of a linear algorithm relatively quickly which is the case for some multivariate problems defined over reproducing kernel Hilbert spaces. It is also sometimes the case for more general spaces if again d and ε^{-1} are not too large.

For the fourth class of problems, we know how to construct algorithms with tractability error bounds. Only in this case we can claim that we know how to solve efficiently multivariate continuous problems.

The early papers in this area often provide non-constructive proofs of tractability. There is much more emphasis later on constructive proofs by designing algorithms whose errors achieve tractability error bounds. For example, there is the *CBC* (*component-by-component*) algorithm designed by the Australian school of Ian Sloan, and the weighted tensor product algorithm, also known as the weighted Smolyak algorithm from [271]. First the CBC algorithm was used for multivariate integration and later for multivariate L_2 and L_∞ approximation in the worst case and average case settings, see Dick [42], Dick and Kuo [43], [44], Kuo [119], Sloan, Kuo and Joe [213] for multivariate integration, and [121], [122], [124], [125] for multivariate approximation. The implementation cost of the CBC algorithm was also significantly reduced due to work of Nuyens and Cools [180], [181], and today is linear (modulo a logarithm) in the number of input data. The weighted Smolyak algorithm can be applied to any linear tensor product problem. We will be analyzing these algorithms in Volumes II and III of the book.

Still, there are many non-constructive tractability results. An important area for tractability research is to continue to move problems from the second and third to the fourth class. That is, we want to provide *constructive* algorithms for problems for which we currently have only existence or semi-construction. Obviously, we are especially interested in a simple construction that does not involve extensive precomputations.

2.1 Notes and Remarks

NR 2:1. Chapter 2 is partially based on the survey [177] as well as on [294]. The concepts of weak tractability and intractability are new and will be studied in the course of this book. Relations between the optimal order of convergence and tractability are also new, although quite straightforward.

Chapter 3
Twelve Examples

This introductory chapter informally introduces the concept of tractability for multivariate continuous problems. We will illustrate tractability by a number of representative examples that have already been thoroughly studied or for which the analysis of tractability is relatively easy. We also hope that these examples will help the reader to develop the proper intuition needed for tractability studies. Section 3.1 deals with the worst case setting, whereas Section 3.2 deals with the average case, probabilistic and randomized settings. In Section 3.3 we present 15 open problems related to the tractability of problems studied in the first two sections.

We add that the 12 examples presented in this chapter can be read in any order. Furthermore, the reader who is more interested in a general study of tractability may skip this chapter and go directly to

- Chapter 4, where we survey information-based complexity results relevant for tractability study, or even may go to
- Chapter 5, where tractability is studied in the worst case or to
- Chapter 6, where tractability in the average case is considered.

3.1 Tractability in the Worst Case Setting

In this section, we illustrate tractability results by presenting several examples in the worst case setting, in which the error and cost of an algorithm is defined by its worst performance with respect to functions from a given class and the cost (complexity) of a multivariate problem is defined as the minimal cost of algorithms that approximate the solution to within ε. As will be explained in this section and more thoroughly in Chapter 4, for many multivariate problems the complexity is proportional to the minimal number $n(\varepsilon, d)$ of function values or linear functionals needed for computing the solution to within ε, and that is why we concentrate on $n(\varepsilon, d)$.

3.1.1 Example 1: Integration of Lipschitz Functions

Consider the class of functions

$$F_d^{\mathrm{lip}} = \left\{ f : [0, 1]^d \to \mathbb{R} \mid \|f\| < \infty \right\},$$

where the norm of f is defined by

$$\|f\| = \max \left(\sup_{x \in [0,1]^d} |f(x)|, \quad \sup_{x,y \in [0,1]^d} \frac{|f(x) - f(y)|}{\|x - y\|_\infty} \right)$$

with $\|x\|_\infty = \max_{j=1,2,\ldots,d} |x_j|$ for $x = [x_1, x_2, \ldots, x_d]$. Hence, F_d^{lip} consists of Lipschitz functions, and its unit ball consists of functions whose absolute values and the Lipschitz constant are bounded by 1.

Many continuous problems have been studied for the class F_d^{lip}. Examples include integration, approximation, optimization, ordinary or partial differential equations, and integral equations. For brevity, we restrict ourselves to integration. Let

$$\text{INT}_d f = \int_{[0,1]^d} f(x) \, dx \quad \text{for all } f \in F_d^{\text{lip}}.$$

We want to approximate $\text{INT}_d f$ to within ε. Algorithms for approximating INT_d must use finitely many function values, and they are of the form

$$A_n(f) = \varphi_n \left(f(x_1), f(x_2), \ldots, f(x_n) \right)$$

for some linear or non-linear mapping $\varphi_n : \mathbb{R}^n \to \mathbb{R}$ and for some points $x_j \in [0, 1]^d$. The points x_j may be chosen adaptively, i.e., x_j may depend on the already computed x_k and $f(x_k)$ for $k = 1, 2, \ldots, j - 1$. The worst case error of A_n is defined as

$$e^{\text{wor}}(A_n) = \sup_{f \in F_d^{\text{lip}}, \|f\| \le 1} |\text{INT}_d f - A_n(f)|.$$

We want to find the smallest n for which the error is at most ε, i.e.,

$$n^{\text{wor}}(\varepsilon, \text{INT}_d, F_d^{\text{lip}}) = \min \{n \mid \text{there exists } A_n \text{ with } e^{\text{wor}}(A_n) \le \varepsilon\}.$$

In the previous chapter, $n^{\text{wor}}(\varepsilon, \text{INT}_d, F_d^{\text{lip}})$ was simply denoted by $n(\varepsilon, d)$. We now use a more accurate notation in which we mention the worst case setting by the superscript "wor" and the multivariate integration problem for the class F_d^{lip} by replacing "d" by "$\text{INT}_d, F_d^{\text{lip}}$".

We stress that $n^{\text{wor}}(\varepsilon, \text{INT}_d, F_d^{\text{lip}})$ can be regarded as the minimal cost (complexity) of solving this integration problem. Indeed, from general results that we present in Chapter 4, we know that adaptive choice of points x_j as well as non-linear mappings φ_n do not help. That is, an algorithm with the minimal n among all algorithms having worst case error at most ε is linear. Such an algorithm requires us to compute n function values; its evaluation can then be done by performing at most $2n$ arithmetic operations. Since the cost of one function evaluation is usually much larger than the cost of one arithmetic operation, the minimal cost is therefore proportional to $n^{\text{wor}}(\varepsilon, \text{INT}_d, F_d^{\text{lip}})$.

The integration problem for Lipschitz functions has been thoroughly studied in the literature. Already in 1959, Bakhvalov [7] proved that the optimal order of convergence is $n^{-1/d}$, i.e.,

$$e(n, d) := \inf_{A_n} e^{\text{wor}}(A_n) = \Theta\left(n^{-1/d}\right),$$

as $n \to \infty$.[1] As we already know, this implies polynomial intractability, but weak tractability is an open question until we know how the factors in the big theta notation depend on d. If we examine Bakhvalov's proof of the lower and upper bounds then we realize that there are two numbers

$$0 < c < 1 < C$$

such that

$$c^d n^{-1/d} \le e(n, d) \le C^d n^{-1/d} \quad \text{for all } n.$$

That is, the lower bound is exponentially small in d, whereas the upper bound is exponentially large in d. Obviously, if the upper bound is sharp then we have intractability. However, if the lower bound is sharp then we have weak tractability. Hence, these two estimates are too weak in their dependence on d, and based on them we cannot say whether the integration problem for Lipschitz functions is intractable or weakly tractable.

Fortunately, Sukharev [227] provided an explicit formula for $e(n, d)$ if $n = m^d$ for some integer m. We have

$$e(n, d) = \frac{d}{2d + 2} \left(\frac{1}{n} \right)^{1/d}. \tag{3.1}$$

This error can be achieved by the midpoint algorithm using function values from a grid with $n = m^d$ points.

The error (3.1) implies that if $\varepsilon = 1/((2 + 2/d)m)$ for some integer m then

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d, F_d^{\mathrm{lip}}) = \frac{1}{(1 + 1/d)^d} \left(\frac{1}{2\varepsilon} \right)^d = \frac{1 + o(1)}{e} \left(\frac{1}{2\varepsilon} \right)^d.$$

Hence, $n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d)$ is exponentially large in d, meaning that the integration problem suffers from the curse of dimensionality and is intractable.

3.1.2 Example 2: Integration of Trigonometric Polynomials

Integration of Lipschitz functions is intractable. This may be interpreted as stating that the Lipschitz class F_d^{lip} is just too large. So, we should study the integration problem

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

---

[1]Here and elsewhere in the book we use the standard notation $\mathcal{O}(f(n))$, $\Omega(f(n))$ and $\Theta(f(n))$ for a function $f : \mathbb{N} \to (0, \infty)$. For completeness, we remind the reader that these notations mean that there are three positive numbers n_0, C_1, and C_2 such that

- $g(n) = \mathcal{O}(f(n))$ means that $g(n) \le C_1 f(n)$ for $n \ge n_0$,
- $g(n) = \Omega(f(n))$ means that $g(n) \ge C_2 f(n)$ for $n \ge n_0$, and
- $g(n) = \Theta(f(n))$ means that $C_2 f(n) \le g(n) \le C_1 f(n)$ for $n \ge n_0$.

for a much smaller class F_d than F_d^{lip}. The elements of the class F_d that we shall study are now trigonometric polynomials of degree at most one in each variable. Note that the class F_d is *finite dimensional*, whereas the class F_d^{lip} is infinite dimensional.

For $d = 1$, the space F_1 is linear and generated by three functions, $e_1 = 1$, $e_2(x) = \cos(2\pi x)$, and $e_3(x) = \sin(2\pi x)$ for $x \in [0, 1]$. We define a scalar product $\langle \cdot, \cdot \rangle_{F_1}$ on F_1 by

$$\langle e_i, e_j \rangle_{F_1} = 0 \quad \text{for } i \neq j, \quad \text{and} \quad \langle e_i, e_i \rangle_{F_1} = 1.$$

Hence, the e_i's are orthonormal. Let F_d be the (d-fold) tensor product of F_1 with the tensor (cross-) scalar product

$$\langle f_1 \otimes f_2 \otimes \cdots \otimes f_d, g_1 \otimes g_2 \otimes \cdots \otimes g_d \rangle_{F_d} = \prod_{j=1}^{d} \langle f_j, g_j \rangle_{F_1},$$

for $f_j, g_j \in F_1$. The tensor product of functions is defined by

$$(f_1 \otimes f_2 \otimes \cdots \otimes f_d)(x) = f_1(x_1) f_2(x_2) \cdots f_d(x_d),$$

where $x = [x_1, x_2, \ldots, x_d]$.

For $j = [j_1, j_2, \ldots, j_d]$ with $j_i \in \{1, 2, 3\}$, define $e_{j,d}(x) = \prod_{i=1}^{d} e_{j_i}(x_i)$ for $x \in [0, 1]^d$. Then $\{e_{j,d}\}$ is an orthonormal basis of F_d, and $\dim(F_d) = 3^d$.

Observe that $\|f\|_{F_d} \geq \|f\|_{L_2}$, i.e., the unit ball of F_d is smaller than the unit ball of F_d with respect to the L_2-norm.

It is easy to see that F_d is a Hilbert space with a reproducing kernel K_d. This means that $K_d : [0, 1]^d \times [0, 1]^d \to \mathbb{R}$, $K_d(\cdot, x) \in F_d$ for all $x \in [0, 1]^d$, the $m \times m$ matrix $(K_d(x_i, x_j))_{i,j=1}^{m}$ is symmetric and semi positive-definite for all m and all choices of points $x_j \in [0, 1]^d$, and most importantly that

$$f(x) = \langle f, K_d(\cdot, x) \rangle_{F_d} \quad \text{for all } f \in F_d \text{ and } x \in [0, 1]^d.$$

Indeed, for $f \in F_d$, the Dirac functional $f \mapsto f(x_i)$ can be written in the form

$$f(x_i) = \langle f, \delta_{x_i} \rangle_{F_d}$$

with

$$\delta_{x_i}(x) = \prod_{j=1}^{d} \left[1 + \cos(2\pi(x_{i,j} - x_j)) \right],$$

where $x_{i,j}$ and x_j are the jth components of the vector x_i and x. This proves that the function

$$K_d(x, y) = \langle \delta_x, \delta_y \rangle_{F_d} = \prod_{j=1}^{d} \left[1 + \cos(2\pi(x_j - y_j)) \right] \quad \text{for all } x, y \in [0, 1]^d$$

is the reproducing kernel of F_d. Observe that the kernel K_d is pointwise non-negative, $K_d(x, y) \geq 0$ for all $x, y \in [0, 1]^d$, and $K_d(x, x) = 2^d$.

For multivariate integration, we now have

$$\text{INT}_d f = \langle f, 1 \rangle_{F_d} \quad \text{and} \quad \|1\|_{F_d} = 1 \quad \text{for all } d \in \mathbb{N}.$$

This example has been studied in [167] with a more general norm $\|f\|_{F_{d,\beta}}$ for $\beta > 0$. The case of this section corresponds to $\beta = 1$, whereas the value $\beta = 2$ corresponds to the L_2-norm.

We consider the worst case error on the unit ball of F_d. Again we know from general results that it is enough to consider linear algorithms (quadrature or cubature formulas) of the form

$$Q_n(f) = \sum_{i=1}^{n} c_i f(x_i).$$

For $c_i = n^{-1}$ we obtain quasi-Monte Carlo algorithms, which are widely used for multivariate integration especially for large d, see also Section 3.1.6. Many people prefer to use algorithms with non-negative coefficients $c_i \geq 0$ that integrate the function 1 exactly; this holds iff $\sum_{i=1}^{n} c_i = 1$. The reason is that $c_i \geq 0$ implies numerical stability, and measurement errors for computation of $f(x_i)$ and rounding errors for computation of $Q_n(f)$ are under tight control.

It is easy to check that the worst case error of Q_n is

$$e^{\text{wor}}(Q_n) = \sup_{f \in F_d, \|f\|_{F_d} \leq 1} |\text{INT}_d f - Q_n(f)| = \left\| 1 - \sum_{i=1}^{n} c_i \delta_{x_i} \right\|$$

$$= \left(1 - 2 \sum_{i=1}^{n} c_i + \sum_{i,j=1}^{n} c_i c_j K_d(x_i, x_j) \right)^{1/2}.$$

Let

$$e(n, d) = \inf_{Q_n} e^{\text{wor}}(Q_n)$$

denote the minimal error when we use n function values.

For $n = 1$, we obtain

$$\left[e^{\text{wor}}(Q_1) \right]^2 = 1 - 2c_1 + c_1^2 K_d(x_1, x_1) = 1 - 2c_1 + c_1^2 2^d.$$

Note that the error is now independent of x_1. For a quasi-Monte Carlo algorithm the error is $(2^d - 1)^{1/2}$, which is huge for large d, although we know a priori that the value of an integral lies in $[-1, 1]$ and the error of the zero algorithm, $Q_n = 0$, is just one. In fact, the best choice of c_1 is to minimize the error which yields $c_1 = 2^{-d}$ and

$$e(1, d) = \left(1 - 2^{-d} \right)^{1/2}.$$

For $n \geq 2$ and $c_i \geq 0$, we can find a lower bound on $e^{\mathrm{wor}}(Q_n)$ using the fact that K_d is pointwise non-negative, and we can drop the terms of the double sum for $i \neq j$. Then we have

$$e^{\mathrm{wor}}(Q_n)^2 \geq 1 - 2 \sum_{i=1}^{n} c_i + \sum_{i=1}^{n} c_i^2 2^d = 1 - n 2^{-d} + 2^d \sum_{i=1}^{n} \left(c_i - 2^{-d} \right)^2.$$

For a quasi-Monte Carlo algorithm, we have

$$[e^{\mathrm{wor}}(Q_n)]^2 \geq \frac{1}{n} 2^d - 1,$$

which is exponentially large in d, whereas for all algorithms with $c_i \geq 0$ we have

$$[e^{\mathrm{wor}}(Q_n)]^2 \geq 1 - n 2^{-d}. \tag{3.2}$$

This lower bound cannot be improved since any algorithm with weights $c_i = 2^{-d}$ and points x_i from the set $\left\{0, \frac{1}{2}\right\}^d$ has an error $e(Q_n, F_d)^2 = 1 - n 2^{-d}$. In particular, for $n = 2^d$ the error is 0. Indeed, if $d = 1$ then $Q_2(f) = \frac{1}{2} \left(f(0) + f\left(\frac{1}{2}\right) \right)$ is exact for F_1. The tensor product of Q_2 yields the algorithm Q_{2^d} which is a quasi-Monte Carlo algorithm and is exact for F_d, as claimed.

From (3.2) we conclude that the integration problem for F_d is *intractable* for the class of algorithms with *non-negative coefficients*. Indeed, to achieve an error ε we need $n \geq 2^d (1 - \varepsilon^2)$ function evaluations, which is exponential in d.

It is not clear whether the same is true if we consider arbitrary algorithms with some negative coefficients, see Open Problem 3.

In general, multivariate integration defined over a Hilbert space whose reproducing kernel is pointwise non-negative is much easier to analyze for algorithms with non-negative coefficients than for algorithms with arbitrary coefficients, see [216], [291].

There are examples shown that intractability for the class of algorithms with non-negative coefficients can be broken by using algorithms with negative coefficients, see Section 5 of [177]. We will report these results and examples in Volume II of this book.

3.1.3 Example 3: 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

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

For large α, the class $F_{d,\alpha}$ consists of smooth functions. Indeed, take

$$\beta = [\beta_1, \beta_2, \ldots, \beta_d]$$

with non-negative integers β_j such that $1 + \beta_j < \alpha$. By $|\beta|$ we mean the sum of all β_j, i.e., $|\beta| = \beta_1 + \beta_2 + \cdots + \beta_d$. Then we have

$$
\begin{aligned}
D^\beta f(x) &:= \frac{\partial^{|\beta|}}{\partial^{\beta_1} x_1 \partial^{\beta_2} x_2 \cdots \partial^{\beta_d} x_d} f(x) \\
&= \sum_{h \in \mathbb{Z}^d} \hat{f}(h)(2\pi i)^{|\beta|} h_1^{\beta_1} h_2^{\beta_2} \cdots h_d^{\beta_d} e^{2\pi i h \cdot x}.
\end{aligned}
$$

This derivative is well defined since the last series is absolutely convergent, i.e.,

$$
\begin{aligned}
|D^\beta f(x)| &\leq (2\pi)^{|\beta|} \sum_{h \in \mathbb{Z}^d} |\hat{f}(h)|(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_d)^\alpha \bar{h}_1^{-(\alpha-\beta_1)} \bar{h}_2^{-(\alpha-\beta_2)} \cdots \bar{h}_d^{-(\alpha-\beta_d)} \\
&\leq (2\pi)^{|\beta|} \sum_{h \in \mathbb{Z}^d} \bar{h}_1^{-(\alpha-\beta_1)} \bar{h}_2^{-(\alpha-\beta_2)} \cdots \bar{h}_d^{-(\alpha-\beta_d)} \\
&= (2\pi)^{|\beta|} \prod_{k=1}^{d} \sum_{h \in \mathbb{Z}} \bar{h}^{-(\alpha-\beta_k)} = (2\pi)^{|\beta|} \prod_{k=1}^{d} (1 + 2\zeta(\alpha - \beta_k)) < \infty.
\end{aligned}
$$

Here, ζ denotes the Riemann zeta function, $\zeta(s) = \sum_{j=1}^{\infty} j^{-s}$ for $s > 1$. For $\beta = 0$, we conclude that functions from $F_{d,\alpha}$ have continuous 1-periodic extensions.

As in the previous examples, we consider the integration problem

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

We define several concepts just as before, namely, 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

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

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, see the books of Niederreiter [158] and Sloan and Joe [214], that

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

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 $z \in \{1, 2, \ldots, n-1\}^d$ 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 in [215], where it was proved that

$$e(n, d) = 1 \quad \text{for } n = 0, 1, \ldots, 2^d - 1, \tag{3.3}$$

which implies that

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

That is, even for arbitrarily large α, despite an excellent order of convergence, this integration problem is *intractable*.

It is interesting to compare (3.3) with the statements in Section 3.1.2, in particular with (3.2). Although the latter equation shows that even with one evaluation we obtain algorithms which are slightly better than the zero algorithm $A_0 = 0$, here the situation is even worse:

for *all* $n < 2^d$ we cannot improve the quality of the zero algorithm *at all*.

The proof that $e(n, d) = 1$ is short and instructive, and therefore we repeat it here. First of all, take the most trivial algorithm, $A_n = 0$. Its worst case error is just

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

Hence, $e(n, d) \leq 1$ and we need to prove that as long as $n < 2^d$, all algorithms behave as badly as the most trivial zero algorithm. Take an arbitrary algorithm $A_n(f) = \varphi(f(x_1), f(x_2), \ldots, f(x_n))$ for some (perhaps non-linear) mapping φ and some (perhaps adaptively) chosen points x_j. We will construct a function $f \in F_{d,\alpha}$ for which

$$f(x_j) = 0 \quad \text{for } j = 1, 2, \ldots, n$$

but $|\mathrm{INT}_d f| = 1$.

More precisely, if points x_j are given adaptively, we take the first point x_1 that is independent of functions from $F_{d,\alpha}$, and construct f such that $f(x_1) = 0$. This implies a choice of the second point x_2 for which we construct f for which $f(x_1) = f(x_2) = 0$, and so on. That is, knowing that $f(x_1) = f(x_2) = \cdots = f(x_k) = 0$, the $(k+1)$st point x_{k+1} is chosen and we again construct f for which additionally $f(x_{k+1}) = 0$. This is done for $k = 1, 2, \ldots, n-1$.

We first choose a trigonometric polynomial of the form

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

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

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

Here, we need the assumption that $n < 2^d$. Indeed, we have 2^d unknowns a_h and n homogeneous linear equations, since 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 $\vartheta(x) = e^{-2\pi i h^* \cdot x}$. Our function f is given as

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

where $c = 1$ if $\Re \varphi(0, 0, \ldots, 0) \leq 0$, and $c = -1$ if $\Re \varphi(0, 0, \ldots, 0) \geq 0$.

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 $\overline{h_j - h_j^*} = \max(1, |h_j - h_j^*|) = 1$ and

$$\overline{h_1 - h_1^*} \, \overline{h_2 - h_2^*} \cdots \overline{h_d - h_d^*} = 1 \quad \text{for all } h \in \{0, 1\}^d.$$

We have $|\hat{f}(h - h^*)| = |a_h| \leq 1$ for all $h \in \{0, 1\}^d$, and $\hat{f}(h - h^*) = 0$ for all $h \notin \{0, 1\}^d$. Hence, $|\hat{f}(h)| \leq (\bar{h}_1 \bar{h}_2 \cdots \bar{h}_d)^\alpha$ for all $h \in \mathbb{Z}^d$. This means that $f \in F_{d,\alpha}$, as claimed.

Clearly, $f(x_j) = 0$ for all $j = 1, 2, \ldots, n$ and therefore $A_n(f) = \varphi(0, 0, \ldots, 0)$. Furthermore, $\mathrm{INT}_d f = \hat{f}(0) = c a_{h^*} = c$, and

$$|\mathrm{INT}_d f - A_n(f)| = |c - \varphi(0, 0, \ldots, 0)| \geq |c - \Re \varphi(0, 0, \ldots, 0)| \geq |c| = 1.$$

Hence, the worst case error of A_n is at least 1, which completes the proof.

It was also observed in [215] that the bound on n in the formula $e(n, d) = 1$ is sharp. Namely, for $n = 2^d$, we may use the product rectangle rule

$$A_{2^d}(f) = \frac{1}{2^d} \sum_{k_1=0}^{1} \sum_{k_2=0}^{1} \cdots \sum_{k_d=0}^{1} f\left(\frac{k_1}{2}, \frac{k_2}{2}, \ldots, \frac{k_d}{2}\right),$$

which has the worst case error

$$e^{\mathrm{wor}}(A_{2^d}) = \left(1 + \frac{\zeta(\alpha)}{2^{\alpha-1}}\right)^d - 1.$$

If $\alpha \to \infty$ then

$$e^{\mathrm{wor}}(A_{2^d}) = \frac{d \, \zeta(\alpha)}{2^{\alpha-1}} + \mathcal{O}\left(4^{-\alpha}\right)$$

goes exponentially fast to 0. Hence, for large α we have very peculiar behavior of the nth minimal errors. Nothing happens for $n = 1, 2, \ldots, 2^d - 1$, whereas for $n = 2^d$ the nth minimal error is almost 0. This implies that for every $\varepsilon_0 \in (0, 1)$ and for every integer d_0 there exists a real α such that

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d, F_{d,\alpha}) = 2^d \quad \text{for all } \varepsilon \in (\varepsilon_0, 1) \text{ and for all } d \leq d_0.$$

3.1.4 Example 4: Approximation of C^∞ Functions

We now consider functions from the class $C^\infty([0, 1]^d)$ of infinitely differentiable functions defined on the d-dimensional cube $[0, 1]^d$. Let $f \in C^\infty([0, 1]^d)$. Obviously for any multi-index

$$\beta = [\beta_1, \beta_2, \ldots, \beta_d] \in \mathbb{N}_0^d \quad \text{with } \mathbb{N}_0 = \{0, 1, 2, \ldots\},$$

the function $D^\beta f$ also belongs to $C^\infty([0, 1]^d)$, where D^β is the differentiation operator defined as in the previous example. For any $p \in [1, \infty]$ we also have $\|D^\beta f\|_{L_p} < \infty$, where L_p is the classical space of functions defined on $[0, 1]^d$, i.e., for $p \in [1, \infty)$ we have

$$\|f\|_{L_p} = \left(\int_{[0,1]^d} |f(x)|^p \, dx \right)^{1/p},$$

whereas for $p = \infty$,

$$\|f\|_{L_\infty} = \operatorname*{ess\,sup}_{x \in [0,1]^d} |f(x)|.$$

We restrict the class $C^\infty([0, 1]^d)$ by taking the linear space

$$F = F_{d,p} := \left\{ f \in C^\infty([0, 1]^d) \mid \|f\|_F := \left(\sum_{\beta \in \mathbb{N}_0^d} \frac{1}{\beta!} \|D^\beta f\|_{L_p}^p \right)^{1/p} < \infty \right\},$$

with $\beta! = \prod_{j=1}^d \beta_j!$.

Hence, we deal with infinitely differentiable functions for which the sum of all normalized derivatives is bounded in L_p. This class is nonempty since $f \equiv 1$ belongs to F. Furthermore, all multivariate polynomials belong to F since the series with respect to β for a polynomial consists of only finitely many positive terms. In any case, we hope the reader agrees that F seems to be a "very small" set of functions.

For a given non-negative integer m, we consider the space $G = G_{d,m,p}$ given by

$$G = \left\{ f \in W_p^m([0, 1]^d) \mid \|f\|_G := \left(\sum_{\beta \in \mathbb{N}_0^d : |\beta_j| \le m} \frac{1}{\beta!} \|D^\beta f\|_{L_p}^p \right)^{1/p} < \infty \right\}.$$

Hence, G is the Sobolev space $W_p^m([0, 1]^d)$ of functions whose partial derivatives up to order m in each variable belong to $L_p([0, 1]^d)$. Note that for $m = 0$, the space $G_{d,0,p}$ is just $L_p([0, 1]^d)$. For any m, and for all $f \in F$ we have $\|f\|_G \le \|f\|_F$. Let $P_{d,m}$ denote the linear space of polynomials of d variables which are of degree at most m in each variable. Clearly, $\dim(P_m) = (m + 1)^d$ and

$$\|f\|_F = \|f\|_G \quad \text{for all } f \in P_{d,m}.$$

Hence, the norms in F and G are the same for this $(m + 1)^d$-dimensional subspace. As we shall see this property will be very important for our analysis.

For the classes $F_{d,p}$ and $G_{d,m,p}$, we consider the multivariate approximation problem APP_d with $\mathrm{APP}_d \colon F_{d,p} \to G_{d,m,p}$ given by

$$\mathrm{APP}_d f = f.$$

This is clearly a well-defined problem. Since

$$\|\mathrm{APP}_d\| := \sup_{f \in F_{d,p}, \|f\|_F \leq 1} \|\mathrm{APP}_d\, f\|_{G_{d,m,p}} = 1,$$

it is properly normalized. We approximate $\mathrm{APP}_d\, f$ by algorithms A_n that may now use not only function values but also arbitrary linear functionals, i.e.,

$$A_n(f) = \varphi_n\left(L_1(f), L_2(f), \ldots, L_d(f)\right),$$

where $\varphi_n \colon \mathbb{R}^n \to G_{d,m,p}$ is some linear or non-linear mapping, and L_j is an arbitrary linear functional whose choice may adaptively depend on the already computed values $L_1(f), L_2(f), \ldots, L_{j-1}(f)$. As before, the worst case error of A_n is defined by

$$e^{\mathrm{wor}}(A_n) = \sup_{f \in F_{d,p}, \|f\|_F \leq 1} \|\mathrm{APP}_d\, f - A_n(f)\|_{G_{d,m,p}}.$$

The minimal number of information operations needed to solve the problem to within ε is now given by

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p}) = \min\left\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{wor}}(A_n) \leq \varepsilon\right\}.$$

Observe that we also mention the range space in the list of the arguments of n^{wor}. As we shall see in a moment, tractability will depend on the range space parameter m. In general, we adopt the strategy in our notation of showing all important parameters of the problem and suppressing the parameters that are clear from the context and do not play a major role.

We first discuss the optimal order of convergence. It is easy to see that for any $d \in \mathbb{N}$ and any $r > 0$ we have

$$e(n, d) = \inf_{A_n} e^{\mathrm{wor}}(A_n) = \mathcal{O}\left(n^{-r}\right) \quad \text{as } n \to \infty.$$

To prove this, consider first the spaces $C^s := C^s([0,1]^d)$ of s times continuously differentiable functions with the norm $\|f\|_s = \max_{x \in [0,1]^d} \max_{\beta : |\beta| \leq s} |D^\beta f(x)|$. Now take

$$s_2 = d(r + m) \quad \text{and} \quad s_1 = dm.$$

Note that the norm of the space C^{s_1} is stronger than the norm of $G_{d,m,p}$. That is, $C^{s_1} \subseteq G_{d,m,p}$ and there exists a number C dependent on d, m and p such that $\|f\|_{G_{d,m,p}} \leq C \|f\|_{s_1}$ for all $f \in C^{s_1}$.

Note that for any positive k, the class $F_{d,p}$ is a subset of the Sobolev space $W_p^k([0,1]^d)$. If the embedding condition $k - s_2 > d/p$ holds then $W_p^k([0,1]^d)$ and $F_{d,p}$ can both be regarded as subsets of C^{s_2}.

It is well-known that we can approximate functions from C^{s_2} in the norm of C^{s_1}, and then in the norm of $G_{d,m,p}$, by algorithms using n function values with worst case error of order $n^{-(s_2-s_1)/d}$. Moreover, $r = (s_2 - s_1)/d$ is the order of convergence; this result was probably first observed by Bakhvalov [7] for $m = 0$, which gives $s_1 = 0$.

For general s_1, which is needed for $m \geq 1$, this result can be found, for instance, in the book of Triebel [246, p. 348].

Take $k = s_2 + 1 + d/p$. Then we conclude that functions from $F_{d,p}$ can be approximated in the norm of $G_{d,m,p}$ with worst case error of order n^{-r}, as claimed.

Since r can be arbitrarily large, the optimal order of convergence of the multivariate approximation problem for the class $F_{d,p}$ is formally infinite. This implies that for a fixed d, the minimal number of information operations goes to infinity slower then any power of ε^{-1}. That is, for any fixed d and any positive η we have

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p}) = o(\varepsilon^{-\eta}) \quad \text{as } n \to \infty.$$

Again, this is very encouraging but one could say that this is possible since the class $F_{d,p}$ is so small.

But how about tractability? How long do we have to wait to see this nice convergence of $n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p})$ to 0? We claim that

$$e(n, d) = 1 \quad \text{for all } n = 0, 1, \ldots, (m+1)^d - 1,$$

which implies that

$$n^{\mathrm{wor}}(\varepsilon, \mathrm{APP}_d, F_{d,p}, G_{d,m,p}) \geq (m+1)^d \quad \text{for all } \varepsilon \in [0, 1).$$

Hence if $m \geq 1$ then we have the curse of dimensionality and the multivariate approximation problem for the classes $F_{d,p}$ and $G_{d,m,p}$ is *intractable*. This means that the set $F_{d,m}$ is *not* so small after all.

The proof that $e(n, d) = 1$ is essentially the same as the proof in the previous subsection. First of all, observe that the zero algorithm $A_n(f) = 0$ has worst case error at most 1 since APP_d is properly normalized. Hence, $e(n, d) \leq 1$. To prove the reverse inequality, take an arbitrary algorithm $A_n(f) = \varphi_n(L_1(f), \ldots, L_n(f))$ that uses adaptive linear functionals L_j. We now show that $e^{\mathrm{wor}}(A_n) \geq 1$.

For $b = [b_1, b_2, \ldots, b_d] \in \{0, 1, \ldots, m\}^d$, define the functions

$$f_b(x) = \prod_{j=1}^{d} \left(x_j - \frac{1}{2} \right)^{b_j}.$$

The functions f_b are polynomials of at most degree m in each variable. Each b yields a new polynomial f_b and the set $\{f_b\}$ consists of $(m+1)^d$ linearly independent polynomials. Note also that $\|f_b\|_F = \|f_b\|_G$ since all terms $D^\beta f_b$ are 0 if there is an index $\beta_j > m$, and hence the summation for the F-norm is the same as the summation for the G-norm. Let

$$g(x) = \sum_{b \in \{0, 1, \ldots, m\}^d} a_b f_b(x)$$

for some real numbers a_b. Again for any choice of a_b we have $\|g\|_F = \|g\|_G$.

We choose a_b's such that $L_1(g) = 0$. Based on this zero value, the second linear functional L_2 is chosen, and we add the second equation for a_b's by requiring that

$L_2(g) = 0$. We do the same for all chosen linear functionals L_j based on the zero information, and we have n homogeneous linear equations for $\{a_b\}$,

$$\sum_{b \in \{0,1,\dots,m\}^d} a_b L_j(f_b) = 0 \quad \text{for } j = 1, 2, \dots, n.$$

Since we have $(m + 1)^d > n$ unknowns, we can choose a non-zero vector $a_b = a_b^*$ satisfying these n equations. The function g with a_b^* is non-zero since the f_b's are linearly independent. Then $\|g\|_{F_{d,p}}$ is well defined and positive. We finally define two functions

$$f_k = (-1)^k \frac{g}{\|g\|_{F_{d,p}}} \quad \text{for } k \in \{0, 1\}.$$

Note that

$$f_k \in F_{d,p} \quad \text{and} \quad \|f_k\|_{F_{d,p}} = \|f\|_{G_{d,m,p}} = 1.$$

Furthermore, $L_j(f_k) = 0$ for all $j = 1, 2, \dots, n$ and therefore $A_n(f_k) = \varphi(0, \dots, 0)$ does not depend on k. Hence,

$$\begin{aligned}
e^{\mathrm{wor}}(A_n) &\geq \max_{f_0, f_1} \left(\|f_0 - \varphi(0, 0, \dots, 0)\|_G, \|f_1 - \varphi(0, 0, \dots, 0)\|_G \right) \\
&\geq \tfrac{1}{2} \left(\|f_0 - \varphi(0, 0, \dots, 0)\|_G + \|f_1 - \varphi(0, 0, \dots, 0)\|_G \right) \\
&\geq \tfrac{1}{2} \|f_0 - f_1\|_G = 1.
\end{aligned}$$

This completes the proof.

Hence, for $m \geq 1$ we have the curse of dimensionality and intractability. How about $m = 0$? In this case we restrict ourselves to $p = 2$ and analyze the problem in detail.

We will need a couple of known general results that will be presented in Chapter 4. The reader may also consult, for instance, the books [242], [244] where these results can be also found. For $m = 0$ and $p = 2$, the space $G_{d,0,2}$ is just the Hilbert space $L_2 = L_2([0, 1]^d)$ with the inner product

$$\langle f, g \rangle_{L_2} = \int_{[0,1]^d} f(x) g(x) \, \mathrm{d}x,$$

whereas $F = F_{d,2}$ is the unit ball of the Hilbert space with the inner product

$$\langle f, g \rangle_F = \sum_{\beta \in \mathbb{N}_0^d} \frac{1}{\beta!} \langle D^\beta f, D^\beta g \rangle_{L_2}.$$

Let $W_d = \mathrm{APP}_d^* \mathrm{APP}_d \colon F_{d,2} \to F_{d,2}$, where $\mathrm{APP}_d^* \colon L_2 \to F_{d,2}$ is the adjoint operator of APP_d. Obviously W_d is a self-adjoint positive semi-definite operator. It is well-known that $\lim_{n \to \infty} e(n, d) = 0$ iff W_d is compact. Since we already know that the limit of $e(n, d)$ is 0, we conclude that W_d is compact. Hence, $F_{d,2}$ has an orthonormal basis of the eigenfunctions $\eta_{d,j}$ of W_d, i.e.,

$$W_d \eta_{d,j} = \lambda_{d,j} \eta_{d,j}$$

with $\langle \eta_{d,j}, \eta_{d,k} \rangle_F = \delta_{j,k}$. We may assume that the non-negative eigenvalues $\lambda_{d,j}$ are ordered, i.e.,

$$\lambda_{d,1} \geq \lambda_{d,2} \geq \cdots \geq \lambda_{d,n} \geq \cdots \geq 0.$$

Obviously $\lim_{n \to \infty} \lambda_{d,n} = 0$. Since we now allow algorithms using arbitrary linear functionals, it is well-known that

$$e(n, d) = \sqrt{\lambda_{d,n+1}} \quad \text{for all } n = 0, 1, \ldots,$$

and that the algorithm

$$A_n(f) = \sum_{j=1}^{n} \langle f, \eta_{d,j} \rangle_F \, \eta_{d,j}$$

has worst case error equal to $e(n, d)$. We stress that although the algorithm A_n is linear and uses non-adaptive information, it minimizes the worst case error in the class of all non-linear algorithms using n arbitrary adaptive linear functionals. This is a typical result that holds for much more general problems, as we shall see later in Chapter 4.

Periodic Case for $m = 0$ and $p = 2$

We now restrict our attention to functions from the space $F = F_{d,2}$ that are periodic. By a periodic function $f \in F_{d,2}$ we mean that for $d = 1$ we have $f^{(\beta)}(1) = f^{(\beta)}(0)$ for all $\beta \in \mathbb{N}_0$, whereas for $d \geq 1$, we have $(D^\beta f)(x) = (D^\beta f)(y)$ if $|x_i - y_i| \in \{0, 1\}$ for all i. That is, the values of all derivatives are the same if a component $x_i = 0$ of x is changed into $x_i = 1$. Hence, let

$$F_{d,2}^{\text{per}} = \{ f \in F_{d,2} \mid f \text{ is periodic} \}.$$

The space $F_{d,2}^{\text{per}}$ is equipped with the same norm as $F_{d,2}$. For example, the functions $\prod_{k=1}^{d} \eta_{j_k}(x_j)$, with $j \in \mathbb{N}_0^d$ and $\eta_{j_k}(x) = \sin(2\pi j_k x)$ or $\eta_{j_k}(x) = \cos(2\pi j_k x)$ belong to $F_{d,2}^{\text{per}}$. Note that the approximation problem is still properly normalized for the subspace $F_{d,2}^{\text{per}}$ since $\|\text{APP}_d\|_{F_{d,2}^{\text{per}} \to L_2} = 1$.

The subspace $F_{d,2}^{\text{per}}$ is much smaller than $F_{d,2}$. So if we establish a negative result for $F_{d,2}^{\text{per}}$, then the same result will be also true for the larger class $F_{d,2}$. Obviously, positive results for $F_{d,2}^{\text{per}}$ do not have to be true for $F_{d,2}$.

To verify tractability of the approximation problem defined over the subspace $F_{d,2}^{\text{per}}$, we need to find the eigenpairs of W_d. It will be instructive to consider first the univariate case $d = 1$. Define $\eta_1(x) = 1$, and for $k = 1, 2, \ldots$, define

$$\eta_{2k}(x) = \sqrt{2} e^{-2(\pi k)^2} \sin(2\pi k x), \quad \eta_{2k+1}(x) = \sqrt{2} e^{-2(\pi k)^2} \cos(2\pi k x).$$

It is easy to check that the sequence $\{\eta_k\}$ is orthonormal in the subspace $F_{1,2}^{\text{per}}$, i.e., $\langle \eta_k, \eta_s \rangle_{F_{1,2}} = \delta_{k,s}$. Define

$$K_1(x, y) = \sum_{j=1}^{\infty} \eta_j(x) \eta_j(y) \quad \text{for } x, y \in [0, 1].$$

We claim that K_1 is the *reproducing kernel* of $F_{d,2}^{\text{per}}$. That is, in particular, $K_1(\cdot, y) \in F_{d,2}^{\text{per}}$ for all $y \in [0, 1]$, and $f(y) = \langle f, K_1(\cdot, y) \rangle_{F_{1,2}}$ for all $f \in F_{d,2}^{\text{per}}$ and all $y \in [0, 1]$. Indeed, it is enough to check the last property. Observe that for arbitrary $\beta \in \mathbb{N}$ and $k \geq 1$, we have

$$\langle f^{(\beta)}, \eta_{2k}^{(\beta)} \rangle_{L_2} \eta_{2k}(y) + \langle f^{(\beta)}, \eta_{2k+1}^{(\beta)} \rangle_{L_2} \eta_{2k+1}(y)$$
$$= (2\pi k)^{2\beta} \left(\langle f, \eta_{2k} \rangle_{L_2} \eta_{2k}(y) + \langle f, \eta_{2k+1} \rangle_{L_2} \eta_{2k+1}(y) \right).$$

Therefore,

$$\langle f, K_1(\cdot, y) \rangle_{F_{1,2}} = \sum_{j=1}^{\infty} \langle f, \eta_j \rangle_{F_{1,2}} \eta_j(y) = \sum_{j=1}^{\infty} \sum_{\beta=0}^{\infty} \frac{1}{\beta!} \langle f^{(\beta)}, \eta_j^{(\beta)} \rangle_{L_2} \eta_j(y)$$

$$= \sum_{j=1}^{\infty} \langle f, \eta_j \rangle_{L_2} \eta_j(y) + \sum_{\beta,k=1}^{\infty} \frac{\langle f^{(\beta)}, \eta_{2k}^{(\beta)} \rangle_{L_2} \eta_{2k}(y) + \langle f^{(\beta)}, \eta_{2k+1}^{(\beta)} \rangle_{L_2} \eta_{2k+1}(y)}{\beta!}$$

$$= \langle f, \eta_1 \rangle_{L_2} + \sum_{k=1}^{\infty} e^{(2\pi k)^2} \left(\langle f, \eta_{2k} \rangle_{L_2} \eta_{2k}(y) + \langle f, \eta_{2k+1} \rangle_{L_2} \eta_{2k+1}(y) \right)$$

$$= \langle f, 1 \rangle_{L_2} + 2 \sum_{k=1}^{\infty} \langle f, \sin 2\pi k \cdot \rangle_{L_2} \sin(2\pi k y) + \langle f, \cos 2\pi k \cdot \rangle_{L_2} \cos(2\pi k y).$$

The last series is the Fourier series for f evaluated at y. Since f is periodic and differentiable, this is equal to $f(y)$.

This also proves that the sequence $\{\eta_k\}$ is an orthonormal basis of the subspace $F_{1,2}^{\text{per}}$. Indeed, it is enough to show that if $f \in F_{1,2}^{\text{per}}$ and $\langle f, \eta_j \rangle_{F_{1,2}} = 0$ for all j then $f = 0$. Orthogonality of f to all η_j implies that $\langle f, K_1(\cdot, y) \rangle_{F_{1,2}} = 0$, and therefore $f(y) = 0$. Since this holds for all $y \in [0, 1]$, we have $f = 0$, as claimed.

Note that for $k \neq s$, we have

$$0 = \langle \eta_k, \eta_s \rangle_{L_2} = \langle \text{APP}_1 \eta_k, \text{APP}_1 \eta_s \rangle_{L_2} = \langle \eta_k, \text{APP}_1^* \text{APP}_1 \eta_s \rangle_F = \langle \eta_k, W_1 \eta_s \rangle_F.$$

This means that $W_1 \eta_s$ is orthogonal to all η_k except $k = s$. Hence,

$$W_1 \eta_s = \lambda_s \eta_s,$$

and $\lambda_s = \langle \eta_s, \eta_s \rangle_{L_2}$. This yields

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_{2k} = \lambda_{2k+1} = e^{-(2\pi k)^2} \quad \text{for } k = 1, 2, \dots.$$

For $d \geq 2$, it is easy to see that $F_{d,2}^{\text{per}}$ is the tensor product of d copies of $F_{1,2}^{\text{per}}$ and W_d is the d-fold tensor product of W_1. This implies that the eigenpairs of W_d are

$$W_d \eta_{d,j} = \lambda_{d,j} \eta_{d,j},$$

where $j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d$ and

$$\eta_{d,j}(x) = \prod_{k=1}^{d} \eta_{j_k}(x_k) \quad \text{and} \quad \lambda_{d,j} = \prod_{k=1}^{d} \lambda_{j_k}.$$

Hence, the eigenvalues for the d-dimensional case are given as the products of the eigenvalues for the univariate case. To find out the nth optimal error $e(n, d)$, we must order the sequence $\{\lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d}\}_{j \in \mathbb{N}^d}$. Then the square root of the $(n + 1)$st largest eigenvalue is $e(n, d)$. Thus, $e(n, d) \leq \varepsilon$ iff n is at least the cardinality of the set of all eigenvalues $\lambda_{d,j} > \varepsilon^2$. If we denote $n(\varepsilon, d) := n^{\text{wor}}(\varepsilon, \text{APP}_d, F_{d,2}^{\text{per}}, L_2)$ as the minimal number of linear functionals needed to solve the problem to within ε, then

$$n(\varepsilon, d) = \left| \left\{ j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d \mid \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2 \right\} \right|.$$

Clearly, $n(\varepsilon, d) = 0$ for all $\varepsilon \geq 1$ since the largest eigenvalue is 1. It is also easy to see that $n(\varepsilon, d) = 1$ for all $\varepsilon \in (e^{-2\pi^2}, 1)$ since the second largest eigenvalue is $\lambda_2 = e^{-4\pi^2}$. For $d = 1$, note that $e^{-(2\pi k)^2} > \varepsilon^2$ iff $k \leq \lceil \sqrt{2 \ln \varepsilon^{-1}}/(2\pi) \rceil - 1$. This yields that

$$n(\varepsilon, 1) = 2 \left\lceil \frac{1}{2\pi} \sqrt{2 \ln \varepsilon^{-1}} \right\rceil - 1 = \frac{\sqrt{2}}{\pi} \sqrt{\ln \frac{1}{\varepsilon}} + \mathcal{O}(1) \quad \text{as } \varepsilon \to 0.$$

For $d \geq 1$, we have the formula

$$n(\varepsilon, d + 1) = \sum_{j=1}^{\infty} n\left(\varepsilon / \sqrt{\lambda_j}, d\right) = n(\varepsilon, d) + 2 \sum_{k=1}^{\infty} n\left(\varepsilon e^{2(\pi k)^2}, d\right),$$

which relates the cases for $d + 1$ and d. The last two series are only formally infinite, since for large j and k the corresponding terms are 0. More precisely, to obtain a positive $n(\varepsilon e^{2(\pi k)^2}, d)$ we need to assume that $\varepsilon e^{2(\pi k)^2} < 1$. Let

$$k_\varepsilon = \left\lceil \frac{\sqrt{2}}{2\pi} \sqrt{\ln \frac{1}{\varepsilon}} \right\rceil - 1.$$

Then

$$n(\varepsilon, d + 1) = n(\varepsilon, d) + 2 \sum_{k=1}^{k_\varepsilon} n\left(\varepsilon e^{2(\pi k)^2}, d\right).$$

We now show by induction on d that

$$n(\varepsilon, d) = \Theta\left(\left(\ln \frac{1}{\varepsilon} \right)^{d/2} \right) \quad \text{as } \varepsilon \to 0. \tag{3.4}$$

This is clearly true for $d = 1$. If it is true for d, then using the formula for $n(\varepsilon, d + 1)$ we easily see that we can bound $n(\varepsilon, d + 1)$ from above by $\mathcal{O}((\ln 1/\varepsilon)^{(d+1)/2})$ since k_ε is of order $(\ln 1/\varepsilon)^{1/2}$. We can estimate $n(\varepsilon, d + 1)$ from below by taking $k_\varepsilon/2$ terms and using the lower bound on $n(\varepsilon e^{2(\pi k)^2}, d)$, which again yields an estimate of order $(\ln 1/\varepsilon)^{(d+1)/2}$.

Let us pause and ask what (3.4) means. From one point of view, this estimate of $n(\varepsilon, d)$ is quite positive since we have weak dependence on ε only through $\ln 1/\varepsilon$. Hence, if d is relatively small, then the multivariate approximation problem with $m = 0$ can be easily solved. But if d is large, (3.4) may suggest that we have an exponential dependence on d, and the problem may be intractable. As we already know the factors in the big theta notation are very important for large d and so we can claim nothing based solely on (3.4). We need more information about how $n(\varepsilon, d)$ behaves. We now prove that

$$C_d := \lim_{\varepsilon \to 0} \frac{n(\varepsilon, d)}{(\ln 1/\varepsilon)^{d/2}} = \frac{1}{(2\pi)^{d/2} \Gamma(1 + d/2)}, \qquad (3.5)$$

establishing the asymptotic behavior of $n(\varepsilon, d)$ as ε tends to 0.

For $d = 1$, we have already shown the formula $C_1 = \sqrt{2}/\pi$. Assume that C_d is the asymptotic constant for d, and consider the case $d + 1$. For every positive δ there exists $\varepsilon_d = \varepsilon_{d,\delta} \in (0, 1)$ such that for all $\varepsilon \in (0, \varepsilon_d]$, we have

$$n(\varepsilon, d) = C_d(1 + g(\varepsilon)) \left(\ln \frac{1}{\varepsilon} \right)^{d/2} \quad \text{with } |g(\varepsilon)| \leq \delta \text{ and } \lim_{\varepsilon \to 0} g(\varepsilon) = 0.$$

Define

$$k_\varepsilon^* = \left\lceil \frac{\sqrt{2}}{2\pi} \sqrt{\ln \frac{\varepsilon_d}{\varepsilon}} \right\rceil - 1.$$

Note that $k_\varepsilon - k_\varepsilon^* = \mathcal{O}(1)$ as $\varepsilon \to 0$. We have

$$n(\varepsilon, d + 1) = C_d(1 + g(\varepsilon)) \left(\ln \frac{1}{\varepsilon} \right)^{d/2}$$

$$+ 2C_d \sum_{k=1}^{k_\varepsilon^*} \left(1 + g\left(\varepsilon e^{2(\pi k)^2} \right) \right) \left(\ln \frac{1}{\varepsilon} - 2(\pi k)^2 \right)^{d/2}$$

$$+ 2 \sum_{k=k_\varepsilon^*+1}^{k_\varepsilon} n(\varepsilon e^{2(\pi k)^2}, d).$$

Note that for $k \in [k_\varepsilon^* + 1, k_\varepsilon]$ we have $n(\varepsilon e^{2(\pi k)^2}, d) \leq n(\varepsilon_d, d)$, and therefore

$$\sum_{k=k_\varepsilon^*+1}^{k_\varepsilon} n(\varepsilon e^{2(\pi k)^2}, d) \leq (k_\varepsilon - k_\varepsilon^*) n(\varepsilon_d, d) = \mathcal{O}\left((\ln \varepsilon^{-1})^{d/2} \right).$$

Now consider the terms for which $k \in [1, k_\varepsilon^*]$. Then $\varepsilon e^{2(\pi k)^2} \leq \varepsilon_d$. For ε tending to 0, we have

$$\sum_{k=1}^{k_\varepsilon^*} \left(1 + g\left(\varepsilon e^{2(\pi k)^2}\right)\right) \left(\ln \frac{1}{\varepsilon} - 2(\pi k)^2\right)^{d/2}$$

$$= (1 + o(1)) \int_1^{k_\varepsilon^*} \left(\ln \frac{1}{\varepsilon} - 2(\pi x)^2\right)^{d/2} dx$$

$$= \frac{1 + o(1)}{\sqrt{2\pi}} \left(\ln \frac{1}{\varepsilon}\right)^{(d+1)/2} \int_0^1 (1 - x^2)^{d/2} dx$$

$$= \frac{1 + o(1)}{\sqrt{2\pi}} \left(\ln \frac{1}{\varepsilon}\right)^{(d+1)/2} \tfrac{1}{2} B\left(\tfrac{1}{2}, 1 + d/2\right),$$

where $B(x, y)$ is the beta function and is related to the Gamma function by $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$. This proves that

$$n(\varepsilon, d + 1) = C_{d+1}(1 + o(1))(\ln 1/\varepsilon)^{(d+1)/2}$$

as ε goes to 0, with

$$C_{d+1} = \frac{B\left(\tfrac{1}{2}, 1 + d/2\right) C_d}{\sqrt{2\pi}} = \frac{\Gamma(1/2)\Gamma(1 + d/2)}{\sqrt{2\pi}\Gamma(1 + (d + 1)/2)} C_d.$$

Solving this recurrence, we obtain

$$C_{d+1} = \frac{\Gamma(1/2)^d \Gamma(3/2)}{(\sqrt{2\pi})^d \Gamma(1 + (d + 1)/2)} C_1 = \frac{1}{(2\pi)^{(d+1)/2}\Gamma(1 + (d + 1)/2)}$$

which agrees with the asymptotic formula (3.5).

We stress that the asymptotic constant C_d in (3.5) is super exponentially small in d due to the presence of $\Gamma(1 + d/2)$ in the denominator. This property raises our hopes that we can beat the apparent exponential dependence on d. Indeed, assume for a moment that the limit in (3.5) is uniform in d. That is, suppose that there exists a positive ε_0 such that for all $\varepsilon \in (0, \varepsilon_0)$ and all d, we have

$$n(\varepsilon, d) \leq 2C_d \left(\ln \frac{1}{\varepsilon}\right)^{d/2} = \frac{2(\ln \varepsilon^{-1})^{d/2}}{(2\pi)^{d/2}\Gamma(1 + d/2)}.$$

It can be easily checked that $x^{d/2}/\Gamma(1 + d/2) \leq \exp(x)$ for all $x \geq 1$. Therefore

$$n(\varepsilon, d) \leq \frac{2}{(2\pi)^{d/2}} \frac{1}{\varepsilon}.$$

Hence, we have *strong polynomial tractability* if (3.5) holds uniformly in d.

We now return to the proof of (3.5) with the new task of checking whether ε_d can be uniformly bounded from below. Unfortunately, this is *not* true. It is enough to take

$\varepsilon^2 \in [\lambda_3, \lambda_2)$ to realize that we can take $d - 1$ indices $j_i = 1$ and the remaining index $j_i = 2$ to obtain the eigenvalue $\lambda_{d,j} = \lambda_2 > \varepsilon^2$. Hence $n(\varepsilon, d) \geq d$ which contradicts strong polynomial tractability.

In fact, even polynomial tractability does not hold. This follows from the general observation that as long as the largest eigenvalue is 1, and the second largest eigenvalue λ_2 for $d = 1$ is positive then there is *no* polynomial tractability. Indeed, for an arbitrary integer k and arbitrary $d > k$, consider the eigenvalues $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}$ with $d - k$ indices j_i equal to 1 and k indices j_i equal to 2. Then we have at least $\binom{d}{k} = \Theta(d^k)$ eigenvalues equal to λ_2^k. It is enough to take now, say, $\varepsilon^2 = \lambda_2^k/2$ to realize that $n\left(\sqrt{\lambda_2^k/2}, d\right)$ is at least of order d^k. Since k can be arbitrary, this contradicts not only strong polynomial tractability but also polynomial tractability.

Well, we are back to square one. Despite the exponentially small asymptotic constants, we have *polynomial intractability* of the multivariate problem for $m = 0$. Hence, the only remaining hope for a positive result is weak tractability. Here we will finally report good news.

As in [288], let $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d} > \varepsilon^2$ and let k be the number of indices $j_i \geq 2$. Then[2] $(d - k)_+$ indices are equal to 1. Note that $\lambda_2^k > \varepsilon^2$ implies that

$$k \leq a(\varepsilon) := \left\lceil \frac{2 \ln \varepsilon^{-1}}{\ln \lambda_2^{-1}} \right\rceil - 1.$$

So we have at least $(d - a(\varepsilon))_+$ indices equal to 1. Observe also that $j_i \leq n(\varepsilon, 1)$. Thus

$$\binom{d}{(d - a(\varepsilon))_+} \leq n(\varepsilon, d) \leq \binom{d}{(d - a(\varepsilon))_+} n(\varepsilon, 1)^{a(\varepsilon)}.$$

For a fixed ε and for d tending to infinity, we have

$$n(\varepsilon, d) = \Theta\left(d^{\lceil 2 \ln \varepsilon^{-1}/\ln \lambda_2^{-1} \rceil - 1}\right)$$

with the factors in the big theta notation depending now on ε^{-1}.

For arbitrary d and $\varepsilon \in (0, 1)$ we conclude that

$$n(\varepsilon, d) \leq \frac{(d + a(\varepsilon))^{a(\varepsilon)}}{a(\varepsilon)!} \left(2 \left\lceil \frac{1}{2\pi} \sqrt{2 \ln \frac{1}{\varepsilon}} \right\rceil - 1\right)^{a(\varepsilon)}.$$

This implies that

$$\lim_{\varepsilon^{-1} + d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0$$

which means that weak tractability indeed holds.

Hence, we have mixed news for the periodic case of the approximation problem. We have polynomial intractability, which obviously implies polynomial intractability for the original non-periodic case. But we have weak tractability for the periodic case, and it is not yet clear whether this good property extends to the non-periodic case.

---

[2] Here and elsewhere in the book we use the standard notation $x_+ = \max(x, 0)$.

Weak tractability for $m = 0$ and $p = 2$

We now show that weak tractability holds not only for the original non-periodic case but it also holds for a much larger space of less smooth functions. Namely, define

$$F_{d,2}^1 = G_{d,1,2} = W_2^1([0,1]^d)$$

as the Sobolev space of functions whose partial derivatives up to order one belong to $L_2 = L_2([0,1]^d)$. The norm in $F_{d,2}^1$ is defined as in $G_{d,1,2}$. Clearly

$$F_{d,2} \subseteq F_{d,2}^1 \quad \text{and} \quad \|f\|_{F_{d,2}^1} \leq \|f\|_{F_{d,2}} \quad \text{for all } f \in F_{d,2}.$$

Again, consider first the case $d = 1$, and the subspace $\widetilde{F}_{1,2}^1$ of periodic functions from $F_{d,2}^1$. Now periodicity means that $f(1) = f(0)$. Proceeding as before, it is easy to check that the functions $\eta_1 = 1$, and

$$\eta_{2k}(x) = \frac{\sqrt{2}}{\sqrt{1 + (2\pi k)^2}} \sin(2\pi k x), \quad \eta_{2k+1}(x) = \frac{\sqrt{2}}{\sqrt{1 + (2\pi k)^2}} \cos(2\pi k x)$$

are orthonormal in $\widetilde{F}_{1,2}^1$, and the function

$$K_1(x, y) = \sum_{j=1}^{\infty} \eta_j(x)\eta_j(y)$$

is the reproducing kernel of $\widetilde{F}_{1,2}$. Therefore the sequence $\{\eta_j\}$ forms a basis of $\widetilde{F}_{1,2}$. The eigenvalues λ_j^{per} of $W_1 = \text{APP}_1^* \text{APP}_1 \colon \widetilde{F}_{1,2}^1 \to \widetilde{F}_{1,2}^1$ are

$$\lambda_1^{\text{per}} = 1 \quad \text{and} \quad \lambda_{2k}^{\text{per}} = \lambda_{2k+1}^{\text{per}} = \frac{1}{1 + (2\pi k)^2} \quad \text{for } k = 1, 2, \ldots.$$

We now turn to the space $F_{1,2}^1$ of non-periodic functions. Define

$$g(x) = x - \tfrac{1}{2} + \sum_{k=1}^{\infty} \frac{1}{\pi k (1 + (2\pi k)^2)} \sin(2\pi k x) \quad \text{for } x \in [0, 1].$$

It is easy to check that g belongs to $F_{1,2}^1$ and is orthogonal to all η_j. Note that $g(1) = -g(0) = \tfrac{1}{2}$, hence $g \notin \widetilde{F}_{1,2}^1$. For $f \in F_{1,2}^1$, let

$$h_f(x) = f(x) - [f(1) - f(0)] g(x).$$

Then $h_f \in \widetilde{F}_{1,2}^1$. Hence,

$$f = [f(1) - f(0)] g + h_f \quad \text{for all } f \in F_{1,2}^1.$$

This decomposition suggests that we first compute $L_1(f) = f(1) - f(0)$ and then approximate the function $h_f = f - L_1(f)g$. Note that

$$\langle f, \eta_j \rangle_{F_{1,2}^1} = L_1(f) \langle g, \eta_j \rangle_{F_{1,2}^1} + \langle h_f, \eta_j \rangle_{F_{1,2}} = \langle h_f, \eta_j \rangle_{F_{1,2}}$$

and

$$\|f\|_{F_{1,2}^1}^2 = L_1(f)^2 \|g\|_{F_{1,2}^1}^2 + \|h_f\|_{F_{1,2}^1}^2.$$

Hence, approximation of functions from the unit ball of $F_{1,2}^1$ with n information evaluations is not harder than approximation of periodic functions from the unit ball of $\tilde{F}_{1,2}^1$ with $n - 1$ information evaluations, and not easier than the periodic case with n evaluations. Let $\lambda_j^{\text{non-per}}$ denote the ordered sequence of eigenvalues of $W_1 = \text{APP}_1^* \text{APP}_1 \colon F_{1,2}^1 \to F_{1,2}^1$ for the non-periodic case. It is easy to check that $\lambda_1^{\text{non-per}} = 1$, and $\lambda_2^{\text{non-per}} < \lambda_1^{\text{non-per}}$, as well as

$$\lambda_j^{\text{per}} \leq \lambda_j^{\text{non-per}} \leq \lambda_{j-1}^{\text{per}} \quad \text{for all } j \geq 2.$$

Hence, $\lambda_j^{\text{non-per}} = \Theta(j^{-1})$.

We turn to the case $d \geq 2$. Since $F_{d,2}^1$ is the d-fold tensor product of $F_{1,2}^1$, the eigenvalues of $W_d = \text{APP}_d^* \text{APP}_d \colon F_{d,2}^1 \to F_{d,2}^1$ are products of $\lambda_{j_1}^{\text{non-per}} \cdots \lambda_{j_d}^{\text{non-per}}$ for $j_i \in \mathbb{N}$. In Theorem 5.5 of Chapter 5 we prove that linear tensor product problems are weakly tractable as long as the eigenvalues for $d = 1$ satisfy the following two conditions:

- the second largest eigenvalue is smaller than the largest eigenvalue,
- the nth largest eigenvalue goes to 0 faster than $(\ln n)^{-2} (\ln \ln n)^{-2}$.

These two assumptions hold in our case, and therefore the approximation problems for the space $F_{d,2}^1$ as well as for the smaller space $F_{d,2}$ are weakly tractable.

We finish this section with a short summary. The multivariate approximation problem studied in this section is defined on infinitely differentiable functions, and we allow arbitrary linear functionals as information operations. The optimal rate of convergence of this problem is infinite. Despite this excellent asymptotic speed of convergence, the problem is *intractable* if the target space is equipped with the L_p-norm involving partial derivatives, i.e., when $m \geq 1$. For $m = 0$ and $p = 2$, the target space is simply L_2, and then the problem remains *polynomially intractable*, but is *weakly tractable*.

3.1.5 Example 5: Discrepancy

We now discuss the notion of *discrepancy*, which is related to multivariate integration for some classes of functions. 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. One of the chapters of Volume II is devoted to discrepancy and its role

for tractability of multivariate integration. Here, we only wish to introduce this subject and illustrate it by a few surprising facts and results.

We begin with the definition of the L_p-*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 $x_1 x_2 \cdots x_d$. 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

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

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 L_p-star discrepancy of the points $t_1, \ldots, t_n \in [0, 1]^d$ is defined by the L_p-norm of the discrepancy function disc, i.e., for $p \in [1, \infty)$,

$$\mathrm{disc}_p^*(t_1, t_2, \ldots, t_n) = \left(\int_{[0,1]^d} \left| x_1 x_2 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|^p \mathrm{d}x \right)^{1/p}, \quad (3.6)$$

and for $p = \infty$,

$$\mathrm{disc}_\infty^*(t_1, t_2, \ldots, t_n) = \sup_{x \in [0,1]^d} \left| x_1 x_2 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|. \quad (3.7)$$

The main problem associated with L_p-star discrepancy is that of finding points $t_1, t_2, \ldots, t_n$ that minimize disc_p^*, and to study how this minimum depends on d and n. There are many deep results for this problem and we will report some of them in Volume II.

We now show that the L_p-star discrepancy is intimately related to multivariate integration. Let $W_q^1 := W_q^{(1,1,\ldots,1)}([0, 1]^d)$ be the Sobolev space of functions defined on $[0, 1]^d$ that are once differentiable in each variable and whose derivatives have finite L_q-norm, where $1/p + 1/q = 1$, see the books of Drmota and Tichy [50] and Niederreiter [158]. 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

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

for $q \in [1, \infty)$ and

$$\|f\|_{d,\infty}^* = \sup_{x \in [0,1]^d} \left| \frac{\partial^d}{\partial x} f(x) \right|$$

for $q = \infty$. Here, $\partial x = \partial x_1 \partial x_2 \ldots \partial x_d$.

That is, we consider the class

$$F_{d,q}^* = \left\{ f \in W_q^1 \mid f(x) = 0 \text{ if } x_j = 1 \text{ for some } j \in [1, d], \text{ and } \| f \|_{d,q}^* \leq 1 \right\}.$$

Consider the multivariate integration problem

$$\text{INT}_d \, f = \int_{[0,1]^d} f(x) \, dx \quad \text{for } f \in F_{d,q}^*.$$

We approximate $\text{INT}_d \, f$ 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 $f(t_1), f(t_2), \ldots, f(t_n)$ are already computed then the computation of $Q_{d,n} f$ requires just $n - 1$ additions and one division. Since the points $t_1, t_2, \ldots, t_n$ are non-adaptive, $Q_{d,n} f$ can be very efficiently evaluated in parallel since each $f(t_j)$ can be computed on a different processor. Obviously, $Q_{d,n}$ integrates constant functions exactly, even though $1 \notin F_{d,q}^*$.

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, and we will spend considerable time explaining a small part of this theory. 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. The reader is referred to the book of Traub and Werschulz [243] for a thorough discussion.

We now recall Hlawka and Zaremba's identity, see Hlawka [98] and Zaremba [296], which states that for $f \in W_q^1$ we have

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

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 $dx_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} dx_j$. 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

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

For $f \in F_{d,q}^*$, due to the boundary conditions, all terms in Hlawka and Zaremba's identity vanish except the term for $\mathfrak{u} = \{1, 2, \ldots, d\}$. Hence, for $f \in F_{d,q}^*$ we have

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

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

$$e^{\text{wor}}(Q_{d,n}) = \sup_{f \in F_{d,q}^*} \left| \text{INT}_d f - Q_{d,n} f \right| = \text{disc}_p^*(t_1, t_2, \ldots, t_n),$$

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

Now take $n = 0$ and define $Q_{d,0} = 0$. In this case we do not sample the function f. The error of this zero algorithm is the initial worst case error, which is the norm of the linear functional INT_d. It is easy to check that

$$e^{\text{wor}}(0) = e^{\text{wor}}(Q_{d,0}) = \| \text{INT}_d \| = \left(\frac{1}{p+1} \right)^{d/p},$$

which is 1 for $p = \infty$.

Assume for now that $p < \infty$, or (equivalently) that we consider the multivariate integration problem for the class $F_{d,q}^*$ with $q > 1$. Then the initial error goes to 0 exponentially fast with d. This means that the multivariate integration problem for the class $F_{d,q}$ is poorly scaled. To see this point better, take $p = q = 2$ and, as in some financial applications, assume that $d = 360$. Then the initial error is $3^{-180} \approx 10^{-85.9}$. Hence, without computing any function value we know a priori, just by the formulation of the problem, that the absolute value of the integral which we want to approximate is at most 3^{-180}. Hence, as long as $\varepsilon \geq 3^{-180}$, the zero algorithm solves the problem, and the minimal number of function values is $n^{\text{wor}}(\varepsilon, d) = 0$.

One can claim that this is because we introduced boundary conditions; perhaps it might be hard to find practical applications for which functions satisfy these boundary conditions. Let us agree with this criticism, and remove the boundary conditions. So we now consider the class

$$F_{d,q} = \{f \in W_q^1 \mid \| f \|_{d,q} \leq 1\},$$

where the norm is given by

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

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

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

$$e^{\text{wor}}(Q_{d,n}) = \sup_{f \in F_{d,q}} \left| \text{INT}_d f - Q_{d,n} f \right| = \text{disc}_p(t_1, t_2, \ldots, t_n),$$

where the *combined L_p-star discrepancy* disc_p for $p \in [1, \infty]$ is given by

$$\mathrm{disc}_p(t_1, t_2, \ldots, t_n) = \Big(\sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\ldots,d\}} \big[\mathrm{disc}_p^*((t_1)_{\mathfrak{u}}, (t_2)_{\mathfrak{u}}, \ldots, (t_n)_{\mathfrak{u}}) \big]^p \Big)^{1/p},$$

with the usual change to the maximum for $p = \infty$.

What is now the initial error? As before, it is the worst case error of the zero algorithm, which is again the norm of INT_d. However, this time the norm is given in the space W_q^1 without boundary conditions, and

$$e^{\mathrm{wor}}(0) := e^{\mathrm{wor}}(Q_{d,0}) = \| \mathrm{INT}_d \|$$

$$= \Big(\sum_{\mathfrak{u} \subseteq \{1,2,\ldots,d\}} (p+1)^{-|\mathfrak{u}|} \Big)^{1/p} = \Big(\sum_{j=0}^{d} \binom{d}{j} (p+1)^{-j} \Big)^{1/p}$$

$$= \Big(1 + \frac{1}{p+1} \Big)^{d/p}.$$

So the initial error is now exponentially *large* in d. For $p = 2$ and $d = 360$, the initial error is $(4/3)^{180} \approx 10^{+22.5}$. Hence, we switch from almost a zero initial error with the boundary conditions, to almost an infinite initial error without the boundary conditions. Either way, the multivariate integration problem is very badly scaled.

What should we then do? A possible solution is to consider *weighted* discrepancy for which the initial error will be reasonable for all d and $p \in [1, \infty)$. For weighted discrepancy each variable or, more generally, each group $x_{\mathfrak{u}}$ of variables, may play a different role measured by some weight $\gamma_{d,\mathfrak{u}}$. With a proper condition on the weights $\gamma_{d,\mathfrak{u}}$, we achieve reasonable initial errors.

Another point which we want to make is that the absolute error that we discussed so far is only reasonable if the initial error is properly scaled. In fact, for all problems studied in the preceding sections, the initial error was 1. If the initial error is poorly scaled, then it is much better to use the *normalized* error criterion instead of the absolute error.

Now we want to solve the problem to within $\varepsilon e^{\mathrm{wor}}(0)$, with the natural assumption that $\varepsilon \in (0, 1)$. That is, for the normalized error we want to *reduce* the initial error by a factor ε. Then we define the minimal number $n(\varepsilon, d)$ of information operations as the minimal number of function values needed to solve the problem to within $\varepsilon e^{\mathrm{wor}}(0)$ and ask again whether the integration problem is polynomially tractable. That is exactly what we will be doing in Volume II, and therefore we do not pursue this point further here.

We now consider the remaining case $p = \infty$, or equivalently $q = 1$. Note that in this case, we have

$$\mathrm{disc}_\infty^*(t_1, t_2, \ldots, t_n) = \mathrm{disc}_\infty(t_1, t_2, \ldots, t_n),$$

and the multivariate problem is properly scaled for both classes $F_{d,1}^*$ and $F_{d,1}$ since the initial error is 1. Then

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

is the same for both classes, it is just the inverse of the L_∞-discrepancy, the latter being simply called 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. In fact, such a bad behavior was conjectured by Larcher [129]. Therefore it was quite a surprise when a positive result was proved in [91]. More precisely, let

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

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

$$\mathrm{disc}^*_\infty(n,d) \leq C d^{1/2} n^{-1/2} \quad \text{for all } n, d = 1, 2, \ldots.$$

The proof of this bound follows directly from deep results of the theory of empirical processes. In particular, we use a result of Talagrand [231] combined with a result of Haussler [78], and a result of Dudley [51] on the Vapnik–Červonenkis dimension of the family of rational cubes $[0, x)$. The proof is unfortunately non-constructive, and we do not know points for which this bound holds.

The slightly worse upper bound

$$\mathrm{disc}^*_\infty(n,d) \leq 2\sqrt{2} n^{-1/2} \left(d \ln \left(\left\lceil \frac{d n^{1/2}}{2(\ln 2)^{1/2}} \right\rceil + 1 \right) + \ln 2 \right)^{1/2}$$

follows from Hoeffding's inequality and is quite elementary, see also Doerr, Gnewuch and Srivastav [48], and Gnewuch [68]. 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 [91] for details.

In fact, it is possible to have a *semi-construction* of such points by selecting them randomly and checking their star discrepancy. If their star discrepancy satisfies the needed bound, say the last displayed bound times 10, we are done; if not we repeat the process. The probability of failure after k trials will be exponentially small in k so with high probability we will find points with a good bound on the star discrepancy. However, there is a problem with this approach. Namely, today's algorithms for computing the star discrepancy for given points are exponential in d. Hence, for large d the cost will be prohibitively expensive. One of the main open problems for the star discrepancy is to *construct* points with a good bound on the star discrepancy.

One can also use the results on the average behavior of the L_p-star discrepancy for an even integer p to obtain upper bounds for the star discrepancy, see again [91] and Gnewuch [67]. For concrete values of d and n, these upper bounds seem to be better than those presented above.

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

$$n(\varepsilon, d) \leq \left\lceil C^2 d \left(\frac{1}{\varepsilon}\right)^2 \right\rceil \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d = 1, 2, \ldots. \qquad (3.8)$$

This means that we have *polynomial tractability*. Furthermore it was also shown in [91] that there exists a positive number c such that

$$n(\varepsilon, d) \geq cd \ln \varepsilon^{-1} \quad \text{for all } \varepsilon \in (0, 1/64] \text{ and } d = 1, 2, \ldots.$$

In fact, this lower bound holds not only for quasi-Monte Carlo algorithms but in full generality for all algorithms. The last bound was improved by Hinrichs [97], who showed that there exist positive numbers c and ε_0 such that

$$n(\varepsilon, d) \geq cd \varepsilon^{-1} \quad \text{for all } \varepsilon \in (0, \varepsilon_0] \text{ and } d = 1, 2, \ldots.$$

The essence of the lower bounds is that we do *not* have strong polynomial tractability, and the factor d in the bounds on $n(\varepsilon, d)$ cannot be removed.

How about the dependence on ε^{-1}? This is open and seems to be a difficult problem. We know that for a fixed d, the minimal star discrepancy $\mathrm{disc}^*_\infty(n, d)$ behaves much better asymptotically in n. More precisely, we know that for arbitrary d, we have

$$\Omega\left(n^{-1}(\ln n)^{(d-1)/2}\right) = \mathrm{disc}^*_\infty(d, n) = \mathcal{O}\left(n^{-1}(\ln n)^{d-1}\right) \quad \text{as } n \to \infty.$$

The lower bound follows from the lower bound on the minimal L_2-star discrepancy due to Roth [202], whereas the upper bound is due to Halton [75], see also Hammersley [74]. Another major open problem for the star discrepancy is to find the proper power of the logarithm of n in the asymptotic formula for $\mathrm{disc}^*_\infty(n, d)$.

Hence, modulo powers of logarithms, the star discrepancy behaves like n^{-1}, which is optimal since such behavior is already present for the univariate case $d = 1$. This means that $n(\varepsilon, d)$ grows at least as ε^{-1}. Furthermore, for any d, we have

$$\lim_{\varepsilon \to 0} \frac{n(\varepsilon, d)}{\varepsilon^{-(1+\delta)}} = 0 \quad \text{for any } \delta > 0.$$

This may suggest that the exponent 2 of ε^{-1} in the upper bound on $n(\varepsilon, d)$ in (3.8) can be lowered. However, we think that as long as we consider upper bounds of the form $n(\varepsilon, d) \leq Cd^k \varepsilon^{-\alpha}$, the exponent $\alpha \geq 2$ and 2 cannot be improved. This is another open problem related to the star discrepancy, see also Section 3.3.

3.1.6 Example 6: Diagonal Problems for Weighted Spaces

The purpose of this section is to introduce the reader to *weighted* spaces, which play a major role in tractability studies. We will present weighted spaces for relatively

simple diagonal multivariate problems to make the analysis simpler; however, the basic tractability proof technique will be the same as for general linear multivariate problems, as long as we use arbitrary linear functionals as information operations.

Let $L_2([0, 1]^d)$ be the space of square integrable real functions defined on the d-dimensional unit cube with the inner product

$$\langle f, g \rangle_d = \int_{[0,1]^d} f(x)g(x)\,dx.$$

Let

$$\eta_1(x) = 1, \quad \eta_{2k}(x) = \sqrt{2}\sin(2\pi kx), \quad \eta_{2k+1}(x) = \sqrt{2}\cos(2\pi kx)$$

for $x \in [0, 1]$ and $k = 1, 2, \ldots$. For $j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d$, let

$$\eta_{d,j}(x) = \prod_{k=1}^d \eta_{j_k}(x_k) \quad \text{for } x = [x_1, x_2, \ldots, x_d] \in [0, 1]^d.$$

Clearly, $\{\eta_{d,j}\}_{j \in \mathbb{N}^d}$ is an orthonormal basis of $L_2([0, 1]^d)$, i.e., we have $\langle \eta_{d,i}, \eta_{d,j} \rangle_d = \delta_{i,j}$ for all $i, j \in \mathbb{N}^d$.

For the univariate case $d = 1$, we define a *diagonal* operator as a continuous linear operator $S_1 : L_2([0, 1]) \to L_2([0, 1])$ such that

$$S_1 \eta_k = \sigma_k \eta_k \quad \text{for all } k = 1, 2, \ldots.$$

Here we assume that

$$1 = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq \cdots \geq 0.$$

Note that $\|S_1\|_{L_2([0,1]) \to L_2([0,1])} = \sigma_1 = 1$, and therefore the problem is well normalized.

For $d \geq 2$, we define a *diagonal* operator as the tensor product of d copies of S_1, i.e., $S_d : L_2([0, 1]^d) \to L_2([0, 1]^d)$ is a continuous linear operator such that

$$S_d \eta_{d,j} = \sigma_{d,j} \eta_{d,j} \quad \text{with } \sigma_{d,j} = \prod_{k=1}^d \sigma_{j_k} \text{ for all } j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d.$$

Clearly, $\|S_d\|_{L_2([0,1]^d) \to L_2([0,1]^d)} = 1$ for all d.

We want to approximate $S_d f$ for f belonging to some class F_d that is the unit ball of a *weighted* space $L_{2,\gamma}([0, 1]^d)$.

To explain the idea of weighted spaces, we first recall the *ANOVA* (for "analysis of variance") decomposition of functions from $L_2([0, 1]^d)$. See Efron and Stein [55], Sobol [222], Ho and Rabitz [99] and Rabitz and Alis [200] for this and related representations of functions. The ANOVA decomposition is widely used in statistics, as well as in the study of quasi-Monte Carlo algorithms for multivariate integration. It represents

a function f of d variables as a sum of functions in which each term depends only on one specific group of variables. More precisely, we have

$$f(x) = f_\emptyset + \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\ldots,d\}} f_\mathfrak{u}(x_\mathfrak{u}), \tag{3.9}$$

where $f_\emptyset = \int_{[0,1]^d} f(x)\, dx$, and the rest of the ANOVA terms $f_\mathfrak{u}$ are defined recursively with respect to the increasing cardinality of $|\mathfrak{u}|$, i.e.,

$$f_\mathfrak{u}(x_\mathfrak{u}) = \int_{[0,1]^{d-|\mathfrak{u}|}} f(x)\, dx_{-\mathfrak{u}} - \sum_{\mathfrak{v} \subseteq \mathfrak{u}, \mathfrak{v} \neq \mathfrak{u}} f_\mathfrak{v}(x_\mathfrak{v}),$$

where $x_{-\mathfrak{u}}$ denotes the vector $x_{\{1,2,\ldots,d\}-\mathfrak{u}}$. For $\mathfrak{u} = \{1, 2, \ldots, d\}$, the integral is replaced by $f(x)$. Here we use the same notation as in the discrepancy section, namely, $x_\mathfrak{u}$ denotes the vector from $[0, 1]^{|\mathfrak{u}|}$ whose components are those components of x whose indices are in $\mathfrak{u}$.

Clearly $f_\mathfrak{u}$ depends only on $x_\mathfrak{u}$, since we integrate over all variables not present in $\mathfrak{u}$ and the summation involves functions that depend only on some variables from $\mathfrak{u}$. The recursive definition is done in a way such that we first know $f_\emptyset$, then $f_{\{i\}}$ for all $i = 1, 2, \ldots, d$, then $f_{\{i,j\}}$ for all $1 \le i < j \le d$, and so on. Finally we get 2^d ANOVA terms. From the definition of $f_\mathfrak{u}$, we easily conclude (3.9) by taking $\mathfrak{u} = \{1, 2, \ldots, d\}$.

We give some examples. The easy case is when $d = 1$. Then $x = x_1$, and we have only two terms, $f_\emptyset = \int_0^1 f(t)\, dt$ and $f_{\{1\}}(x_1) = f(x) - f_\emptyset$. Let $d = 2$. Then $x = [x_1, x_2]$ and we have four terms

$$f_\emptyset = \int_0^1 \int_0^1 f(t_1, t_2)\, dt_1\, dt_2,$$

$$f_{\{1\}}(x_1) = \int_0^1 f(x_1, t)\, dt - f_\emptyset,$$

$$f_{\{2\}}(x_2) = \int_0^1 f(t, x_2)\, dt - f_\emptyset,$$

$$f_{\{1,2\}}(x_1, x_2) = f(x_1, x_2) - f_\emptyset - f_{\{1\}}(x_1) - f_{\{2\}}(x_2).$$

For $f(x) = 1 + x_1 x_2 + \sin(2\pi x_1) + \cos(2\pi x_2)$ we have

$$f_\emptyset = \tfrac{5}{4},$$

$$f_{\{1\}}(x_1) = -\tfrac{1}{4} + \tfrac{1}{2}x_1 + \sin(2\pi x_1),$$

$$f_{\{2\}}(x_2) = -\tfrac{1}{4} + \tfrac{1}{2}x_2 + \cos(2\pi x_2),$$

$$f_{\{1,2\}}(x_1, x_2) = f(x_1, x_2) - \tfrac{3}{4} - \tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 - \sin(2\pi x_1) - \cos(2\pi x_2).$$

Let us return to the case of general d. It follows easily that the ANOVA terms are orthogonal in the L_2 sense, i.e.,

$$\int_{[0,1]^d} f_\mathfrak{u}(x_\mathfrak{u}) f_\mathfrak{v}(x_\mathfrak{v})\, dx = 0 \quad \text{for all } \mathfrak{u} \neq \mathfrak{v}.$$

In particular, this means that the integrals of all $f_{\mathfrak{u}}$ are 0 if $\mathfrak{u} \neq \emptyset$. We also have

$$\|f\|_d^2 = \sum_{\mathfrak{u} \subseteq \{1,2,\ldots,d\}} \|f_{\mathfrak{u}}\|_d^2. \qquad (3.10)$$

We see then that the contributions of all $f_{\mathfrak{u}}$ to the norm of f are the same. Furthermore, if we compute the variance $\sigma^2(f)$, defined as

$$\sigma^2(f) = \int_{[0,1]^d} f^2(x)\,dx - \left(\int_{[0,1]^d} f(x)\,dx\right)^2,$$

then the variance of f is a sum of the variances of the separate ANOVA terms $f_{\mathfrak{u}}$,

$$\sigma^2(f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\ldots,d\}} \sigma^2(f_{\mathfrak{u}}).$$

The last property was used in many papers to explain the efficient error behavior of quasi-Monte Carlo algorithms for multivariate integration, see for example the papers of Caflisch, Morokoff and Owen [25], Sobol [222], and Wang and Fang [256]. The point was that some $\sigma(f_{\mathfrak{u}})$ are small or even 0 for functions arising in practical computations, especially for sets $\mathfrak{u}$ of large cardinality.

This is our point of departure on the road to weighted spaces. We want to treat the influence of each $f_{\mathfrak{u}}$ on the norm of f separately, and to model situations for which we may know a priori that some terms $f_{\mathfrak{u}}$ in the ANOVA decomposition of f are small or even negligible. In particular, we wish to be able to model cases for which we know that f is a sum of functions of at most ω variables with ω much smaller than d, or that the influence of the first variable is more important than the second variable which is, in turn, more important than the third variable and so on.

This is achieved by introducing *weights* $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ for $\mathfrak{u} \subseteq \{1, 2, \ldots, d\}$ and $d = 1, 2, \ldots$. We always assume that $\gamma_{d,\mathfrak{u}} \geq 0$. We define the *weighted* space $L_{2,\gamma}([0,1]^d)$ as the space of square integrable functions with the finite inner product

$$\langle f, g \rangle_{d,\gamma} = \sum_{\mathfrak{u} \subseteq \{1,2,\ldots,d\}} \frac{1}{\gamma_{d,\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} f_{\mathfrak{u}}(x_{\mathfrak{u}}) g_{\mathfrak{u}}(x_{\mathfrak{u}})\,dx_{\mathfrak{u}},$$

and with the convention that $0/0 = 0$. That is, $\gamma_{d,\mathfrak{u}} = 0$ implies that $f_{\mathfrak{u}} = g_{\mathfrak{u}} = 0$ and the term for $\mathfrak{u}$ disappears from the sum. For all positive $\gamma_{d,\mathfrak{u}}$, we have $L_{2,\gamma}([0,1]^d) = L_2([0,1]^d)$ but the norms of f in these two spaces can be quite different since

$$\frac{1}{\max_{\mathfrak{u}} \sqrt{\gamma_{d,\mathfrak{u}}}} \|f\|_d \leq \|f\|_{d,\gamma} \leq \frac{1}{\min_{\mathfrak{u}} \sqrt{\gamma_{d,\mathfrak{u}}}} \|f\|_d \quad \text{for all } f \in L_{2,\gamma}([0,1]^d).$$

However, if there is a weight $\gamma_{d,\mathfrak{u}} = 0$ then the spaces $L_{2,\gamma}([0,1]^d)$ and $L_2([0,1]^d)$ are *not* equivalent, and $L_{2,\gamma}([0,1]^d)$ is a proper subset of $L_2([0,1]^d)$. In the extreme case, when all $\gamma_{d,\mathfrak{u}} = 0$ we have $L_{2,\gamma}([0,1]^d) = \{0\}$.

Take the class $F_d = F_{d,\gamma}$ as the unit ball of the space $L_{2,\gamma}([0, 1]^d)$, i.e.,

$$F_{d,\gamma} = \{f \in L_{2,\gamma}([0, 1]^d) \mid \|f\|_{d,\gamma} \leq 1\}.$$

Suppose for a moment that $\gamma_{d,\mathfrak{u}}$ is small for some $\mathfrak{u}$. Then for $f \in F_{d,\gamma}$ we know a priori that $\|f_{\mathfrak{u}}\|_2 \leq \sqrt{\gamma_{d,\mathfrak{u}}}$, and that the ANOVA term $f_{\mathfrak{u}}$ does not play much of a role. Furthermore, if $\gamma_{d,\mathfrak{u}} = 0$ we know a priori that $f_{\mathfrak{u}} = 0$.

By a proper choice of weights $\gamma_{d,\mathfrak{u}}$, we can model our a priori knowledge about f. For instance, suppose we know that f is a sum of functions depending on at most ω variables. Then we choose *finite-order* weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ which are defined by assuming that

$$\gamma_{d,\mathfrak{u}} = 0 \quad \text{for all } d \text{ and } \mathfrak{u} \text{ for which } |\mathfrak{u}| > \omega.$$

If we know that the ith variable is more important than the $(i + 1)$st variable then we choose *product* weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ which are defined by assuming that

$$\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$$

for some non-increasing $\gamma_{d,1} \geq \gamma_{d,2} \geq \cdots \geq \gamma_{d,d} \geq 0$. Finally, if we choose $\gamma_{d,\mathfrak{u}} = 1$ for all $\mathfrak{u}$, then we are back to the unweighted case for which all ANOVA terms play the same role and $L_{2,\gamma}([0, 1]^d) = L_2([0, 1]^d)$ with $\|f\|_{d,\gamma} = \|f\|_d$ for all f.

We are ready to return to the diagonal multivariate problem, defined as approximation of $S_d f$ for f from $F_{d,\gamma}$. We want to determine conditions on an arbitrary set of weights γ that guarantee tractability. For simplicity, we restrict ourselves in this section to polynomial tractability.

We stress that if we consider the diagonal operator S_d over the class $F_{d,\gamma}$, the norm $\|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)}$, which is the initial error and the worst case error of the zero algorithm, depends on γ. Indeed, for the extreme case of all $\gamma_{d,\mathfrak{u}} = 0$ we have $\|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)} = 0$ since $F_{d,\gamma} = \{0\}$. On the other hand, if we take $\gamma_{d,\mathfrak{u}} \equiv \alpha > 0$ then $\|f\|_{d,\gamma} = \alpha^{-1/2}\|f\|_d$ and $\|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)} = \sqrt{\alpha}$. To avoid the scaling problem, we will use the normalized error and compute approximation of $S_d f$ to within $\varepsilon \|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)}$.

As in Section 3.1.4, we assume that arbitrary linear functionals can be used. As before, an algorithm using n information operations has the form

$$A_n(f) = \varphi(L_1(f), L_2(f), \dots, L_n(f))$$

for some mapping φ, where $\varphi \colon \mathbb{R}^n \to L_2([0, 1]^d)$. The worst case error of A_n is now given by

$$e^{\text{wor}}(A_n) = \sup_{f \in F_{d,\gamma}} \|S_d f - A_n(f)\|_d,$$

whereas the minimal[3] number of information operations is given by

$$n^{\text{wor}}(\varepsilon, S_d, F_{d,\gamma}) = \min\{n \mid e^{\text{wor}}(A_n) \leq \varepsilon \|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)} \text{ for some } A_n\}.$$

---

[3]Here and elsewhere in the book, by the minimum over an empty set we mean ∞.

We want to find necessary and sufficient conditions on the diagonal operators S_d in terms of the numbers σ_j and on the weights $\gamma_{d,\mathfrak{u}}$ to obtain *polynomial tractability*. More precisely, we say that the problem $S = \{S_d\}$ is *polynomially tractable* if there exist three non-negative numbers C, p and q such that

$$n^{\mathrm{wor}}(\varepsilon, S_d, F_{d,\gamma}) \le C\varepsilon^{-p} d^q \quad \text{for all } \varepsilon \in (0, 1) \text{ and for all } d = 1, 2, \dots. \quad (3.11)$$

If $q = 0$ in the bound above, then we say that the problem is *strongly polynomially tractable*.

As in Section 3.1.4, we need to know the ordered eigenvalues $\{\lambda_{d,\gamma,j}\}_{j \in \mathbb{N}}$ of the operator $W_d = S_d^* S_d : F_{d,\gamma} \to F_{d,\gamma}$ since $\|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)} = \sqrt{\lambda_{d,\gamma,1}}$ and

$$n^{\mathrm{wor}}(\varepsilon, S_d, F_{d,\gamma}) = \min\left\{ n \mid \sqrt{\lambda_{d,\gamma,n+1}} \le \varepsilon \sqrt{\lambda_{d,\gamma,1}} \right\}.$$

Polynomial tractability depends on the summability of some powers of $\lambda_{d,\gamma,j}$. More precisely, for a positive τ, define

$$M_{d,\gamma}(\tau) = \left(\sum_{j=1}^{\infty} \left(\frac{\lambda_{d,\gamma,j}}{\lambda_{d,\gamma,1}} \right)^{\tau} \right)^{1/\tau},$$

with the convention that if the series above is not convergent then we formally set $M_{d,\gamma}(\tau) = \infty$.

It is known that polynomial tractability holds iff there exist two numbers s and τ such that $s \ge 0$ and $\tau > 0$ and

$$M := \sup_{d=1,2,\dots} d^{-s} M_{d,\gamma}(\tau) < \infty,$$

see Theorem 5.2 of Chapter 5. If this holds then (3.11) holds with $C = M^{\tau}$, $p = 2\tau$, and $q = s\tau$. Furthermore, strong polynomial tractability holds iff $s = 0$ in the formula above.

We first compute the ANOVA terms of $\eta_{d,j}$ for a given $j = [j_1, j_2, \dots, j_d] \in \mathbb{N}^d$. Define the set $\mathfrak{u}(j)$ as the set of all indices k for which $j_k \ge 2$, i.e.,

$$\mathfrak{u}(j) = \{k \mid j_k \ge 2\}.$$

In particular, for $j = [1, 1, \dots, 1]$ we have $\mathfrak{u}(j) = \emptyset$, and for $j = [j_1, j_2, \dots, j_d]$ with all $j_k \ge 2$, we have $\mathfrak{u}(j) = \{1, 2, \dots, d\}$.

We claim that

$$\left(\eta_{d,j} \right)_{\mathfrak{u}} = \delta(\mathfrak{u}, \mathfrak{u}(j)) \eta_{d,j},$$

i.e., all ANOVA terms of $\eta_{d,j}$ are 0 for $\mathfrak{u} \ne \mathfrak{u}(j)$, and the ANOVA term for $\mathfrak{u} = \mathfrak{u}(j)$ is just the function $\eta_{d,j}$ itself. Indeed, if we take $\mathfrak{u} \subseteq \mathfrak{u}(j)$ and $\mathfrak{u} \ne \mathfrak{u}(j)$, then all terms in the definition of $(\eta_{d,j})_{\mathfrak{u}}$ involve integration over a variable (say, k) for which $j_k \ge 2$ and therefore the corresponding integral is 0, and hence $(\eta_{d,j})_{\mathfrak{u}} = 0$. For $\mathfrak{u} = \mathfrak{u}(j)$, we integrate only over variables not present in $\mathfrak{u}$ and the corresponding

integral reproduces $\eta_{d,j}$ and therefore $(\eta_{d,j})_{\mathfrak{u}(j)} = \eta_{d,j}$. From (3.10) we conclude that all ANOVA terms $(\eta_{d,j})_{\mathfrak{u}}$ are 0 for $\mathfrak{u} \neq \mathfrak{u}(j)$, as claimed.

For all $j \in \mathbb{N}^d$, define

$$\eta_{d,\gamma,j} = \sqrt{\gamma_{d,\mathfrak{u}(j)}}\,\eta_{d,j}.$$

Then for all $i, j \in \mathbb{N}^d$, we have

$$\langle \eta_{d,\gamma,i}, \eta_{d,\gamma,j} \rangle_{d,\gamma} = \sum_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \frac{1}{\gamma_{d,\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \left(\eta_{d,\gamma,i}\right)_{\mathfrak{u}}(x_{\mathfrak{u}}) \left(\eta_{d,\gamma,j}\right)_{\mathfrak{u}}(x_{\mathfrak{u}})\,dx_{\mathfrak{u}} = \delta_{i,j}.$$

Hence, $\{\eta_{d,\gamma,j}\}_{j \in \mathbb{N}^d}$ is an orthonormal basis of $L_{2,\gamma}([0,1]^d)$, which obviously is orthogonal in $L_2([0,1]^d)$. We then have

$$\begin{aligned}
\sqrt{\gamma_{d,\mathfrak{u}(i)}\gamma_{d,\mathfrak{u}(j)}}\,\sigma_{d,i}\sigma_{d,j}\delta_{i,j} &= \langle S_d\,\eta_{d,\gamma,i}, S_d\,\eta_{d,\gamma,j} \rangle_d \\
&= \langle \eta_{d,\gamma,i}, S_d^* S_d\,\eta_{d,\gamma,j} \rangle_{d,\gamma} = \langle \eta_{d,\gamma,i}, W_d\,\eta_{d,\gamma,j} \rangle_{d,\gamma}.
\end{aligned}$$

Thus, $W_d\,\eta_{d,\gamma,j}$ is orthogonal to all $\eta_{d,\gamma,i}$ with $i \neq j$, and therefore

$$W_d\,\eta_{d,\gamma,j} = \gamma_{d,\mathfrak{u}(j)}\sigma_{d,j}^2\,\eta_{d,\gamma,j} \quad \text{for all } j \in \mathbb{N}^d.$$

This proves that the eigenvalues of W_d are

$$\{\lambda_{d,\gamma,j}\}_{j \in \mathbb{N}} = \{\gamma_{d,\mathfrak{u}(j)}\sigma_{d,j}^2\}_{j \in \mathbb{N}^d}.$$

Note that $\mathfrak{u}(j) = \emptyset$ only for $j = [1,1,\dots,1]$, whereas if $\mathfrak{u}$ is not empty then $\mathfrak{u}(j) = \mathfrak{u}$ for all $j = [j_1, j_2, \dots, j_d]$ for which $j_k \geq 2$ only for all $k \in \mathfrak{u}$. Recall that $\sigma_{d,j} = \prod_{k=1}^d \sigma_{j_k}$ with the ordered σ_j. From this we conclude that

$$\|S_d\|_{F_{d,\gamma} \to L_2([0,1]^d)} = \max_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \sqrt{\gamma_{d,\mathfrak{u}}}\,\sigma_2^{|\mathfrak{u}|}.$$

This allows us to compute the sums of powers of the eigenvalues as

$$M_{d,\gamma}(\tau) = \frac{\left(\gamma_{d,\emptyset}^{\tau} + \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\dots,d\}} \gamma_{d,\mathfrak{u}}^{\tau} \left(\sum_{j=2}^{\infty} \sigma_j^{2\tau}\right)^{|\mathfrak{u}|}\right)^{1/\tau}}{\max_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \gamma_{d,\mathfrak{u}}\sigma_2^{2|\mathfrak{u}|}}.$$

Hence, polynomial tractability holds iff

$$\sup_{d=1,2,\dots} d^{-s} \frac{\left(\gamma_{d,\emptyset}^{\tau} + \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,2,\dots,d\}} \gamma_{d,\mathfrak{u}}^{\tau} \left(\sum_{j=2}^{\infty} \sigma_j^{2\tau}\right)^{|\mathfrak{u}|}\right)^{1/\tau}}{\max_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \gamma_{d,\mathfrak{u}}\sigma_2^{2|\mathfrak{u}|}} < \infty$$

for some $s \geq 0$ and $\tau > 0$. Strong polynomial tractability holds iff we can take $s = 0$ in the formula above.

We first provide some trivial conditions which imply strong polynomial tractability. The first is when all $\gamma_{d,\mathfrak{u}} = 0$ and the second when $\sigma_2 = 0$. For the first case, $n^{\text{wor}}(\varepsilon, S_d, F_{d,\gamma}) = 0$ whereas for the second case $n^{\text{wor}}(\varepsilon, S_d, F_{d,\gamma}) \leq 1$. From now on assume then that not all $\gamma_{d,\mathfrak{u}}$ are 0 and that $\sigma_2 > 0$. Then polynomial tractability requires, in particular, that $\sum_{j=2}^{\infty} \sigma_j^{2\tau}$ be finite. This means that $\sigma_n = \mathcal{O}(n^{-1/(2\tau)})$. Hence, polynomial decay of $\{\sigma_j\}_{j=1}^{\infty}$ is a necessary condition for polynomial tractability.

Assume then also that $\{\sigma_j\}$ decays polynomially. Consider now the case of finite-order weights, i.e., $\gamma_{d,\mathfrak{u}} = 0$ if $|\mathfrak{u}| > \omega$. Observe that the maximal number of non-zero $\gamma_{d,\mathfrak{u}}$ is now of order d^{ω}. In fact, it can be shown that it is no more than $2d^{\omega}$. Let $C_{\tau} = \left(\sum_{j=2}^{\infty} \sigma_j^{2\tau}\right)^{1/\tau} < \infty$. Then

$$\sup_d d^{-s} M_{d,\gamma}(\tau) \leq 2^{1/\tau} \sup_d d^{-s+\omega/\tau} \frac{\max_{\mathfrak{u}:|\mathfrak{u}|\leq\omega}\left(\gamma_{d,\mathfrak{u}} C_{\tau}^{|\mathfrak{u}|}\right)}{\max_{\mathfrak{u}:|\mathfrak{u}|\leq\omega}\left(\gamma_{d,\mathfrak{u}} \sigma_2^{2|\mathfrak{u}|}\right)}$$

$$\leq \max_d d^{-s+\omega/\tau} \frac{\max\left(1, C_{\tau}^{\omega}\right)}{\min\left(1, \sigma_2^{2\omega}\right)} < \infty$$

if we take $s \geq \omega/\tau$. Hence, finite-order weights *imply* polynomial tractability.

Now assume product weights, $\gamma_{d,\mathfrak{u}} = \prod_{j\in\mathfrak{u}}^{d} \gamma_{d,j}$ with $\gamma_{d,\emptyset} = 1 \geq \gamma_{d,1}$. Then

$$\sup_d d^{-s} M_{d,\gamma}(\tau) \leq \sup_d d^{-s} \prod_{j=1}^{d}\left(1 + C_{\tau}^{\tau}\gamma_{d,j}^{\tau}\right)^{1/\tau}.$$

It is easy to check that we get strong polynomial tractability iff

$$\limsup_d \sum_{j=1}^{d} \gamma_{d,j}^{\tau} < \infty,$$

and polynomial tractability iff

$$\limsup_d \frac{\sum_{j=1}^{d} \gamma_{d,j}^{\tau}}{\ln d} < \infty.$$

Hence, polynomial or even strong polynomial tractability is indeed possible for some weights such as finite-order or product weights. On the other hand, note that for $\gamma_{d,\mathfrak{u}} \equiv 1$, we have polynomial intractability since $M_{d,\gamma}(\tau) = (1 + C_{\tau}^{\tau})^d$ is exponentially large in d. Furthermore, if we take $\sigma_2 = \sigma_1 = 1$, then for the d-dimensional case we have 2^d eigenvalues of W_d equal to 1, and

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

which means that the problem is intractable.

We hope we have convinced the reader that weighted spaces may be a natural choice of spaces to obtain tractability. Finally we want to stress that the ANOVA decomposition is just one possible source of weighted spaces. As we shall see, any Hilbert space of functions can be used for an ANOVA-type decomposition of functions and for introducing weighted spaces for which tractability holds under some conditions on the weights.

3.2 Tractability in Other Settings

In the previous section, we illustrated tractability of selected multivariate problems in the worst case setting. In this setting, the error of an algorithm is defined by its worst performance or, speaking more mathematically, by taking the supremum over all problem elements. In our case, problem elements are d-variate functions from some class. We stress that the worst case setting, although quite conservative, has a big advantage for problems for which we can establish positive results. Suppose we have, say, polynomial tractability in the worst case setting with the absolute error criterion. Then we can claim that the error is at most ε for *all* problem elements using polynomially many information operations. In other settings, an error of at most ε will *not* hold for all problem elements but will hold only on the average or with high probability with respect to all problem elements. Only in the worst case setting can we guarantee that the error is at most ε for a specific problem element and for one run of the algorithm. But there is a price for this desirable property. Namely, many results in the worst case setting may be overly pessimistic. In particular, the reader may suspect that at least some negative tractability results are due to the fact that the error defined in the worst case setting is too pessimistic. Indeed, if we switch to more lenient ways of defining the error of an algorithm, some negative results of the worst case setting disappear. Again we stress that this may happen due to a weaker error assurance, which unfortunately tells us nothing about the error for a particular problem element.

We indicated in the previous section that intractability (or the curse of dimensionality) of a multivariate problem in the worst case setting may sometimes be broken by using weights to shrink the class of problem elements. This led us to study weighted spaces, seeking necessary and sufficient conditions on the weights to obtain weak or polynomial tractability still in the worst case setting. Another way of breaking worst case intractability is to switch to a different setting for the same class of problem elements. As we shall see, for some multivariate problems this approach indeed works, and a more lenient definition of an algorithm's error allows us to attain weak or polynomial tractability. However, some other multivariate problems are so difficult that they remain intractable in *all* settings studied here. We will encounter such examples in the course of the book but here we only mention that multivariate approximation for the class of continuous d-variate functions is intractable in the worst case, average case (with the isotropic Wiener measure), randomized, and probabilistic settings. So switching to another setting is not always a remedy for intractability. In any case, it

is of practical and mathematical interest to identify multivariate problems for which intractability in the worst case setting can be broken by switching to another setting, as well as multivariate problems which remain intractable in all settings. This will be a major subject of this book. Here, we only illustrate what can happen in other settings for some problems.

We will study three more settings in this book, and illustrate them in this section for selected multivariate problems. We begin with the *average case* setting in which the error of an algorithm is defined by its average performance, and the *probabilistic* setting in which the error of an algorithm is defined by its worst performance over a set of measure $1 - \delta$, where δ is a (small) positive parameter from $(0, 1)$. In both the average case and probabilistic settings, we need to assume that problem elements are distributed according to some probability measure. In our case, we deal with problem elements that are d-variate functions from a linear space, which is typically infinite dimensional. Therefore we use measure theory over infinite dimensional spaces. We usually assume a Gaussian measure; the reader who wants to know more about Gaussian measures on infinite dimensional spaces is referred to the books of Kuo [118], and Vakhania, Tarieladze and Chobanyan [251]. For the reader's convenience we also survey the properties of Gaussian measures needed for tractability studies in Appendix B.

Tractability in the average case setting is defined in the same way as in the worst case setting. The only difference is what we mean by the error of an algorithm. In the probabilistic setting, the situation is a little more complicated since we have one more parameter δ, which is used in the definition of the error of an algorithm. It is desirable to also vary δ and study tractability in terms of three parameters ε, δ and d, the first two parameters with respect to the error performance, and the third with respect to the number of variables. As we shall see, sometimes the parameter δ plays a less important role than the other parameters. Indeed, sometimes we may have positive results even if we define tractability as polynomial dependence on ε^{-1} and d and only logarithmic dependence on δ^{-1}. This will depend on the error criterion. Indeed, as long as we use the absolute or normalized error criteria, the dependence on δ is usually expressed by a small power of $\ln \delta^{-1}$ or *no* dependence on δ at all. But as we shall also see, for some other error criteria, such as the very important relative error criteria, the situation changes, and the parameter δ will play more or less the same role as ε. In this case, positive tractability results will be possible only if we allow polynomial dependence on δ^{-1}. These points will be illustrated in Sections 3.2.5 and 3.2.6. The three settings – worst case, average case and probabilistic – deal with *deterministic* algorithms. That is, there is no random element in the choice of an algorithm and no random element in the choice of output. Put differently, we will get always the same output for the same problem element.

But *randomized* algorithms are widely used and have proved to be efficient for a number of computational problems. The most famous example is probably the classical Monte Carlo algorithm and its many modifications for multivariate integration. Monte Carlo is widely used, especially in computational physics and chemistry. Hence we will be also studying the *randomized* setting, in which we allow randomized algorithms

whose error is defined by the average performance with respect to randomization and the worst case performance with respect to problem elements. We also study tractability in this setting, with the hope that randomization will allow us to find positive tractability results that are impossible to obtain in the worst case setting. As we shall see, this is indeed the case for some multivariate problems such as multivariate integration over some classes of functions. However, it is also true that for some other multivariate problems, such as multivariate approximation defined over Hilbert spaces, randomization does not help and we have essentially the same results as in the worst case setting. These points will be illustrated in Sections 3.2.7 and 3.2.8.

In general, we will see that error estimates and tractability results in the average case setting are no worse than those in the randomized setting which, in turn, are no worse than those in the worst case setting. This may be schematized as

$$\text{AVERAGE} \ \leq \ \text{RANDOMIZED} \ \leq \ \text{WORST}.$$

Sometimes we have equality or near-equality between all three of them, whereas in other cases error estimates and tractability results are much better in one or two of the settings.

We finally remark that tractability results in the average case setting obviously depend on the choice of a measure. For some measures, we can trivialize the problem. Indeed, take an extreme case of an atomic measure on one problem element. Then although the original problem may have been defined over a large infinite dimensional space, the atomic measure makes it trivial as the problem defined only on one problem element. That is why we must be careful to choose a proper measure in the average case setting, avoiding measures that trivialize otherwise interesting problems.

3.2.1 Average Case Setting

We present two multivariate problems in the average case setting. The first is Gaussian integration with the isotropic Wiener measure and with algorithms using only function values. The second is approximation with the folded Wiener sheet measure, with algorithms using arbitrary linear functionals.

3.2.2 Example 7: Isotropic Wiener Measure

We take F_d as the Banach space of those continuous real functions defined over $\mathbb{R}^d$ for which

$$\|f\| = \sup_{x \in \mathbb{R}^d} |f(x)|(1 + \|x\|_2)^{-\beta} < \infty,$$

where $\|\cdot\|_2$ denotes the Euclidean norm of real vectors, i.e., for $x = [x_1, x_2, \ldots, x_d]$ we have $\|x\|_2 = \sqrt{\sum_{j=1}^{d} x_j^2}$. Here β is any real number such that $\beta > \frac{1}{2}$.

The space F_d is equipped with the *isotropic Wiener* measure w_d, which is a zero-mean Gaussian measure whose *covariance kernel* is

$$K_d(x, y) = \tfrac{1}{2} \left(\|x\|_2 + \|y\|_2 - \|x - y\|_2 \right).$$

Knowing the mean element and the covariance kernel of w_d, we know how to integrate linear and bilinear forms. Namely,

$$\int_{F_d} f(x) w_d(\mathrm{d}f) = 0 \qquad \text{for all } x \in \mathbb{R}^d,$$

$$\int_{F_d} f(x) f(y) w_d(\mathrm{d}f) = K_d(x, y) \quad \text{for all } x, y \in \mathbb{R}^d.$$

The isotropic Wiener measure is also called *Brownian motion in Lévy's sense*[4]. We add that we will be also studying the *Wiener sheet* measure, which is a zero-mean Gaussian measure with the covariance kernel $\widetilde{K}_d(x, y) = \prod_{j=1}^{d} \widetilde{K}_1(x_j, y_j)$, where $\widetilde{K}_1(x, y) = \min(|x|, |y|)$ for $xy > 0$, and $\widetilde{K}_1(x, y) = 0$ otherwise. As we shall see, the Wiener sheet measure is related to discrepancy. For $d = 1$, we have $K_1 = \widetilde{K}_1$, and so the isotropic Wiener and Wiener sheet measures are the same. For $d \geq 2$, we have $K_d \neq \widetilde{K}_d$, and so the isotropic Wiener and Wiener sheet measures are different. As we shall see, we obtain quite different tractability results for these two measures.

We consider *Gaussian* integration, which is defined as the approximation of

$$\mathrm{INT}_d f = \int_{\mathbb{R}^d} f(x) \varrho_d(x) \, \mathrm{d}x \quad \text{for all } f \in F_d,$$

with the standard Gaussian weight,

$$\varrho_d(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\tfrac{1}{2} \|x\|_2^2\right) \quad \text{for all } x \in \mathbb{R}^d.$$

We approximate $\mathrm{INT}_d f$ by algorithms using function values. As in the worst case setting, let $A_n(f) = \varphi_n(f(x_1), f(x_2), \ldots, f(x_n))$ be an algorithm using n adaptive function values. In the average case setting, we assume for simplicity that A_n is measurable[5] and define the average case error of A_n by its average performance

$$e^{\mathrm{avg}}(A_n) = \left(\int_{F_d} (\mathrm{INT}_d f - A_n(f))^2 \, w_d(\mathrm{d}f) \right)^{1/2}.$$

We would like to convince the reader that the analysis of the average case setting with a Gaussian measure is not too difficult, perhaps being even easier than the analysis of

---

[4]Lévy [133], see also Ciesielski [30], proved that $w_d(F_d) = 1$ for $\beta > \tfrac{1}{2}$. That is, the isotropic Wiener measure is concentrated on continuous functions that can go to infinity no faster than $(1 + \|x\|_2)^{\beta}$.

[5]It is also possible to define the average case error of A_n without assuming its measurability, which will be discussed in Chapter 4. In any case, the measurability assumption is quite weak and there is not much loss in studying only measurable algorithms.

the worst case setting. As a partial proof of this claim, we now show that as long as A_n is a linear algorithm, its average case error can be easily computed. Indeed, assume that

$$A_n(f) = \sum_{j=1}^{n} a_j f(x_j),$$

for some $a_j \in \mathbb{R}$ and for some non-adaptively chosen $x_j \in \mathbb{R}^d$. Then

$$e^{\mathrm{avg}}(A_n)^2 = \int_{F_d} (\mathrm{INT}_d f)^2 \, w_d(\mathrm{d}f) - 2 \sum_{j=1}^{n} a_j \int_{F_d} f(x_j) \mathrm{INT}_d f \, w_d(\mathrm{d}f)$$

$$+ \sum_{i,j=1}^{n} a_i a_j \int_{F_d} f(x_i) f(x_j) w_d(\mathrm{d}f).$$

We note that the last integral within the double sum is simply $K_d(x_i, x_j)$, whereas the integral within the single sum is

$$\int_{F_d} f(x_j) \mathrm{INT}_d f \, w_d(\mathrm{d}f) = \int_{\mathbb{R}^d} \left(\int_{F_d} f(x_j) f(x) w_d(\mathrm{d}f) \right) \varrho_d(x) \, \mathrm{d}x$$

$$= \int_{\mathbb{R}^d} K_d(x_j, x) \varrho_d(x) \, \mathrm{d}x.$$

Finally, the first integral can be computed by noting

$$\int_{F_d} (\mathrm{INT}_d f)^2 \, w_d(\mathrm{d}f) = \int_{F_d} \int_{\mathbb{R}^d} f(x) \varrho_d(x) \, \mathrm{d}x \int_{\mathbb{R}^d} f(y) \varrho_d(y) \, \mathrm{d}y \, w_d(\mathrm{d}f)$$

$$= \int_{\mathbb{R}^{2d}} \left(\int_{F_d} f(x) f(y) w_d(\mathrm{d}f) \right) \varrho_d(x) \varrho_d(y) \, \mathrm{d}x \, \mathrm{d}y$$

$$= \int_{\mathbb{R}^{2d}} K_d(x, y) \varrho_d(x) \varrho_d(y) \, \mathrm{d}x \, \mathrm{d}y.$$

Hence, without much work, we have the explicit formula

$$e^{\mathrm{avg}}(A_n)^2 = \int_{\mathbb{R}^{2d}} K_d(x, y) \varrho_d(x) \varrho_d(y) \, \mathrm{d}x \, \mathrm{d}y - 2 \sum_{j=1}^{n} a_j \int_{\mathbb{R}^d} K_d(x_j, x) \varrho_d(x) \, \mathrm{d}x$$

$$+ \sum_{i,j=1}^{n} a_i a_j K_d(x_i, x_j)$$

for the average case error of the linear algorithm A_n. This formula allows us to compute the initial average case error. As always, the initial error depends only on the formulation of the problem without sampling the function. Clearly, this is the average case error for the zero algorithm (i.e., $a_j = 0$ for all j), which can be also called the average case norm of INT_d, so that

$$e^{\mathrm{avg}}(0)^2 = \int_{F_d} (\mathrm{INT}_d f)^2 \, w_d(\mathrm{d}f) = \int_{\mathbb{R}^{2d}} K_d(x, y) \varrho_d(x) \varrho_d(y) \, \mathrm{d}x \, \mathrm{d}y.$$

We now elaborate on the initial average case error. Using the formula for the covariance kernel K_d we obtain

$$e^{\mathrm{avg}}(0)^2 = \int_{\mathbb{R}^d} \|x\|_2 \varrho_d(x)\,\mathrm{d}x - \frac{1}{2} \int_{\mathbb{R}^{2d}} \|x - y\|_2 \varrho_d(x)\varrho_d(y)\,\mathrm{d}x\,\mathrm{d}y.$$

For the last integral we use coordinate rotations by changing variables to

$$v = \frac{\sqrt{2}}{2}(x - y),$$

$$w = \frac{\sqrt{2}}{2}(x + y).$$

Clearly, $\|v\|_2^2 + \|w\|_2^2 = \|x\|_2^2 + \|y\|_2^2$ and $\varrho_d(v)\varrho_d(w) = \varrho_d(x)\varrho_d(y)$ and therefore

$$\int_{\mathbb{R}^{2d}} \|x - y\|_2 \varrho_d(x)\varrho_d(y)\,\mathrm{d}x\,\mathrm{d}y = \sqrt{2} \int_{\mathbb{R}^d} \|v\|_2 \varrho_d(v)\,\mathrm{d}v.$$

Hence,

$$e^{\mathrm{avg}}(0)^2 = \left(1 - \frac{\sqrt{2}}{2}\right) \int_{\mathbb{R}^d} \|x\|_2 \varrho_d(x)\,\mathrm{d}x.$$

Finally changing variables to $t = \|x\|_2$, see the book of Gradshtein and Ryzhik [66, 3.461 and 4.642], we obtain

$$e^{\mathrm{avg}}(0)^2 = \left(1 - \frac{\sqrt{2}}{2}\right) \frac{2\pi^{d/2}}{(2\pi)^{d/2}\Gamma(d/2)} \int_0^\infty t^d e^{-t^2/2}\,\mathrm{d}t,$$

which for even d is equal to

$$e^{\mathrm{avg}}(0)^2 = \sqrt{2\pi}\left(1 - \frac{\sqrt{2}}{2}\right) \frac{(d-1)!!}{2^{d/2}(-1 + d/2)!}$$

$$= \sqrt{2\pi}\left(1 - \frac{\sqrt{2}}{2}\right) \frac{(d-1)!}{2^{d-1}[(-1 + d/2)!]^2},$$

and for odd d is equal to

$$e^{\mathrm{avg}}(0)^2 = \sqrt{2}\left(1 - \frac{\sqrt{2}}{2}\right) \frac{[(d-1)/2]!}{\Gamma(d/2)} = \frac{2^{d/2}}{\pi^{1/2}}\left(1 - \frac{\sqrt{2}}{2}\right) \frac{[(d-1)/2]!}{(d-2)!!},$$

with $(-1)!! = 1$.[6] For odd d with $d \geq 3$ we have

$$e^{\mathrm{avg}}(0)^2 = \frac{2^{d-3/2}}{\pi^{1/2}}\left(1 - \frac{\sqrt{2}}{2}\right) \frac{[(d-1)/2]![(d-3)/2]!}{(d-2)!}.$$

---

[6] Here $n!!$ denotes the product of all even (if n is even) or odd (if n is odd) integers from $[1, n]$. That is, $(2k)!! = 2 \cdot 4 \cdots (2k)$ and $(2k+1)!! = 1 \cdot 3 \cdots (2k+1)$.

Using Stirling's formula, $n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}$, it is possible to check that for large d we have

$$e^{\text{avg}}(0) = \sqrt{1 - \frac{\sqrt{2}}{2}d^{1/4}}\,(1 + o(1)) \quad \text{as } d \to \infty.$$

It is also easy to see that $e^{\text{avg}}(0)$ is an increasing function of d. This follows from the fact that

$$\int_{\mathbb{R}^{d+1}} \sqrt{x_1^2 + x_2^2 + \cdots + x_{d+1}^2}\,\varrho_{d+1}(x)\,\mathrm{d}x$$

$$\geq \int_{\mathbb{R}^{d+1}} \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}\,\varrho_{d+1}(x)\,\mathrm{d}x$$

$$= \int_{\mathbb{R}^d} \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}\,\varrho_d(x)\,\mathrm{d}x,$$

i.e., the integral $\int_{\mathbb{R}^d} \|x\|_2 \varrho_d(x)\,\mathrm{d}x$ increases with d. Furthermore, since

$$\int_{\mathbb{R}^d} \sqrt{x_1^2 + \cdots + x_d^2}\,\varrho_d(x)\,\mathrm{d}x \leq \left(\int_{\mathbb{R}^d} \left(x_1^2 + \cdots + x_d^2\right)\varrho_d(x)\,\mathrm{d}x\right)^{1/2} = \sqrt{d},$$

we conclude that

$$e^{\text{avg}}(0) \leq \sqrt{1 - \frac{\sqrt{2}}{2}d^{1/4}} \quad \text{for all } d \in \mathbb{N}.$$

The asymptotic formula for $e^{\text{avg}}(0)$ tells us that the last estimate is sharp for large d.

We choose the normalized error criterion and define

$$n^{\text{avg-nor}}(\varepsilon, \text{INT}_d) = \min \left\{ n \mid \text{there exists } A_n \text{ with } e^{\text{avg}}(A_n) \leq \varepsilon e^{\text{avg}}(0) \right\}$$

as the minimal number of function values needed to solve the Gaussian integration problem in the average case setting to within $\varepsilon e^{\text{avg}}(0)$.

As we have already said, tractability in the average case setting is defined similarly to the worst case setting. In particular, *polynomial* tractability means that there exist three non-negative numbers C, p and q such that

$$n^{\text{avg-nor}}(\varepsilon, \text{INT}_d) \leq C\varepsilon^{-p}d^q \quad \text{for all } \varepsilon \in (0,1) \text{ and } d = 1, 2, \ldots.$$

If $q = 0$ then we have *strong* polynomial tractability.

Surprisingly, it is easy to establish *strong polynomial tractability* of Gaussian integration in the average case setting. The proof is short and based on the averaging argument that is often used in tractability studies. The downside of this approach is that the proof is *not* fully constructive and usually gives too large an exponent of ε^{-1}. However, in our case, the exponent of ε^{-1} is sharp.

To prove strong polynomial tractability, let A_n be a quasi-Monte Carlo algorithm using sample points x_j. Hence $a_1 = a_2 = \cdots = a_n = 1/n$. To stress the role of x_j,

let us denote $A_n = A_n(x_1, x_2, \ldots, x_n)$. We know the formula for its average case error, which obviously depends on the sample points x_j. We now treat these points x_j as independent identically distributed points over $\mathbb{R}^d$ with the standard Gaussian distribution. We compute the average value of the squares of the average case error of $A_n(x_1, x_2, \ldots, x_n)$ with respect to such x_j, defined as

$$\left[e_n^{\mathrm{avg}}\right]^2 := \int_{\mathbb{R}^{dn}} \left[e^{\mathrm{avg}}\left(A_n(x_1, \ldots, x_n)\right)\right]^2 \varrho_d(x_1)\varrho_d(x_2)\ldots\varrho_d(x_n)\,\mathrm{d}x_1 \ldots \,\mathrm{d}x_n.$$

So we need to integrate

$$e^{\mathrm{avg}}(0)^2 - \frac{2}{n}\sum_{k=1}^n \int_{\mathbb{R}^d} K_d(x_k, x)\varrho_d(x)\,\mathrm{d}x + \frac{1}{n^2}\sum_{i,k=1}^n K_d(x_i, x_k).$$

The integration of $\int_{\mathbb{R}^d} K_d(x_k, x)\varrho_d(x)\,\mathrm{d}x$ with respect to the points x_j will give us $e^{\mathrm{avg}}(0)^2$, whereas the integration of $K_d(x_i, x_k)$ depends on whether i and k are equal or different. If they are equal, then we obtain

$$\int_{\mathbb{R}^d} K_d(x, x)\varrho_d(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \|x\|_2\varrho_d(x)\,\mathrm{d}x = \left(2 + \sqrt{2}\right)e^{\mathrm{avg}}(0)^2,$$

and if they are different we obtain $e^{\mathrm{avg}}(0)^2$. This yields

$$\left[e_n^{\mathrm{avg}}\right]^2 = \left(1 - 2 + \frac{(2+\sqrt{2})n}{n^2} + \frac{n^2 - n}{n^2}\right)e^{\mathrm{avg}}(0)^2 = \frac{1 + \sqrt{2}}{n}e^{\mathrm{avg}}(0)^2.$$

Hence,

$$e_n^{\mathrm{avg}} = \frac{\sqrt{1 + \sqrt{2}}}{n^{1/2}}e^{\mathrm{avg}}(0).$$

We now apply the mean value theorem. If the average of $[e^{\mathrm{avg}}(A_n(x_1, x_2, \ldots, x_n))]^2$ is $\left[e_n^{\mathrm{avg}}\right]^2$ then there must exist points $x_1^*, x_2^*, \ldots, x_n^*$ for which

$$e^{\mathrm{avg}}(A_n(x_1^*, x_2^*, \ldots, x_n^*)) \le e_n^{\mathrm{avg}} = \frac{\sqrt{1 + \sqrt{2}}}{n^{1/2}}e^{\mathrm{avg}}(0).$$

This means that we can reduce the initial average case error by a factor ε by using a quasi-Monte Carlo algorithm with n chosen such that $(1 + \sqrt{2})/n \le \varepsilon^2$. Hence we can choose

$$n = \left\lceil \frac{1 + \sqrt{2}}{\varepsilon^2} \right\rceil,$$

which does not depend on d and depends quadratically on ε^{-1}. This implies *strong polynomial tractability* of Gaussian integration, as well as the estimate

$$n^{\mathrm{avg\text{-}nor}}(\varepsilon, \mathrm{INT}_d) \le \left\lceil \frac{1 + \sqrt{2}}{\varepsilon^2} \right\rceil, \tag{3.12}$$

see [94] where this result was originally shown.

The exponent of ε^{-1} in the last estimate is 2. We can hope to improve it since the exponent 2 measures the behavior of $A_n(x_1, x_2, \ldots, x_n)$ for average points $x_1, x_2, \ldots, x_n$, whereas the best exponent p cannot be worse than the behavior of $A_n(x_1, x_2, \ldots, x_n)$ for points $x_1, x_2, \ldots, x_n$ minimizing its average case error. This is, however, *not* the case as shown by Wasilkowski [263][7]. Hence, the bound above is sharp with respect to the exponent of ε^{-1}.

Although this is a good result, the choice of x_j^* is *not* constructive. We can do a little better in terms of constructivity by using Chebyshev's inequality. More precisely, let us take a number $C > 1$ and form the set

$$A_C = \left\{ [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{dn} \mid e^{\mathrm{avg}}(A_n(x_1, x_2, \ldots, x_n)) \leq C e_n^{\mathrm{avg}} \right\}$$

of points with the average case error bounded by $C e_n^{\mathrm{avg}}$. Then Chebyshev's inequality states that

$$\lambda_{dn}(A_C) \geq 1 - C^{-2},$$

where λ_{dn} is the standard Gaussian measure over $\mathbb{R}^{dn}$. Hence if we take, say, $C = 10$ then we can select good points x_j with probability at least 0.99. This leads us to a *semi-construction* of sample points x_j. We select the points randomly from $\mathbb{R}^{dn}$, compute the average case error of $A_n(x_1, x_2, \ldots, x_n)$ and accept them if

$$e^{\mathrm{avg}}(A_n(x_1, x_2, \ldots, x_n)) \leq C e_n^{\mathrm{avg}}.$$

If not, then we select the points x_j again. As long as we select them independently and identically distributed according to the standard Gaussian measure, see e.g., Box and Muller [18] how it can be done, we succeed after a few trials, since the probability of failure of all k trials is C^{-2k}; for $C = 10$ this probability is 10^{-2k}. In this way we can construct points and solve the problem as long as

$$n = \left\lceil \frac{C^2(1 + \sqrt{2})}{\varepsilon^2} \right\rceil.$$

We now comment on the absolute error criterion. Since

$$\varepsilon e^{\mathrm{avg}}(0) \leq \varepsilon \sqrt{1 - \frac{\sqrt{2}}{2}} d^{1/4}$$

and this estimate is sharp for large d, we conclude that for the absolute error criterion we have

$$n^{\mathrm{avg\text{-}abs}}(\varepsilon, \mathrm{INT}_d) \leq \left\lceil \frac{\sqrt{2d}}{2\varepsilon^2} \right\rceil,$$

---

[7]Wasilkowski [263] studied multivariate integration over the unit ball with weight equal to 1 and proved that the optimal exponent for the d-variate case is $2/(1 + 1/d)$. Since our integration problem is not easier we conclude that the optimal exponent must be at least $2/(1 + 1/d)$. Since this holds for all d, we see that the optimal exponent is indeed 2.

and that the exponents of ε^{-1} and d are sharp. Hence, we lost strong polynomial tractability but we still have *polynomial tractability* with the (small) exponent $\frac{1}{2}$ with respect to d.

We end this example by a remark that will be fully explained and used in the course of this book. There is a useful relationship between the average case and worst case errors for approximating continuous linear functionals. We explain this relation for Gaussian integration and for the isotropic Wiener measure. Namely, having the covariance kernel K_d we can take the Hilbert space $H(K_d)$ that has the same reproducing kernel K_d as the Gaussian measure.

Reproducing kernel Hilbert spaces play a major role in tractability studies. In Appendix A, the reader may find many useful properties of such spaces, which will be used throughout this book. See also Appendix B as well as Aronszajn [2] and the book of Berlinet and Thomas-Agnan [10]. Here, we only mention that the space $H(K_d)$ is the completion of the linear space of functions $f = \sum_{k=1}^{m} \alpha_k K_d(\cdot, t_k)$ for any choice of integer m, real coefficients α_k and points $t_k \in \mathbb{R}^d$. If g is of the same form as f with coefficients β_k then the inner product in $H(K_d)$ of f and g is

$$\langle f, g \rangle_{H(K_d)} = \sum_{k,j=1}^{m} \alpha_k \beta_j K_d(t_k, t_j).$$

Clearly, $H(K_d)$ is a subset of F_d. In fact, it is known that $w_d(H(K_d)) = 0$.

The space $H(K_d)$ was characterized by Molchan [150] for odd d, and later by Ciesielski [30] for arbitrary d. In particular, it was shown that the inner product of f and g that have finite support, vanish at 0, and are infinitely differentiable is given by

$$\langle f, g \rangle_{H(K_d)} = a_d \langle (-\Delta)^{(d+1)/4} f, (-\Delta)^{(d+1)/4} g \rangle_{L_2(\mathbb{R}^d)},$$

with known numbers a_d. Here, Δ is the Laplace operator, and for $d + 1$ not divisible by 4, the operator $(-\Delta)^{(d+1)/4}$ is understood in the generalized sense, see the book of Stein [224].

For the reproducing kernel Hilbert space $H(K_d)$ we have $K_d(\cdot, x) \in H(K_d)$ and

$$f(x) = \langle f, K_d(\cdot, x) \rangle_{H(K_d)} \quad \text{for all } f \in H(K_d) \text{ and all } x \in \mathbb{R}^d.$$

This property allows us to compute the worst case error of any linear algorithm. Indeed, consider Gaussian integration for functions from the unit ball of $H(K_d)$ in the *worst case* setting. Then

$$\text{INT}_d f = \int_{\mathbb{R}^d} \langle f, K_d(\cdot, x) \rangle_{H(K_d)} \, \varrho_d(x) \, \mathrm{d}x = \langle f, h_d \rangle_{H(K_d)},$$

where

$$h_d(t) = \int_{\mathbb{R}^d} K_d(t, x) \varrho_d(x) \, \mathrm{d}x.$$

Take a linear algorithm $A_n(f) = \sum_{j=1}^n a_j f(x_j)$. Then

$$\text{INT}_d f - A_n(f) = \left\langle f, h_d - \sum_{j=1}^n a_j K_d(\cdot, x_j) \right\rangle_{H(K_d)},$$

and the worst case error of A_n is

$$e^{\text{wor}}(A_n) = \sup_{f \in H(K_d): \|f\|_{H(K_d)} \leq 1} |\text{INT}_d f - A_n(f)| = \left\| h_d - \sum_{j=1}^n a_j K_d(\cdot, x_j) \right\|_{H(K_d)}.$$

Due to the fact that $H(K_d)$ is a Hilbert space, we can compute the last norm to conclude that

$$e^{\text{wor}}(A_n)^2 = \|h_d\|^2_{H(K_d)} - 2 \sum_{j=1}^n a_j h_d(x_j) + \sum_{i,j=1}^n a_i a_j K_d(x_i, x_j).$$

Observe finally that

$$\|h_d\|^2_{H(K_d)} = \int_{\mathbb{R}^{2d}} K_d(x, y) \varrho_d(x) \varrho_d(y) \, dx \, dy.$$

Hence, the worst case error of A_n for the unit ball of $H(K_d)$ is *exactly the same* as the average case error of A_n for the space F_d equipped with the zero-mean Gaussian measure with the covariance function K_d.

   This duality between the average case setting for a zero-mean Gaussian measure with a covariance kernel K_d and the worst case setting for the unit ball of a reproducing kernel Hilbert space $H(K_d)$ holds for all continuous linear functionals. Hence, it is enough to analyze the problem in one setting and claim the results in the other. In our case, we can claim the results for Gaussian integration in the worst case setting for the unit ball of the reproducing kernel Hilbert space $H(K_d)$. Then we have strong polynomial tractability with the normalized error criterion and the exponent of ε^{-1} is 2, and polynomial tractability with the absolute error criterion and the exponent of ε^{-1} is 2, and the exponent of d is $\frac{1}{2}$. All these exponents are sharp.

3.2.3 Example 8: Folded Wiener Sheet Measure

Let r be a non-negative integer. In this section we consider the class

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

of functions that satisfy the boundary conditions and that are r times continuously differentiable with respect to all variables. Namely, we assume that all partial derivatives up to order r are 0 if one component of x is 0, i.e., $(D^{j_1,j_2,\ldots,j_d} f)(x) = 0$ for all $j_i = 0, 1, \ldots, r$ whenever $x_k = 0$ for some k. For example, $F_{d,0}$ is the class of

continuous functions f such that $f(x) = 0$ if some component of x is 0. The space $F_{d,r}$ is equipped with the sup norm, i.e.,

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

$F_{d,r}$ is a separable Banach space. We equip the space $F_{d,r}$ with the r-folded Wiener sheet measure, which is the classical Wiener sheet measure placed on partial derivatives of order r, i.e., for any Borel set B of $F_{d,r}$ we have

$$\mu_d(B) = w_d(D^{r,r,\dots,r}(B)),$$

where w_d is the Wiener sheet measure defined on Borel sets of the space $C([0,1]^d)$ of continuous functions. Recall that w_d is a zero-mean Gaussian measure whose covariance function is

$$K_d(x, y) = \prod_{j=1}^{d} \min(x_j, y_j).$$

The covariance kernel of the measure μ_d is

$$K_{d,r}(x, y) = \prod_{j=1}^{d} \int_0^1 \frac{(x_j - u)_+^r}{r!} \frac{(y_j - u)_+^r}{r!} \, du.$$

Note that for $r = 0$, we have $\mu_d = w_d$; functions from $F_{d,0}$ are distributed according to the Wiener sheet measure. For $r > 0$, the rth partial derivatives of functions from $F_{d,r}$ are distributed according to the Wiener sheet measure.

We consider multivariate approximation, $\mathrm{APP}_d \colon F_{d,r} \to L_2 := L_2([0,1]^d)$, defined as

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

and we consider algorithms using arbitrary linear functionals. This problem has been studied by Papageorgiou and Wasilkowski [188], and we now report their results. Let $\nu_d = w_d \mathrm{APP}_d^{-1}$ be the a priori measure on the target space $\mathrm{APP}_d(F_{d,r})$. Then ν_d is a zero-mean Gaussian measure whose covariance operator $C_{\nu_d} \colon L_2 \to L_2$, defined as

$$\langle C_{\nu_d} g_1, g_2 \rangle_{L_2} = \int_{L_2} \langle f, g_1 \rangle_{L_2} \langle f, g_2 \rangle_{L_2} \, \nu_d(\mathrm{d}f) \quad \text{for all } g_1, g_2 \in L_2,$$

is given by

$$C_{\nu_d} = T_{0,r} T_{1,r},$$

where $T_{0,r}, T_{1,r} \colon L_2 \to L_2$ are

$$(T_{0,r} f)(x) = \int_{[0,1]^d} \prod_{j=1}^{d} \frac{(x_j - t_j)_+^{r-1}}{(r-1)!} f(t) \, \mathrm{d}t,$$

$$(T_{1,r} f)(x) = \int_{[0,1]^d} \prod_{j=1}^{d} \frac{(t_j - x_j)_+^{r-1}}{(r-1)!} f(t) \, \mathrm{d}t.$$

The operator C_{ν_d} is self-adjoint, positive definite, and has a finite trace. Let us denote the eigenpairs of C_{ν_d} by $(\lambda_{d,j}, \eta_{d,j})$ with $C_{\nu_d}\eta_{d,j} = \lambda_{d,j}\eta_{d,j}$, where the eigenvalues $\lambda_{d,j}$ are ordered, $\lambda_{d,1} \geq \lambda_{d,2} \geq \cdots$, and the eigenfunctions $\eta_{d,j}$ are orthonormal, $\langle \eta_{d,i}, \eta_{d,j}\rangle_{L_2} = \delta(i,j)$. Then

$$\text{trace}(C_{\nu_d}) = \sum_{j=1}^{\infty} \lambda_{d,j} < \infty.$$

The eigenpairs of C_{ν_d} are the non-zero solutions of the differential equation

$$\lambda D^{2r+2,2r+2,\ldots,2r+2} z - (-1)^{d(r+1)} z = 0,$$

with the boundary conditions

$$\left.\frac{\partial^i z(x_1, x_2, \ldots, x_d)}{\partial x_j^i}\right|_{x_j=0} = 0, \quad i = 0, 1, \ldots, r,$$

$$\left.\frac{\partial^i z(x_1, x_2, \ldots, x_d)}{\partial x_j^i}\right|_{x_j=1} = 0, \quad i = r+1, r+2, \ldots, 2r+1$$

for $j = 1, 2, \ldots, d$. For the univariate case, $d = 1$, let us denote $\lambda_{1,j} = \lambda_j$ and $\eta_{1,j} = \eta_j$. It is known, see the book of Tikhomirov [237], that

$$\lambda_j = \left(\frac{1}{\pi j}\right)^{2(r+1)} (1 + o(1)) \quad \text{as } j \to \infty.$$

For $d \geq 2$, it is easy to see that the eigenpairs are products of the eigenpairs of the univariate case, i.e.,

$$\{\lambda_{d,j}\}_{j=1}^{\infty} = \{\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}\}_{j_i=1,2,\ldots} \quad \text{and} \quad \{\eta_{d,j}\}_{j=1}^{\infty} = \{\eta_{j_1}\eta_{j_2}\cdots\eta_{j_d}\}_{j_i=1,2,\ldots}.$$

Hence,

$$\text{trace}(C_{\nu_d}) = \left(\sum_{j=1}^{\infty} \lambda_j\right)^d < \infty.$$

On the other hand,

$$\sum_{j=1}^{\infty} \lambda_j = \int_{F_{1,r}} \|f\|_{L_2}^2 \mu_1(df) = \int_0^1 \left(\int_{F_{1,r}} f^2(x)\mu_1(df)\right) dx$$

$$= \int_0^1 K_{1,r}(x,x)\,dx = \int_0^1 \int_0^1 \frac{(x-u)_+^{2r}}{(r!)^2}\,du\,dx$$

$$= \int_0^1 \frac{x^{2r+1}}{(r!)^2(2r+1)}\,dx = \frac{1}{(r!)^2(2r+1)(2r+2)}.$$

This yields

$$\text{trace}(C_{\nu_d}) = \frac{1}{[(r!)^2(2r+1)(2r+2)]^d}.$$

The asymptotic behavior for $\lambda_{d,j}$ follows from Micchelli and Wahba [147], and we have

$$\lambda_{d,j} = \frac{1}{[(d-1)!]^{2(r+1)}\pi^{2d(r+1)}}\left(\frac{(\ln j)^{d-1}}{j}\right)^{2r+2}(1+o(1)) \quad \text{as } j \to \infty.$$

These facts are needed to analyze multivariate approximation for the folded Wiener sheet measure. We consider algorithms of the form

$$A_n(f) = \varphi_n(L_1(f), L_2(f), \ldots, L_{n(f)}(f))$$

for some adaptively chosen continuous linear functionals L_j and $n(f)$. Here we assume that the average value of $n(f)$ is n, i.e.,

$$n = \int_{F_d} n(f)w_d(\mathrm{d}f).$$

For simplicity, as in the previous section, we assume that A_n is measurable and that n is a non-negative integer. The average case error of A_n is now given as

$$e^{\text{avg}}(A_n) = \left(\int_{F_d} \|f - A_n(f)\|_{L_2}^2 w_d(\mathrm{d}f)\right)^{1/2}.$$

For $n = 0$, we take the zero algorithm $A_0 = 0$ and obtain the initial average case error

$$e^{\text{avg}}(0)^2 = \int_{F_d} \|f\|_{L_2}^2 \mu_d(\mathrm{d}f) = \int_{L_2} \|f\|_{L_2}^2 \nu_d(\mathrm{d}f)$$

$$= \text{trace}(C_{\nu_d}) = \frac{1}{[(r!)^2(2r+1)(2r+2)]^d}.$$

We analyze the normalized error. Let $n^{\text{avg-nor}}(\varepsilon, d) = n^{\text{avg-nor}}(\varepsilon, \text{APP}_d, F_{d,r})$ denote the minimal number of information operations needed to reduce the initial error by a factor ε, i.e.,

$$n^{\text{avg-nor}}(\varepsilon, d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\text{avg}}(A_n) \leq \varepsilon e^{\text{avg}}(0)\}.$$

From general results on the average case setting, see Wasilkowski [260] and Chapter 4, we know that

$$n^{\text{avg-nor}}(\varepsilon, d) = \min\left\{n \mid \sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j}\right\}.$$

From the behavior of the eigenvalues $\lambda_{d,j}$ we have

$$n^{\text{avg-nor}}(\varepsilon, d) = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{1/(r+1/2)}\left(\ln\frac{1}{\varepsilon}\right)^{(d-1)(r+1)/(r+1/2)}\right)$$

for fixed r and d. Hence, modulo a logarithmic factor, we have a polynomial dependence on ε^{-1}. Furthermore, the exponent of ε^{-1} is at most 2 and goes to 0 with increasing r.

From the asymptotic formula for the eigenvalues $\lambda_{d,j}$ we conclude that the asymptotic formula for the minimal number $n^{\mathrm{avg\text{-}nor}}(\varepsilon, d)$ is given by

$$n^{\mathrm{avg\text{-}nor}}(\varepsilon, d) = C_{d,r} \left(\frac{1}{\varepsilon}\right)^{1/(r+1/2)} \left(\ln \frac{1}{\varepsilon}\right)^{(d-1)(r+1)/(r+1/2)} (1 + o(1))$$

as $\varepsilon \to 0$, where the asymptotic constant is equal to

$$C_{d,r} = \left(\frac{1}{2(2r+1)}\right)^{(d-1)p_r} \frac{[(r!)^2(2r+1)(2r+2)]^{d/(2r+1)}}{[(d-1)!]^{p_r} \pi^{dp_r} (2r+1)^{1/(2r+1)}}$$

with $p_r = (r+1)/(r+\frac{1}{2})$. We stress that $C_{d,r}$ goes exponentially fast to 0 with increasing d due to the presence of $(d-1)!$ in the denominator.

So far, all looks promising. The minimal number $n^{\mathrm{avg\text{-}nor}}(\varepsilon, d)$ behaves nicely as ε goes to 0, the asymptotic constant is exponentially small in d, so it is natural to expect that we have polynomial tractability. But we *don't*. In fact, it is easy to show the opposite, namely that we have *intractability*, since $n^{\mathrm{avg\text{-}nor}}(\varepsilon, d)$ depends exponentially on d for a fixed ε. Indeed, suppose that we perform n information operations. The best we can do is to use the algorithm

$$A_n(f) = \sum_{j=1}^{n} \langle f, \eta_{d,j} \rangle_{L_2} \, \eta_{d,j},$$

for which the square of the average case error is

$$e^{\mathrm{avg}}(A_n)^2 = \sum_{j=n+1}^{\infty} \lambda_{d,j} \geq \sum_{j=1}^{\infty} \lambda_{d,j} - n\lambda_{d,1}$$
$$= \left(\sum_{j=1}^{\infty} \lambda_j\right)^d - n\lambda_1^d,$$

Hence, $e^{\mathrm{avg}}(A_n) \leq \varepsilon e^{\mathrm{avg}}(0)$ implies that

$$n^{\mathrm{avg\text{-}nor}}(\varepsilon, d) = n \geq \left(1 + \sum_{j=2}^{\infty} \frac{\lambda_j}{\lambda_1}\right)^d (1 - \varepsilon^2).$$

Since λ_2 is positive, $1 + \sum_{j=2}^{\infty} \lambda_j/\lambda_1 > 1$, and therefore n is exponential in d for all $\varepsilon \in (0, 1)$. This indeed means intractability. As we shall see in this book, this is an instance of a general result that intractability holds for linear tensor product problems in the average case setting.

Hence, as in many previous examples, good asymptotic behavior of $n^{\mathrm{avg\text{-}nor}}(\varepsilon, d)$ does not prevent exponential behavior in d. To get tractability, we must again introduce weights and reduce the role of at least some groups of variables.

3.2.4 Probabilistic Setting

In the probabilistic setting, the error of an algorithm is defined as in the worst case setting, but disregarding a set of preassigned measure $\delta \in [0, 1)$. In this introductory chapter, we only wish to show the role of the parameter δ for different error criteria and how δ affects tractability results. To simplify technical considerations, we limit ourselves to Gaussian integration, discussed in Section 3.2.2. We first consider the absolute and normalized error criteria. Then we turn, for the first time in this book, to the relative error criterion, and explain why we have waited to analyze this important error criterion until the probabilistic setting.

3.2.5 Example 9: Absolute and Normalized Errors

Recall the definition of Gaussian integration INT_d for the class F_d and the isotropic Wiener measure w_d from Section 3.2.2. For $\delta \in [0, 1)$, the probabilistic error of an algorithm A_n using n function values is defined as

$$e^{\text{prob}}(A_n; \delta) = \inf_{B : w_d(B) \leq \delta} \sup_{f \in F_d - B} |\text{INT}_d f - A_n(f)|.$$

For the absolute error criterion, we want to find the minimal n for which we solve the problem to within ε. That is, this minimal number is now equal to

$$n^{\text{prob-abs}}(\varepsilon, \delta, \text{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\text{prob}}(A_n; \delta) \leq \varepsilon\}.$$

Note that in the probabilistic setting, the minimal number of function values also depends on δ and, as we shall see in a moment, goes to infinity as δ goes to 0.

The probabilistic setting is closely related to the average case setting, see Wasilkowski [261] and Chapter 8 in [242], as well as Chapter 4, where the relation between the two settings will be explained. In particular, this relation is especially pleasing for problems specified by continuous linear functionals, such as our introductory example of Gaussian integration. (In fact, that was the reason for choosing such an example.) Namely, let

$$n^{\text{avg-abs}}(\varepsilon, \text{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\text{avg}}(A_n) \leq \varepsilon\}$$

be the corresponding minimal number of function values needed to solve the same problem to within ε with the absolute error criterion in the average case setting. Finally, let

$$\psi(z) = \sqrt{\frac{2}{\pi}} \int_0^z \exp\left(-t^2/2\right) \, \mathrm{d}t \quad \text{for } z \in [0, \infty)$$

be the probability integral. Then it is shown in Chapter 8 of [242] that

$$n^{\text{prob-abs}}(\varepsilon, \delta, \text{INT}_d) = n^{\text{avg-abs}}(\varepsilon/\psi^{-1}(1 - \delta), \text{INT}_d).$$

Thus, the minimal number of function values in the probabilistic setting is the same as the minimal number of function values in the average case setting if we replace ε by $\varepsilon/\psi^{-1}(1-\delta)$. Note that $\psi(z) = 1 - \sqrt{2/\pi}\, z^{-1} \exp(-z^2/2)(1 + o(1))$ as z goes to infinity. This implies that

$$\psi^{-1}(1-\delta) = \sqrt{2\ln\delta^{-1}}\,(1 + o(1)) \quad \text{as } \delta \to 0.$$

Hence, for small δ, we have

$$\varepsilon/\psi^{-1}(1-\delta) \approx \varepsilon/\sqrt{2\ln\delta^{-1}}.$$

The average case setting for the normalized error criterion was studied in Section 3.2.2. This obviously corresponds to the absolute error criterion with ε replaced by $\varepsilon e^{\mathrm{avg}}(0)$. Using the estimate (3.12) we therefore conclude that

$$n^{\mathrm{avg\text{-}abs}}(\varepsilon/\psi^{-1}(1-\delta), \mathrm{INT}_d) \leq \left\lceil \frac{(1+\sqrt{2})\left[\psi^{-1}(1-\delta)\right]^2}{\varepsilon^2}\left[e^{\mathrm{avg}}(0)\right]^2 \right\rceil.$$

We know that $e^{\mathrm{avg}}(0)$ is of order $d^{1/4}$. Therefore

$$n^{\mathrm{prob\text{-}abs}}(\varepsilon, \delta, \mathrm{INT}_d) = \mathcal{O}\!\left(\frac{d^{1/2}\ln\delta^{-1}}{\varepsilon^2}\right)$$

with the factor in the big $\mathcal{O}$ notation independent of ε, δ and d. Observe that all exponents of ε^{-1}, $\ln\delta^{-1}$ and d are sharp. We also stress the weak logarithmic dependence on δ.

We are ready to address tractability of Gaussian integration in the probabilistic setting. Suppose we adopt the following definition of tractability, which is quite conservative as far as the dependence on δ^{-1} is concerned. Namely, we say that the problem is *tractable in the probabilistic setting* iff there exist non-negative numbers C, p, q and s such that

$$n^{\mathrm{prob\text{-}abs}}(\varepsilon, \delta, \mathrm{INT}_d) \leq C\varepsilon^{-p}d^q\left(\ln\delta^{-1}\right)^s \quad \text{for all } \varepsilon, \delta \in (0, 1) \text{ and all } d \in \mathbb{N},$$
(3.13)

and *strongly tractable* with respect to d if $q = 0$, and *strongly tractable* with respect to δ if $s = 0$.

Then Gaussian integration is *tractable* in the probabilistic setting for the absolute error with $p = 2$, $q = \frac{1}{2}$ and $s = 1$, and all exponents are sharp.

This illustrates how positive tractability results from the average case setting can be easily transferred to positive tractability results in the probabilistic setting for continuous linear functionals.

We now turn to the normalized error criterion. We need to consider the initial error in the probabilistic setting, i.e., the error of the zero algorithm. It is known, again see Chapter 8 of [242], that

$$e^{\mathrm{prob}}(0; \delta) = \psi^{-1}(1-\delta)e^{\mathrm{avg}}(0) \approx \sqrt{2\ln\delta^{-1}}\,e^{\mathrm{avg}}(0).$$

Note that the initial error now depends on δ, and it goes (slowly) to infinity as δ approaches 0. Let

$$n^{\mathrm{prob\text{-}nor}}(\varepsilon, \delta, \mathrm{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{prob}}(A_n; \delta) \le \varepsilon e^{\mathrm{prob}}(0; \delta)\}$$

be the minimal number of function values needed to reduce the initial error by a factor ε. Observe that for decreasing δ, the probabilistic error of A_n increases but the error bound $\varepsilon e^{\mathrm{prob}}(0; \delta)$ also increases. So there is a trade-off, and it is not yet clear which of these two behaviors is more important. In fact, they cancel and there is *no* dependence on δ. Indeed,

$$n^{\mathrm{prob\text{-}nor}}(\varepsilon, \delta, \mathrm{INT}_d) = n^{\mathrm{prob\text{-}abs}}(\varepsilon e^{\mathrm{prob}}(0; \delta), \delta, \mathrm{INT}_d),$$

and the formula for the initial error and the estimate for the absolute error yields

$$n^{\mathrm{prob\text{-}nor}}(\varepsilon, \delta, \mathrm{INT}_d) = n^{\mathrm{avg\text{-}nor}}(\varepsilon, \mathrm{INT}_d) \le \left\lceil \frac{\sqrt{2d}}{2\varepsilon^2} \right\rceil.$$

Hence, δ does *not* play any role for the normalized error in the probabilistic setting for continuous linear functionals. We now have strong tractability with respect to δ.

3.2.6 Example 10: Relative Error

For the first time in our book we now consider the *relative* error criterion. That is, for Gaussian integration and an algorithm A_n that uses n function values, we define the relative errors

$$\frac{|\mathrm{INT}_d f - A_n(f)|}{|\mathrm{INT}_d f|} \quad \text{for all } f \in F_d$$

with the convention that $0/0 = 0$.

We first consider the worst case error,

$$e^{\mathrm{wor-rel}}(A_n) = \sup_{f \in F_d} \frac{|\mathrm{INT}_d f - A_n(f)|}{|\mathrm{INT}_d f|}.$$

Obviously, the initial error, which is the worst case error for $A_n = 0$, is now 1. Unfortunately, there is no way to reduce the initial error, no matter how large n is and how sophisticated A_n is chosen, see [242, p. 105]. Indeed, take a function f with n zero values at the points used by A_n. What can we say about f? Since $f = 0$ is one such function, we must take $A_n(f) = \varphi(0) = 0$, since otherwise the worst case error will be infinite. But f can be also a non-zero function with a non-zero integral $\mathrm{INT}_d f$. In this case, since our algorithm cannot distinguish this function from the zero function, the relative error is 1. Hence, there is no way to solve the problem with the relative error criterion in the worst case setting. We showed this negative property for Gaussian integration but it is clear that this property holds for all problems that cannot be recovered exactly by the use of finitely many function values.

Let us then switch to the average case setting, and define the average case error as

$$e^{\text{avg−rel}}(A_n) = \int_{F_d} \frac{|\text{INT}_d f - A_n(f)|}{|\text{INT}_d f|} w_d(\mathrm{d}f).$$

Again, the initial error is 1, and again we *cannot* reduce it no matter how n and A_n are chosen, see [242, p. 268]. However, the proof is not as obvious as in the worst case setting. Moreover, unlike the worst case setting, the result is *not* true for general problems; our problem must be a continuous linear functional. We now sketch the proof for $A_n(f) = \varphi(N(f))$ with $N(f) = [f(x_1), f(x_2), \ldots, f(x_n)]$ for some adaptive points x_j's. Knowing $y = N(f)$ we decompose the Gaussian measure w_d such that

$$w_d(B) = \int_{N(F_d)} \mu_{d,2}(B|y)\mu_{d,1}(\mathrm{d}y) \quad \text{for any Borel set } B \text{ of } F_d$$

with $\mu_{d,1} = w_d N^{-1}$ and $\mu_{d,2}(\cdot \mid y)$ being the conditional measure that is concentrated on functions sharing the information y, i.e., $\mu_{d,2}(N^{-1}(y) \mid y) = 1$ for almost all y. It is known that both $\mu_{d,1}$ and $\mu_{d,2}(\cdot \mid y)$ are Gaussian measures. Then we define $\nu_d(\cdot \mid y) = \mu_{d,2}(\text{INT}_d^{-1}\cdot \mid y)$, which is a univariate Gaussian measure with a non-zero variance. This decomposition allows us to present the average case error of A_n as

$$e^{\text{avg−rel}}(A_n) = \int_{N(F_d)} \left(\int_{\mathbb{R}} \frac{|x - \varphi(y)|}{|x|} \nu_d(\mathrm{d}x|y) \right) \mu_{d,1}(\mathrm{d}y).$$

Now consider the inner integral. This integral is 1, if $\varphi(y) = 0$, and it is infinite if $\varphi(y) \neq 0$. Hence, the average case error of A_n is at least one, as claimed.

Thus, there is no way to solve the problem with the relative error in the average case setting. This holds for Gaussian integration, as well for all continuous linear functionals that cannot be recovered exactly by the use of finitely many function values.

This explains why we did not consider the relative error in the worst and average case settings so far. We now switch to the probabilistic setting and show that Gaussian integration *can* be solved in this setting. The probabilistic error is now given by

$$e^{\text{prob-rel}}(A_n; \delta) = \inf_{B: w_d(B) \leq \delta} \sup_{f \in F_d - B} \frac{|\text{INT}_d f - A_n(f)|}{|\text{INT}_d|}.$$

Positive results in the probabilistic setting are possible because we can disregard functions with small $|\text{INT}_d|$, which caused the trouble in the worst and average case settings.

The initial error is still 1. Let

$$n^{\text{prob-rel}}(\varepsilon, \delta, \text{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\text{prob-rel}}(A_n; \delta) \leq \varepsilon\}$$

be the minimal number of function values needed to solve the problem to within ε in the probabilistic setting with the relative error criterion. It was shown in [102], see also Section 6.1 of Chapter 6 in [242] as well as Volume II, that

$$n^{\text{prob-rel}}(\varepsilon, \delta, \text{INT}_d) \leq n^{\text{avg-nor}} \left(\frac{\varepsilon \tan(\delta\pi/2)}{\sqrt{1 + \varepsilon^2 \tan^2(\delta\pi/2)}} \right),$$

and the last inequality is sharp for small ε.

Using the estimate (3.12) we obtain

$$n^{\text{prob-rel}}(\varepsilon, \delta, \text{INT}_d) \leq \left\lceil \frac{(\sqrt{2d})(1 + \varepsilon^2 \tan^2(\delta\pi/2))}{2\varepsilon^2 \tan^2(\delta\pi/2)} \right\rceil = \mathcal{O}\left(\frac{d^{1/2}}{\varepsilon^2 \delta^2}\right),$$

with the factor in the big $\mathcal{O}$ notation independent of ε, δ and d. Furthermore, the exponents of ε^{-1}, δ^{-1} and d are sharp.

So now ε and δ play the same role and the minimal number of function values depends polynomially on both of them. Does it mean that we have tractability? No, it does not if we want to work with (3.13) which requires a poly-log dependence on δ^{-1}. However, if we relax this definition and switch to *polynomial* tractability with respect to all parameters then Gaussian integration will be tractable. More precisely, let us say that the problem is *polynomially tractable in the probabilistic setting* if

$$n^{\text{prob-rel}}(\varepsilon, \delta, \text{INT}_d) \leq C\varepsilon^{-p} d^q \delta^{-s} \quad \text{for all } \varepsilon, \delta \in (0, 1) \text{ and all } d \in \mathbb{N}. \quad (3.14)$$

Then Gaussian integration is indeed polynomially tractable with $p = s = 2$ and $q = \frac{1}{2}$.

3.2.7 Randomized Setting

In the worst case, average case and probabilistic settings, we consider *deterministic* algorithms. We now turn to the *randomized* setting, in which randomized algorithms are considered. Randomized algorithms have proved to be very efficient for many discrete and continuous problems. Probably one of the first randomized algorithm for a computational mathematical problem was the classical Monte Carlo algorithm of Metropolis and Ulam [145] for multivariate integration invented in the 1940's. Today the Monte Carlo algorithm and its many modifications are widely used in computational practice, especially in physics and chemistry. This algorithm is so popular that sometimes all randomized algorithms are also called Monte Carlo, since today the phrase "Monte Carlo" is a synonym for using randomization. We prefer that "Monte Carlo" only refers to the classical Monte Carlo algorithm for multivariate integration, referring to "randomized algorithms" for all other cases.

Obviously, tractability can also be studied in the randomized setting. We feel obliged first to address the question whether the Monte Carlo algorithm for multivariate integration has a tractability error bound. The reader probably knows that the rate of convergence of Monte Carlo does *not* depend on the number d of variables, but as we know, this is not enough for tractability. The randomized error of Monte Carlo depends also on the variance of a function; here we may have a good or bad dependence on d, which in turn depends on the class of functions. This point will be further explained in Section 3.2.8. In Section 3.2.9 we then present a multivariate problem for which randomization does *not* help and for which we have roughly the same results as in the worst case setting. This shows that the power of randomization very much depends on the problem we want to solve.

3.2.8 Example 11: Monte Carlo Algorithm

We introduce the randomized setting by studying multivariate integration

$$\text{INT}_d f = \int_{[0,1]^d} f(x)\,dx \quad \text{for } f \in F_d.$$

We assume that the class F_d consists of square integrable real functions, $F_d \subseteq L_2([0,1]^d)$.

We use randomized algorithms based on function values at randomly selected sample points. The general form of algorithms A_n is now

$$A_n(f, \omega) = \varphi_\omega \left(f(t_{1,\omega}), f(t_{2,\omega}), \ldots, f(t_{n(\omega),\omega}) \right).$$

Here, $\omega \in \Omega$ is a random element distributed according to some probability measure μ. The sample points $t_{j,\omega}$, their number $n(\omega)$, and the mapping φ_ω may depend on the random element ω; moreover adaptive choice of $t_{j,\omega}$ and $n(\omega)$ is allowed. In particular, this means that $n(\omega) = n(f, \omega)$ depends on f through its computed function values. Finally,

$$n = \sup_{f \in F_d} \int_\Omega n(f, \omega)\mu(d\omega)$$

denotes the average value of function values used by the algorithm A_n for a worst f.

For example, the classical Monte Carlo algorithm

$$\text{MC}_n(f, \omega) = \frac{1}{n} \sum_{j=1}^{n} f(\omega_j)$$

is of this form, with $\omega = [\omega_1, \omega_2, \ldots, \omega_n]$ for independent and uniformly distributed ω_j over $[0,1]^d$. That is, the measure $\mu = \lambda_{dn}$ is the Lebesgue measure on $[0,1]^{dn}$, the sample points $t_{j,\omega} = \omega_j$, the cardinality $n(\omega)$ is always equal to n, and the mapping φ_ω is deterministic in this case, being given by $\varphi_\omega(y_1, y_2, \ldots, y_n) = n^{-1} \sum_{j=1}^{n} y_j$.

For a randomized algorithm A_n we first select a random element ω, and we then proceed as before with this particular ω. That is, we compute function values at adaptively chosen points, adaptively decide how many function values we need and finally combine the computed function values to obtain the final approximation.

The randomized error of the algorithm A_n is defined as the average error with respect to randomization for a worst function from the class F_d, i.e.,

$$e^{\text{ran}}(A_n) = \sup_{f \in F_d} \left(\int_\Omega |\text{INT}_d f - A_n(f, \omega)|^2 \mu(d\omega) \right)^{1/2}.$$

The initial error in the randomized setting is the error of the zero algorithm, see Chapter 4 for a more general discussion. Hence we have

$$e^{\text{ran}}(0) = e^{\text{wor}}(0) = \|\text{INT}_d\|_{F_d \to \mathbb{R}}.$$

Having defined the randomized error, we can address tractability of multivariate integration in the randomized setting with the absolute and normalized error criteria. Namely, let

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{ran}}(A_n) \leq \varepsilon\, \mathrm{CRI}_d\},$$

with CRI_d indicated which *error criterion* is used. We have $\mathrm{CRI}_d = 1$ for the absolute error criterion, and $\mathrm{CRI}_d = e^{\mathrm{ran}}(0)$ for the normalized error criterion. Hence, $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ is the minimal number of function values needed to solve the problem to within ε in the randomized setting with a given error criterion.

Then tractability is defined as in other settings. In particular, *weak tractability* of multivariate integration in the randomized setting means that

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)}{\varepsilon^{-1} + d} = 0,$$

whereas *polynomial tractability* of multivariate integration in the randomized setting means that there exist three non-negative numbers C, p and q such that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \leq C \varepsilon^{-p} d^q \quad \text{for all } \varepsilon \in (0,1) \text{ and } d \in \mathbb{N}.$$

Strong polynomial tractability means that $q = 0$ in the estimate above.

Observe that the randomized setting is not harder than the worst case setting. Indeed, all algorithms that we want to use in the worst case setting may be also used in the randomized setting by taking all sample points $t_{j,\omega}$ and the mapping φ_ω independently of ω. This means that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \leq n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d).$$

Obviously, we hope that $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ is much smaller than $n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d)$; in particular, we hope that intractability in the worst case setting can be broken by switching to the randomized setting.

It is instructive and easy to illustrate the randomized error by considering the Monte Carlo algorithm and to rediscover the famous and very useful formula for its randomized error. We first need to compute

$$e^{\mathrm{ran}}(f)^2 = \int_{[0,1]^{dn}} \left(\mathrm{INT}_d f - \frac{1}{n} \sum_{j=1}^{n} f(\omega_j)\right)^2 \lambda_{dn}(\mathrm{d}\omega).$$

By squaring the expression for the integrand, and remembering that $\omega_1, \omega_2, \ldots, \omega_n$ are independent and uniformly distributed, as well as by performing integration over

dummy variables, we obtain

$$e^{\mathrm{ran}}(f)^2 = \mathrm{INT}_d^2 f - \frac{2}{n} \sum_{j=1}^{n} \int_{[0,1]^d} \mathrm{INT}_d f f(\omega_j) \, \mathrm{d}\omega_j$$

$$+ \frac{1}{n^2} \sum_{i,j=1, i \neq j}^{n} \int_{[0,1]^{2d}} f(\omega_i) f(\omega_j) \, \mathrm{d}\omega_i \, \mathrm{d}\omega_j$$

$$+ \frac{1}{n^2} \sum_{j=1}^{n} \int_{[0,1]^d} f^2(\omega_j) \, \mathrm{d}\omega_j.$$

This obviously simplifies to

$$e^{\mathrm{ran}}(f)^2 = \left(1 - 2 + \frac{n^2 - n}{n^2}\right) \mathrm{INT}_d^2 f + \frac{n}{n^2} \mathrm{INT}_d f^2.$$

In this way we rediscover the well-known formula

$$e^{\mathrm{ran}}(f) = \frac{1}{\sqrt{n}} \left[\mathrm{INT}_d f^2 - \mathrm{INT}_d^2(f)\right]^{1/2}.$$

The rate of convergence of the Monte Carlo algorithm, although not great, does *not* depend on d and its randomized error is proportional to $n^{-1/2}$. Does this imply tractability? It does not. The reason is that $n^{-1/2}$ is multiplied by the square root of the variance $\sigma^2(f)$ of the function f,

$$\sigma^2(f) = \mathrm{INT}_d f^2 - \mathrm{INT}_d^2 f,$$

which may depend on d in an arbitrary way. In any case, the randomized error of the Monte Carlo algorithm is equal to

$$e^{\mathrm{ran}}(\mathrm{MC}_n) = \frac{1}{\sqrt{n}} \sup_{f \in F_d} \sigma(f).$$

Hence, if the variances of functions from F_d divided by CRI_d^2 are polynomially dependent on d then the Monte Carlo algorithm has a tractability error bound. That is, if there are two numbers C and q such that

$$\sup_{f \in F_d} \sigma(f) \leq (Cd^q)^{1/2} \, \mathrm{CRI}_d \tag{3.15}$$

then setting

$$n = \left\lceil \frac{Cd^q}{\varepsilon^2} \right\rceil,$$

we obtain $e^{\mathrm{ran}}(\mathrm{MC}_n) \leq \varepsilon \mathrm{CRI}_d$. This implies

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) \leq \left\lceil \frac{Cd^q}{\varepsilon^2} \right\rceil,$$

and so multivariate integration is polynomially tractable in the randomized setting. Furthermore, if $q = 0$ we obtain strong polynomial tractability.

For simplicity, consider now the absolute error criterion, $\mathrm{CRI}_d = 1$. Then (3.15) holds if F_d is a subset of the ball in $L_2([0,1]^d)$ of radius Cd^q. This simply follows from the fact that

$$\sigma(f) \le \|f\|_{L_2([0,1]^d)}.$$

We now consider several spaces that we studied in the worst case setting previously in this chapter. In Section 3.1.1, we studied Lipschitz functions F_d^{lip} for which multivariate integration is intractable in the worst case setting. The problem is strongly polynomially tractable in the randomized setting, since $F_d = F_d^{\mathrm{lip}}$ is a subset of the unit ball of $L_2([0,1]^d)$. One can ask if the exponent 2 of ε^{-1} in the estimate on $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ can be improved. This problem was studied by Bakhvalov [7] who proved that the minimal randomized error of algorithms using n function values in the randomized setting is $\Theta(n^{-(1/2+1/d)})$ which means that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{2/(1+2/d)}\right)$$

but the factors in the Θ-notation may depend on d. Hence for $d = 1$, the optimal exponent of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_1)$ is 2/3 instead of the Monte Carlo exponent of 2. However, if we want the estimate of $n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d)$ to hold for all d, then the exponent 2 is the best possible. Similar results hold for smoother functions. For instance, if we take $F_d = C^r([0,1]^d)$ then we are still in the unit ball of $L_2([0,1]^d)$ and we find that multivariate integration is strongly polynomially tractable in the randomized setting, although polynomial tractability does not hold in the worst case setting. The minimal randomized error is of order $n^{-(r/d+1/2)}$, see again Bakhvalov [7] or Heinrich [80] or [160], [242], which implies that

$$n^{\mathrm{ran}}(\varepsilon, \mathrm{INT}_d) = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{2/(1+2r/d)}\right).$$

Again for large d, the exponent of ε^{-1} is close to 2, which means that the Monte Carlo algorithm not only gives strong polynomial tractability but also minimizes the exponent of ε^{-1}.

If (3.15) does *not* hold, then all bets are off for the Monte Carlo algorithm. We illustrate this by an (admittedly artificial) example, for which the randomized error of Monte Carlo is infinite, even though the problem itself is trivial. Indeed, let

$$F_d = \{f : [0,1]^d \to \mathbb{R} \mid f(x) = cx_1 \text{ for some } c \in \mathbb{R}\}.$$

Clearly, for $f(x) = cx_1$, we have $\sigma^2(f) = c^2/12$ and since c can be arbitrarily large, (3.15) does not hold and the randomized error of the Monte Carlo algorithm is indeed infinite. On the other hand,

$$\int_{[0,1]^d} f(x)\,\mathrm{d}x = f\left(\tfrac{1}{2}, 0, 0, \ldots, 0\right) \quad \text{for all } f \in F_d.$$

This means that the integration problem can be exactly solved even in the worst case setting using just one function value.

What is the lesson from this artificial example? We learn that the a priori information given by the fact that $f \in F_d$ may be very powerful. The Monte Carlo algorithm may be very bad for some classes F_d because it is *not* using the information that integrands enjoy properties hidden in the definition of F_d.

We now turn to the normalized error. As in [219], we take F_d as the unit ball of the weighted Sobolev $H(K_d)$ space, which is the reproducing kernel Hilbert space with the kernel

$$K_d(x, y) = \prod_{j=1}^{d} \left(1 + \gamma_j \min(x_j, y_j)\right)$$

for some sequence of weights γ_j with $\gamma_1 \geq \gamma_2 \geq \cdots \geq 0$, see also Appendix A 2.2.

It is known and easy to verify that the initial error is given by

$$e^{\mathrm{ran}}(0) = \left(\int_{[0,1]^{2d}} K_d(x, y) \, dx \, dy\right)^{1/2} = \prod_{j=1}^{d} \left(1 + \tfrac{1}{3}\gamma_j\right)^{1/2}.$$

Observe that for the unweighted case $\gamma_j = 1$, we have $e^{\mathrm{ran}}(0) = (4/3)^{d/2}$, which is exponentially large in d. Furthermore, it is easy to check that $e^{\mathrm{ran}}(0)$ are uniformly bounded in d iff $\sum_{j=1}^{\infty} \gamma_j < \infty$.

The assumption (3.15) does not hold for the unit ball of $H(K_d)$ for general weights. One can ask what are necessary and sufficient conditions on the weights $\{\gamma_j\}$ to guarantee strong polynomial tractability or polynomial tractability error bounds for the Monte Carlo algorithm using the normalized error criterion. This problem was solved in [219]. Namely,

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

is a necessary and sufficient condition for strong polynomial tractability of the Monte Carlo algorithm, and

$$\limsup_{d \to \infty} \frac{\sum_{j=1}^{d} \gamma_j^2}{\ln d} < \infty$$

is a necessary and sufficient condition for polynomial tractability of the Monte Carlo algorithm.

For example, this means that for the unweighted case $\gamma_j = 1$, the Monte Carlo algorithm requires exponentially many randomized function values in d to reduce the initial error by a factor of ε, whereas for decaying weights such that $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$, this number is independent of d and is of order ε^{-2}.

Hence, $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$ implies that multivariate integration is strongly polynomially tractable for the unit ball of $H(K_d)$ in the randomized setting for the normalized error criterion. It should be added that for the same problem in the worst case setting, we obtain strong polynomial tractability iff $\sum_{j=1}^{\infty} \gamma_j < \infty$, see [178]. Hence, the randomized setting relaxes the conditions on the weights, see also Open Problem 13.

3.2.9 Example 12: Class Λ^{all}

In the previous subsection we indicated that randomization can be very powerful for multivariate integration. In particular, we showed that randomization can break intractability of multivariate integration in the worst case setting for the class Λ^{std} of function values.

In this subsection we consider the class Λ^{all} of all linear functionals and consider linear operators $S \colon F \to G$, where F is the unit ball of a Hilbert space, and the target space G is also a Hilbert space. Without loss of generality we also assume that S is compact, since otherwise the problem cannot be solved to within ε for small ε in the worst case and randomized settings.

Randomized algorithms are defined analogously to the previous subsection. Since we can use linear functionals, the general form of A_n is now

$$A_n(f, \omega) = \varphi_\omega \left(L_{1,\omega}(f), L_{2,\omega}(f, y_1), \dots, L_{n(\omega),\omega}(f, y_1, y_2, \dots, y_{n(\omega)-1}) \right),$$

where $\omega \in \Omega$ is a random element distributed according to some probability measure μ, $y_i = L_{i,\omega}(f, y_1, y_2, \dots, y_{i-1})$, and $L_{i,\omega}(\cdot, y_1, y_2, \dots, y_{i-1}) \in \Lambda^{\mathrm{all}}$. The mapping φ_ω now goes to the target space G, and n is the average value of information operations this time from Λ^{all} for a worst f, i.e.

$$n = \sup_{f \in F} \int_\Omega n(f, \omega) \mu(\mathrm{d}\omega),$$

where $n(\omega) = n(f, \omega)$ may depend adaptively on f through its information operations values.

The randomized error of A_n now takes the form

$$e^{\mathrm{ran}}(A_n) = \sup_{f \in F} \left(\int_\Omega \| Sf - A_n(f, \omega) \|_G^2 \mu(\mathrm{d}\omega) \right)^{1/2}.$$

Finally, define the minimal randomized error in the class $\mathcal{A}_n$ of all randomized algorithms using n information operations from Λ^{all} on the average as

$$e_n^{\mathrm{ran}}(S) = \inf_{A_n \in \mathcal{A}_n} e^{\mathrm{ran}}(A_n).$$

We stress that the class $\mathcal{A}_n$ contains all randomized algorithms that we can get by varying the randomized adaptive choices of linear functionals, mappings φ_ω, and distributions μ.

We compare the numbers $e_n^{\mathrm{ran}}(S)$ to the minimal errors $e_n^{\mathrm{wor}}(S)$ that we can achieve in the worst case setting. Formally, $e_n^{\mathrm{wor}}(S)$ is defined as above if we assume that all algorithms $A_n(f, \omega)$ do *not* depend on ω.

One of the most important and difficult research problems of information-based complexity is to study the power of randomization and compare it to the power of deterministic algorithms. We report on this line of research in the course of this book.

It turns out that as long as we deal with compact operators between Hilbert spaces and use the class Λ^{all}, randomization does *not* really help. More precisely, modulo the measurability assumption of algorithms in the randomized setting, we have the inequality

$$\tfrac{1}{2} e_{4n-1}^{\mathrm{wor}}(S) \le e_n^{\mathrm{ran}}(S) \le e_n^{\mathrm{wor}}(S), \tag{3.16}$$

which will be proved in Chapter 4. Results similar to (3.16) have been already obtained in Mathé [141], Heinrich [79], and Wasilkowski [262], as well as in [161].

Obviously, the second inequality is trivial since the class of randomized algorithms is simply larger than the class of deterministic algorithms in the worst case setting. The essence of (3.16) is the first inequality. It states that the minimal randomized errors of algorithms with n information operations from Λ^{all} on the average must be at least as large as a half of the minimal worst case errors of algorithms using at most $4n - 1$ information operations. Since the constants usually do not play a significant role, (3.16) says that randomization does not really help.

Indeed, let us consider $n^{\mathrm{ran}}(\varepsilon, S, \Lambda^{\mathrm{all}})$ and $n^{\mathrm{wor}}(\varepsilon, S, \Lambda^{\mathrm{all}})$ as the minimal number of information operations from Λ^{all} needed to compute the solution to within ε in the randomized setting and in the worst case setting, respectively, using the absolute or normalized error criterion. As in the previous subsection we note that the initial errors in both the worst case and randomized settings are the same, which is why we may consider both error criteria.

Then (3.16) implies that

$$\tfrac{1}{4} n^{\mathrm{wor}}(2\varepsilon, S, \Lambda^{\mathrm{all}}) \le n^{\mathrm{ran}}(\varepsilon, S, \Lambda^{\mathrm{all}}) \le n^{\mathrm{wor}}(\varepsilon, S, \Lambda^{\mathrm{all}}) \quad \text{for all } \varepsilon > 0. \tag{3.17}$$

Again the second inequality is obvious, and the essential part of (3.17) is the first inequality. We now show that (3.17) can be used for tractability studies.

Assume that we have a sequence of linear operators $S_d : H_d \to G_d$ defined between Hilbert spaces H_d and G_d of functions of d variables. We take F_d as the unit ball of H_d and assume without loss of generality that each S_d is compact. As before we use Λ^{all} as the class of information operations.

We may now talk about tractability of the problem $S = \{S_d\}$. We define

$$n^{\mathrm{ran}}(\varepsilon, S_d, \Lambda^{\mathrm{all}}) \quad \text{and} \quad n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda^{\mathrm{all}})$$

as above, and say that S is *weakly tractable* in the randomized/worst case setting iff

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n^{\mathrm{sett}}(\varepsilon, S_d, \Lambda^{\mathrm{all}})}{\varepsilon^{-1} + d} = 0,$$

where sett $\in \{\mathrm{ran}, \mathrm{wor}\}$.

Then (3.17) implies that S is weakly tractable in the worst case setting iff S is weakly tractable in the randomized setting.

The same is true for polynomial tractability. Obviously, S is *polynomially tractable* in the randomized/worst case setting iff

$$n^{\mathrm{sett}}(\varepsilon, S_d, \Lambda^{\mathrm{all}}) \le C \varepsilon^{-p} d^q \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d \in \mathbb{N},$$

where sett $\in$ {ran, wor}. Strong polynomial tractability holds if $q = 0$.

Then (3.17) implies that strong polynomial and polynomial tractability of S are equivalent in the worst case and randomized settings with the same exponents for ε^{-1} and d.

We stress that this tractability equivalence holds for the absolute and normalized error criteria and for the class Λ^{all}. As we know from the previous subsection, it is *not* true for the class Λ^{std}.

3.3 Open Problems

We end this introductory chapter by presenting fifteen open problems related to multivariate problems discussed in this chapter. We will continue to present open problems in later chapters and we will number them accordingly to their occurrence. We hope that our readers will find these open problems challenging.

In Section 3.1.1 we considered multivariate integration for Lipschitz functions. Assume now that we have smoother functions that are r times continuously differentiable and consider the class

$$F_{d,r} = \big\{ f : [0,1]^d \to \mathbb{R} \mid \|f\|_{d,r} := \max_{\alpha:|\alpha|\leq r} \max_{x\in[0,1]^d} |D^\alpha f(x)| \leq 1 \big\}.$$

This class has been studied for many linear multivariate problems $S_d : F_{d,r} \to G_d$ for specific continuous linear operators S_d and normed linear spaces G_d. Examples include multivariate integration,

$$S_d f := \text{INT}_d f = \int_{[0,1]^d} f(x)\, dx \quad \text{with } G_d = \mathbb{R},$$

and multivariate approximation

$$S_d f := \text{APP}_d f = f \quad \text{with } G_d = L_p([0,1]^d) \text{ for some } p \in [1,\infty].$$

Let $e(n,d)$ denote the minimal worst case error that can be achieved by algorithms using n function values, and let

$$n(\varepsilon,d) = n(\varepsilon, S_d, F_{d,r}, G_d)$$

denote the minimal number of function values needed to solve the problem to within $\varepsilon\|S_d\|$. This means we now consider the normalized error criterion; as always, this coincides with the absolute error criterion whenever $\|S_d\| = 1$.

For multivariate integration and approximation, i.e., for $S_d \in \{\text{INT}_d, \text{APP}_d\}$, it is known, see Bakhvalov [7], Heinrich [80] and [160], [242], that $e(n,d) = \Theta(n^{-r/d})$, which implies that

$$n(\varepsilon,d) = \Theta(\varepsilon^{-d/r}).$$

For a fixed smoothness parameter r and for varying d, this means that the problem is *not* polynomially tractable.

The question is whether the problem is *weakly* tractable, i.e., whether

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0.$$

If we go back to Bakhvalov's proof, then we conclude that indeed there are two sequences of positive numbers $C_{d,1}$ and $C_{d,2}$ such that

$$C_{d,1}\varepsilon^{-d/r} \le n(\varepsilon, d) \le C_{d,2}\varepsilon^{-d/r} \quad \text{for all } \varepsilon \in (0, 1).$$

However, $C_{d,1}$ is exponentially small in d whereas $C_{d,2}$ is exponentially large in d. This means that these bounds are too weak to establish whether weak tractability holds. This leads us to the first open problem.

Open Problem 1.

- Consider the class $F_{d,r}$ for fixed r and algorithms using function values. Let $S_d \in \{\text{INT}_d, \text{APP}_d\}$. Verify whether $S = \{S_d\}$ is weakly tractable.

- Consider the class $F_{d,r}$ for a fixed r and algorithms using arbitrary linear functionals. Verify whether multivariate approximation $\text{APP} = \{\text{APP}_d\}$ is weakly tractable.

Polynomial intractability was shown for fixed r by noting that the exponent d/r of ε^{-1} goes to infinity with d. Assume now that r can vary with d, i.e., $r = r(d)$, and consider the class $F_{d,r(d)}$. Clearly, if the sequence $d/r(d)$ is unbounded then we cannot have polynomial tractability. Assume then that $\sup_d d/r(d) < \infty$. In particular, we can even consider the case $r(d) = \infty$, which means that we deal with infinitely differentiable functions. Obviously, $F_{d,\infty} \subseteq F_{d,r(d)}$.

What happens now with polynomial tractability? There are a few interesting negative results for multivariate integration and approximation with $r(d) = \infty$. Obviously, these results also apply for all $F_{d,r(d)}$ with finite $r(d)$.

For multivariate integration, $S_d = \text{INT}_d$, it was proved by J. Wojtaszczyk [284] that

$$\lim_{d\to\infty} e(n, d) = 1 \quad \text{for any } n.$$

This implies that

$$\lim_{d\to\infty} n(\varepsilon, \text{INT}_d, F_{d,\infty}, \mathbb{R}) = \infty \quad \text{for any } \varepsilon \in (0, 1),$$

and so this problem is *not* strongly polynomially tractable. In fact, this means even more, namely that multivariate integration cannot be strongly T-tractable; that is the inequality

$$n(\varepsilon, \text{INT}_d, F_{d,\infty}, \mathbb{R}) \le C T(\varepsilon^{-1})^p \quad \text{for all } \varepsilon \in (0, 1),$$

cannot hold for any function $T : [1, \infty) \rightarrow [1, \infty)$ and numbers C and p, see Chapter 8 where generalized tractability is studied.

For multivariate approximation, $S_d = \text{APP}_d$ with $G_d = L_\infty([0, 1]^d)$, and for algorithms using arbitrary linear functionals, the same result was proved by Huang and Zhang [100]. Hence, strong polynomial tractability also does *not* hold for this approximation problem.

But it is still open whether polynomial or weak tractability holds for the class $F_{d,\infty}$ or for the class $F_{d,r(d)}$ with some finite $r(d)$. This leads us to the next open problem.

Open Problem 2.

- Consider the class $F_{d,r(d)}$ with bounded $d/r(d)$, and algorithms using function values. Let $S_d \in \{\text{INT}_d, \text{APP}_d\}$. Verify whether $S = \{S_d\}$ is polynomially or weakly tractable.

- Consider the class $F_{d,r(d)}$ with bounded $d/r(d)$ and algorithms using arbitrary linear functionals. Verify whether multivariate approximation $\text{APP} = \{\text{APP}_d\}$ is polynomially or weakly tractable.

- What is the answer to these two questions if $r(d) = \infty$?

In Section 3.1.2 we considered the integration problem for a finite dimensional space F_d of trigonometric functions. It was relatively easy to show that for positive quadrature formulas Q_n we obtain

$$[e(Q_n)]^2 = 1 - n2^{-d} \quad \text{for } n \le 2^d.$$

It is not clear whether the problem is weakly or polynomially tractable since we do not know whether positive quadrature formulas are optimal. This leads to the following open problem.

Open Problem 3.

- Prove (or disprove) the same bound for general quadrature formulas, i.e.,

$$[e(n, d)]^2 = 1 - n2^{-d} \quad \text{for all } n \le 2^d.$$

This example is discussed in [167], where one can also find two equivalent conjectures:

- For any given points $x_1, x_2, \ldots, x_n \in [0, 1]^d$ there is a trigonometric polynomial f of degree 1 in each variable such that

$$f(x_1) = f(x_2) = \cdots = f(x_n) = 1 \quad \text{and} \quad \|f\| \le 2^{-d/2} \cdot n^{1/2}$$

with the norm as in Example 2.

– Any matrix $A \in \mathbb{R}^{n \times n}$ with entries a_{ij} defined by

$$a_{ij} = \left(\prod_{k=1}^{d} \frac{1 + \cos(\alpha_{i,k} - \alpha_{j,k})}{2} \right) - \frac{1}{n},$$

for arbitrary real $\alpha_{i,k}$, is positive semi-definite.

- If $e(n, d)$ is of the form conjectured here, then the problem is intractable. However, if $e(n, d)$ is *not* of this form, tractability of the problem may be still open. This suggests the next problem of verifying polynomial and weak tractability of this integration problem independently of the form of $e(n, d)$.

In Section 3.1.3 we considered multivariate integration for the *unweighted* Korobov space with bounded Fourier coefficients. Consider now a *weighted* Korobov space by assuming that

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

Here, $\mathfrak{u}(h) = \{ j \mid h_j \neq 0 \}$ and $\gamma = \{ \gamma_{d,\mathfrak{u}} \}$ is a sequence of non-negative weights with $\mathfrak{u} \subseteq \{1, 2, \ldots, d\}$ and $d \in \mathbb{N}$.

For $\gamma_{d,\mathfrak{u}} \equiv 1$, we have the unweighted Korobov space previously studied, for which we know that multivariate integration is intractable. What happens for general weights γ? What conditions on the weights imply tractability?

Multivariate integration for weighted Korobov spaces has been studied in many papers for the case where bounds on the Fourier coefficients are given in the L_2 sense, i.e., for

$$\left(\sum_{h \in \mathbb{Z}^d} \gamma_{d,\mathfrak{u}(h)}^{-1} |\hat{f}(h)|^2 \left(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_d \right)^{\alpha} \right)^{1/2} \leq 1,$$

see e.g., Dick [42], Dick and Kuo [43], [44], Kuo [119], Wang, Sloan and Dick [257] as well as [45], [95], [217]. The case with bounds in the L_∞ case, i.e.,

$$\sup_{h \in \mathbb{Z}^d} \gamma_{d,\mathfrak{u}(h)}^{-1} |\hat{f}(h)| \left(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_d \right)^{\alpha} \leq 1,$$

has not been yet studied. It would be interesting to see what we must assume about the sequence γ to obtain different types of tractability for multivariate integration. This leads us to the next open problem.

Open Problem 4.

- Consider the class $F_{d,\gamma,\alpha}$ with the sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ of weights. Find necessary and sufficient conditions on γ to obtain strong polynomial, polynomial and weak tractability of multivariate integration.

In Section 3.1.4 we considered multivariate approximation $\text{APP}_d : F_{d,p} \to G_{d,m,p}$. Here, $F_{d,p}$ is the class of infinitely differentiable functions with derivatives bounded

in the L_p-norm. The norm in the target space $G_{d,m,p}$ depends on m, which determines how many derivatives of functions we want to approximate. For $m \geq 1$, we have intractability for any p, whereas for $m = 0$ and $p = 2$ we have weak tractability but polynomial intractability. This leaves the case $m = 0$ and $p \neq 2$.

Furthermore, the partially positive result for $m = 0$ and $p = 2$ was obtained by assuming that algorithms use arbitrary linear functionals. It is not clear what happens if we allow algorithms that can only use function values. This leads us to the next open problem.

Open Problem 5.

- Consider multivariate approximation as in Section 3.1.4 with $m = 0$ and $p \neq 2$, and algorithms using arbitrary linear functionals. Verify whether weak or polynomial tractability hold.

- Consider multivariate approximation as in Section 3.1.4 with $m = 0$ and arbitrary $p \in [1, \infty]$, and algorithms using only function values. Verify whether weak tractability holds.

In Section 3.1.5 we discussed discrepancy and its relation to multivariate integration. For the star discrepancy, which corresponds to $p = \infty$, we have polynomial tractability with the exponents of n^{-1} and d equal to $\frac{1}{2}$. However, the proof of this result is non-constructive. It is important to find a construction of points even if this results in larger exponents. The best constructions known at this time have running time exponential in d, see Doerr and Gnewuch [46], Doerr, Gnewuch, Kritzer and Pillichshammer [47], Doerr, Gnewuch and Srivastav [48], Kritzer and Pillichshammer [116], and Thiémard [235], which is infeasible for large d. This leads us to the next open problem.

Open Problem 6.

- Construct points $t_1, t_2, \ldots, t_n \in [0, 1]^d$ for which

$$\mathrm{disc}^*_\infty(t_1, t_2, \ldots, t_n) = \mathcal{O}(d^q n^{-r}) \quad \text{for all } n, d = 1, 2, \ldots,$$

  with the factor in the big $\mathcal{O}$ notation independent of d and n. The running time of the construction should be polynomial in d and n. Here q and r are positive numbers. Obviously, the construction for which q is small and r is relatively large, i.e., $q = r = \frac{1}{2}$, would be especially interesting.

We already mentioned that it is unknown whether the exponent $\frac{1}{2}$ of n^{-1} in the star discrepancy bound can be improved. In fact, there is a conjecture, see Heinrich [83], that states that the exponent $\frac{1}{2}$ is sharp as long as we have a polynomial dependence on d in bounds on the star discrepancy. We repeat this conjecture as the next open problem.

Open Problem 7.

- Consider the minimal star discrepancy $\text{disc}_\infty^*(n, d)$. Let

$$p^* = \sup\{p \mid \text{disc}_\infty^*(n, d) \le C d^k n^{-p} \text{ for some } C, k \text{ and all } n, d \in \mathbb{N}\}.$$

    Verify whether $p^* = \frac{1}{2}$. If not, determine p^*.

We want to add that there are many more open problems concerning different types of discrepancy. These will be presented in Volume II, which deals with discrepancy in greater depth.

In Section 3.1.6 we considered diagonal multivariate problems for weighted spaces related to the ANOVA decomposition, and we analyzed algorithms using arbitrary linear functionals.

We could not analyze algorithms using function values for the class $F_{d,\gamma}$ with positive weights. The reason is that these spaces are built on L_2 spaces, for which function values are not even well defined. To allow the study of algorithms using function values we need to restrict the class $F_{d,\gamma}$ to functions for which function values are well defined. This can be done in many different ways. Here, we propose one such restriction. Using the ANOVA decomposition of f from $L_2([0,1]^d)$ we know that $f_{\mathfrak{u}} \in L_2([0,1]^{|\mathfrak{u}|})$. Assume additionally that $f_{\mathfrak{u}}$ is smooth, say, it belongs to the space $C^r([0,1]^{|\mathfrak{u}|})$ for some $r \ge 0$. The norm in the space $C^r([0,1]^{|\mathfrak{u}|})$ is given by

$$\|f_{\mathfrak{u}}\|_r = \max_{\alpha:|\alpha| \le r} \max_{x_u \in [0,1]^{|\mathfrak{u}|}} |D^\alpha f_{\mathfrak{u}}(x_{\mathfrak{u}})|.$$

Then we can restrict the class $F_{d,\gamma}$ by assuming that $\|f_{\mathfrak{u}}\|_r \le 1$. That is, the new class is of the form

$$F_{d,\gamma,r} = \Big\{ f \in F_{d,\gamma} \mid \max_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \|f_{\mathfrak{u}}\|_r \le 1 \Big\}.$$

Function values of f from $F_{d,\gamma,r}$ are now well defined. Hence we can study tractability for algorithms using only function values, as well as for algorithms using arbitrary linear functionals. In particular, we can study how the smoothness parameter r helps. This leads us to the next open problem.

Open Problem 8.

- Consider the diagonal multivariate problem for the class $F_{d,\gamma,r}$ as in Section 3.1.6, and algorithms using only function values. Give necessary and sufficient conditions on γ for weak and polynomial tractability.

- Consider the diagonal multivariate problem for the class $F_{d,\gamma,r}$ as in Section 3.1.6, and algorithms using arbitrary linear functionals. Give necessary and sufficient conditions on γ for weak and polynomial tractability.

In Section 3.2.2 we considered Gaussian integration for the class of continuous d-variate functions equipped with the isotropic Wiener measure. We know that the problem is strongly polynomially tractable, with a sharp exponent of ε^{-1} equal to 2.

However, the proof was semi-constructive, since the selection of points at which we compute the function with good average case error bounds was done probabilistically. It is of interest to construct such points deterministically. This is our next open problem.

Open Problem 9.

- Consider Gaussian integration for the class of continuous d-variate functions equipped with the isotropic Wiener measure. For all d, construct points

$$x_1, x_2, \ldots, x_n \quad \text{from } \mathbb{R}^d$$

    for which the quasi-Monte Carlo algorithm $A_n(f) = n^{-1} \sum_{j=1}^{n} f(x_j)$ has the average case error of order $\mathcal{O}(d^q n^{-r})$, with the factor in the big $\mathcal{O}$ notation independent of d and n. The running time of the construction should be polynomial in d and n.

In Section 3.2.3 we considered multivariate approximation for functions that are r times continuously differentiable functions in each variable. This class was equipped with the folded Wiener sheet measure, defined as the classical Wiener sheet measure placed on rth derivatives. We showed that as long as r is fixed, the problem is intractable. It would be of interest to allow r to depend on d, i.e., $r = r(d)$, as in Open Problem 2. Then a natural question is whether we can break intractability by making $r(d)$ sufficiently large. This is our next open problem.

Open Problem 10.

- Consider multivariate approximation for the class $C_0^{r(d),r(d),\ldots,r(d)}$ equipped with the folded Wiener sheet measure as in Section 3.2.3. Give necessary and sufficient conditions on γ for weak or polynomial tractability.

Another way of dealing with smoothness is to consider different smoothness in each variable. That is, instead of assuming that functions are r times continuously differentiable in each variable, let us assume that they are $r_{d,j}$ times continuously differentiable in the jth variable. This corresponds to the class $C_0^{r_{d,1},r_{d,2},\ldots,r_{d,d}}$ equipped with the folded Wiener sheet measure placed on $r_{d,j}$ derivatives. This is a zero-mean Gaussian measure whose covariance function is

$$K_d(x, y) = \prod_{j=1}^{d} \int_0^1 \frac{(x_j - u)_+^{r_{d,j}}}{r_{d,j}!} \frac{(y_j - u)_+^{r_{d,j}}}{r_{d,j}!} \, du.$$

This class was analyzed by Papageorgiou and Wasilkowski [188] who found the optimal order of convergence and the asymptotic constant; however, they did not study tractability. A natural question is whether we can obtain tractability for some $r_{d,j}$. Obviously, tractability may hold only if the $r_{d,j}$'s are large enough and the point is to verify whether sufficiently large $r_{d,j}$'s indeed lead to tractability. If so we would have an example for which increasing smoothness implies tractability. This is our next open problem.

Open Problem 11.

- Consider multivariate approximation for the class $C_0^{r_{d,1}, r_{d,2}, \ldots, r_{d,d}}$ equipped with the folded Wiener sheet measure similarly as in Section 3.2.3. Find necessary and sufficient conditions on $\{r_{d,j}\}$ to obtain weak tractability and polynomial tractability.

In Section 3.2.6 we studied the relative error for Gaussian integration. We mentioned that the problem cannot be solved under this error criterion in the worst and average case settings. The reason why this result holds is that when finitely many function values are 0, we were unable to check whether the integral was 0. In fact, a small relative error, say,

$$\frac{|\mathrm{INT}_d f - A_n(f)|}{|\mathrm{INT}_d f|} \leq \frac{1}{10}$$

for small integrals, say $|\mathrm{INT}_d f| \leq 10^{-k}$, means that the absolute error of $A_n(f)$ must be at least 10^{-k-1}. Since k can be arbitrarily large this requires that A_n almost exactly recovers small integrals, which as we know, can be done only in the probabilistic setting.

From this point of view, there is a valid criticism of the relative error criterion for problems with small solutions. A possible remedy is to study a combination of the absolute and relative error by taking a positive (and small) number η and measure the error by

$$\frac{|\mathrm{INT}_d f - A_n(f)|}{|\mathrm{INT}_d f| + \eta},$$

see [242] where this modified relative error has been studied. Hence, if $|\mathrm{INT}_d f|$ is larger than η, this is close to the relative error, if $|\mathrm{INT}_d f|$ is comparable to or smaller than η, this is close to the absolute error modulo a factor of order η^{-1}.

It is easy to see that now we can solve the problem in the worst case and average case setting with ε replaced by $\varepsilon \eta$; however, the minimal number of function values needed for the solution will go to infinity as η goes to 0.

Consider the probabilistic setting with this modified error criterion. The probabilistic error of A_n is now given by

$$e^{\mathrm{prob\text{-}mod}}(A_n, \delta, \eta) = \inf_{B: w_d(B) \leq \delta} \sup_{f \in F_d - B} \frac{|\mathrm{INT}_d f - A_n(f)|}{|\mathrm{INT}_d f| + \eta}.$$

Let

$$n^{\mathrm{prob\text{-}mod}}(\varepsilon, \delta, \eta, \mathrm{INT}_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{prob\text{-}mod}}(A_n, \delta, \eta) \leq \varepsilon\}$$

be the minimal number of function values needed to solve the problem to within ε with this modified error criterion.

Observe that we obtain an obvious upper bound on $n^{\mathrm{prob\text{-}mod}}(\varepsilon, \delta, \eta, \mathrm{INT}_d)$ by switching to the absolute error with ε replaced by $\varepsilon \eta$, finding that

$$n^{\mathrm{prob\text{-}mod}}(\varepsilon, \delta, \eta, \mathrm{INT}_d) \leq n^{\mathrm{prob\text{-}abs}}(\varepsilon \eta, \delta, \mathrm{INT}_d) = \mathcal{O}\left(\frac{\ln \delta^{-1}}{\varepsilon^2 \eta^2}\right).$$

So the dependence on δ^{-1} is only through $\ln \delta^{-1}$; however, the dependence on η is through η^{-1}. For small η, we can propose another upper bound by switching to the relative error criterion and by neglecting the parameter η. That is,

$$n^{\text{prob-mod}}(\varepsilon, \delta, \eta, \text{INT}_d) \leq n^{\text{prob-rel}}(\varepsilon, \delta, \text{INT}_d) = \mathcal{O}\left(\frac{1}{\varepsilon^2 \delta^2}\right).$$

So the dependence on δ^{-1} is now much more severe; however, the dependence on η disappears.

There is thus a trade-off between the dependence on δ^{-1} and η^{-1}, which will be the essence of our next open problem.

We briefly discuss tractability. We have now one more parameter η for tractability study. Define tractability when there exist five non-negative numbers C, p, q, s and t such that

$$n^{\text{prob-mod}}(\varepsilon, \delta, \eta, \text{INT}_d) \leq C \varepsilon^{-p} d^q \left(g_1(\delta^{-1})\right)^s \left(g_2(\eta^{-1})\right)^t$$

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

Here, the functions g_i are either $g_i(x) = x$ or $g_i(x) = \ln x$. That is, for $g_1(x) = g_2(x) = x$ we allow polynomial dependence on all four parameters, whereas for $g_1(x) = g_2(x) = \ln x$ we allow poly-log dependence on δ^{-1} and η^{-1}, and polynomial dependence on ε^{-1} and d.

Open Problem 12.

- Find sharp estimates of
$$n^{\text{prob-mod}}(\varepsilon, \delta, \eta, \text{INT}_d)$$
  as a function of the four variables ε, δ, η and d.

- Do we have tractability for all four choices of (g_1, g_2)?

In Section 3.2.8 we studied the randomized setting for multivariate integration. In particular, we mentioned that for the weighted Sobolev space $H(K_d)$ with the kernel

$$K_d(x, y) = \prod_{j=1}^{d} \left(1 + \gamma_j \min(x_j, y_j)\right)$$

for some sequence of weights γ_j with $\gamma_1 \geq \gamma_2 \geq \cdots \geq 0$, we obtain strong polynomial tractability in the randomized setting for the normalized error criterion if

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

If the latter condition holds then the Monte Carlo reduces the initial error to within ε using $\mathcal{O}(\varepsilon^{-2})$ randomized function values, with the factor in the big $\mathcal{O}$ notation independent of d. The open question is whether this condition is also necessary for strong polynomial tractability. We know that it is necessary for good behavior of the Monte Carlo algorithm, but perhaps a more relaxed condition on the weights would suffice for a more sophisticated algorithm. This leads us to the next open problem

Open Problem 13.

- Find necessary and sufficient conditions on the weights $\{\gamma_j\}$ of the Sobolev space $H(K_d)$ for strong polynomial and polynomial tractability of multivariate integration in the randomized setting for the normalized error criterion.

- Find the smallest exponent of ε^{-1} when this problem is strongly polynomially tractable.

It is interesting to add that for the weighted Sobolev space $H(\widetilde{K}_d)$ of periodic functions, see Section A.2.2, with the kernel

$$\widetilde{K}_d(x, y) = \prod_{j=1}^{d} \left(1 + \gamma_j \left(\min(x_j, y_j) - x_j y_j\right)\right),$$

the Monte Carlo algorithm can be improved as shown by Wasilkowski [264]. Namely, the Monte Carlo algorithm enjoys a strong polynomial tractable randomized error bound in this space of periodic functions iff

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

as shown in [219]. Hence, we have the same condition as for the non-periodic case of the space $H(K_d)$.

For the space $H(\widetilde{K}_d)$, Wasilkowski proposed a randomized algorithm that enjoys a strong polynomial tractable randomized error bound if

$$\sum_{j=1}^{\infty} \gamma_j^3 < \infty.$$

Hence, it is enough to assume that the weight sequence is cubic-summable, which obviously relaxes the Monte Carlo condition of square-summability. But it is unknown whether cubic-summability is necessary for strong polynomial tractability. This is our next open problem.

Open Problem 14.

- Find conditions on the weights $\{\gamma_j\}$ of the Sobolev space $H(\widetilde{K}_d)$ of periodic functions that are necessary and sufficient for strong polynomial and polynomial tractability of multivariate integration in the randomized setting for the normalized error criterion.

- Find the exponent of strong polynomial tractability of this problem.

In Section 3.2.9 we showed that randomization does not really help for the class Λ^{all} as long as we consider compact linear problems defined over Hilbert spaces. The

assumption that we consider Hilbert spaces is essential since randomization may indeed help if we switch to Banach spaces as proved by Mathé [141] and Heinrich [79]. One of the major problems of information-based complexity is to characterize linear problems for which randomization essentially helps, as compared to the worst case setting. By "essentially", we mean by more than a constant. This should be done for both classes Λ^{std} and Λ^{all} and for various error criteria. It is of great interest to characterize multivariate problems that are weakly or polynomially tractable in the randomized setting, and to characterize multivariate problems which are intractable or polynomially intractable in the worst case setting but weakly or polynomially tractable in the randomized setting. We summarize these questions in the next open problem.

Open Problem 15.

- Characterize linear operators, their source and target spaces for which the randomized setting is essentially easier than the worst case setting for the two classes of information Λ^{std} and Λ^{all} and for the absolute and normalized error criteria.

- Characterize linear multivariate problems that are weakly tractable or polynomially tractable in the randomized setting.

- Characterize linear multivariate problems for which randomization breaks intractability or polynomial intractability in the worst case setting.

3.4 Notes and Remarks

NR 3.1.1:1. Intractability of multivariate integration for the class of Lipschitz functions implies intractability of other problems that are at least as hard as multivariate integration. For example, this holds for multivariate approximation, $\mathrm{APP}_d f = f \in L_p([0,1]^d)$ for $f \in F_d^{\mathrm{lip}}$, where $p \in [1,\infty]$. This follows from the simple fact that $|\mathrm{INT}_d f| \le \|f\|_p$ and from general results of information-based complexity. Indeed, we can solve the multivariate integration problem to within ε iff

$$\inf_{x_1,x_2,\ldots,x_n \in [0,1]^d} \sup_{f \in F_D^{\mathrm{lip}} \ f(x_j)=0} |\mathrm{INT}_d f| \le \varepsilon,$$

which can happen iff $n \ge n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d, F_d^{\mathrm{lip}})$. Hence, for $n < n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d, F_d^{\mathrm{lip}})$ and for any points $x_1, x_2, \ldots, x_n$, there exists a function from F_d^{lip} such that $f(x_j) = 0$ and $|\mathrm{INT}_d f| > \varepsilon$. Since the best approximation of multivariate approximation for zero function values is 0, we conclude that the worst case error of any algorithm for multivariate approximation is also larger than ε. Hence, we can solve the multivariate approximation problem only if $n \ge n^{\mathrm{wor}}(\varepsilon, \mathrm{INT}_d, F_d^{\mathrm{lip}})$, which implies intractability, as claimed.

In turn, intractability of multivariate approximation for the class F_d^{lip} implies intractability of all problems (with respect to the absolute error) that are at least as hard

as approximation. We add that relations of various computational problems to the approximation problem have been studied by Wasilkowski [259], Werschulz [277], and in [160].

NR 3.1.1:2. We considered the absolute error criterion for multivariate integration of Lipschitz functions. Observe that the initial error of this problem and the initial error of the multivariate approximation problem mentioned above are both 1. Hence, the same intractability results hold for the normalized error criterion.

NR 3.1.3:1. For the Korobov class, the initial error of multivariate integration is 1, so the same intractability result holds for the normalized error criterion. Now consider the multivariate approximation problem $APP_d f = f \in L_p([0, 1]^d)$ for $f \in F_{d,\alpha}$ and $p \in [1, \infty]$. As before, this problem is intractable if we use the absolute error criterion. It is not clear, however, what happens with the normalized error criterion, since the initial error may be exponentially large in d, as happens for $p = 2$.

NR 3.1.4:1. This section is based on [179].

NR 3.1.5:1. The notion of discrepancy plays an important role in the study of tractability. The first tractability papers aimed at understanding the empirical success of QMC algorithms for finance applications of 360-dimensional integrals, see the book of Traub and Werschulz [243] for a survey. Since the error of a QMC algorithm can be expressed by various notions of discrepancy, it was natural to try to understand why discrepancy behaves so well for large d. As already indicated, this led us to weighted spaces and to product and finite-order weights. The story of discrepancy and tractability will be continued in Volume II.

NR 3.1.6:1. This section is new, although technical results are based on the existing papers indicated in the text.

NR 3.2.2:1. As mentioned in the text, strong polynomial tractability of Gaussian integration for the isotropic Wiener measure and for the normalized error criterion was shown in [94]. The use of Chebyshev's inequality for semi-construction of good sample points is standard. The intriguing dependence of the initial error on d, and polynomial tractability for the absolute error criterion is new, although quite straightforward.

NR 3.2.3:1. The major technical part of this section is based on Papageorgiou and Wasilkowski [188], but the tractability analysis is new.

NR 3.2.4:1. Tractability in the probabilistic setting has not been yet thoroughly studied. There are, however, interesting papers by Lifshits and Tulyakova [135], and Lifshits and Zani [136], with negative results for multivariate approximation. For linear functionals, there is a very close relation between the probabilistic and average case settings, which is why we studied a special linear functional given by Gaussian integration. We concentrated on the dependence on the new error parameter δ appearing in the probabilistic setting.

NR 3.2.5:1. The dependence on $\ln \delta^{-1}$ is typical for the absolute error criterion. We showed that there is *no* dependence on δ for the normalized error criterion. Again this is true for all linear functionals, but not for general linear operators.

   We want to stress once more that the notion of the initial error, which is quite natural in the worst case and average case settings, is not so natural in the probabilistic setting as long as δ goes to 0. The reason is that the initial error goes (slowly) to infinity, making the problem easier under the normalized error criterion. This would change if instead of the whole space, we consider a ball of finite radius. Then the initial error in the probabilistic setting with δ tending to 0 will go to the worst case initial error. The study of the probabilistic and average case settings over balls of finite radius and the use of normalized Gaussian measure would allow us to compare the worst, average and probabilistic settings.

NR 3.2.6:1. The relative error may seem to be the most practical error criterion. Unfortunately, as indicated in this section, the relative error is the source of misleading negative results in the worst case setting for linear problems, and in the average case setting for linear functionals. Only in the probabilistic setting can we solve linear functionals under the relative error criterion, although the price is paid by a much more severe dependence on δ^{-1}. We believe that a good alternative to the relative error is the modified error, which is a combination of the absolute and relative error criteria, see Open Problem 12.

NR 3.2.8:1. We discussed the Monte Carlo algorithm for multivariate integration in the randomized setting. There are many variants of this algorithm. One stream of work deals with reducing the variance of the function, which occurs in the randomized error of the Monte Carlo. The idea is to replace the function f by a function g. The function g should have the same integral of f, or we should know a simple formula that allows to compute the integral of f if we know the integral of g. Furthermore, function values $g(x)$ should be easily computed by function values of f. And the main point is that g should have the variance as small as possible. Different transformations for g have been proposed in the literature.

NR 3.2.9:1. We indicated that randomization does not really help for linear problems defined over Hilbert spaces as long as the class Λ^{all} of all linear functionals is allowed. A natural question is what happens if instead of Λ^{all} only the class Λ^{std} can be used. This problem will be studied in Volumes II and III.

   We wish to stress that there are many open problems in the randomized setting. It seems to us that the randomized setting is especially difficult to analyze; this should be regarded as an additional challenge for researchers. The technical difficulty of this setting is the reason that today we have a relatively small number of results. This explains why tractability of many multivariate problems in the randomized setting is unknown or only partially known.

NR 3.3:1. We decided to introduce open tractability problems as soon as possible in this book. We want the reader to start thinking about solving these open problems

without the necessity of studying the rest of this book and grasping many technical results that are sometimes not so easy. In the further course of this book we will be presenting additional open problems. The total number of open problems in Volume I is 30. We hope that this number will quickly decrease.

Chapter 4
Basic Concepts and Survey of IBC Results

In this chapter we present basic definitions and survey results from information-based complexity (IBC) that are needed to study tractability. In Sections 4.1–4.3, we assume that

$$S \colon F \to G \tag{4.1}$$

is any mapping, called the *solution operator*. The set F is a subset of a normed space, such as the unit ball, and the problem elements of F are usually functions defined on a set $D \subseteq \mathbb{R}^d$. The set G is a normed space.

We will discuss the complexity of computing values of S to within some error. In Section 4.1 we recall the basic definitions for the worst case setting and in Section 4.2 we state and prove certain important facts about this setting.

In Section 4.3 we consider different settings. We first study the average case and probabilistic settings followed by the randomized setting. In the worst case, average case and probabilistic settings we use deterministic algorithms, whereas in the randomized setting we use randomized algorithms which are sometimes also called Monte Carlo algorithms.

In Section 4.4 we assume that a whole family of such operators

$$S_d \colon F_d \to G_d, \quad d \in \mathbb{N}, \tag{4.2}$$

is given. The number d is usually the number of variables of functions from F_d. Often F_d is the unit ball of an infinite dimensional space.

For much of the following the reader can think of the examples $S_d = \mathrm{INT}_d$ and $S_d = \mathrm{APP}_d$, where the *integration problem* is given by

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

with $G_d = \mathbb{R}$ and the *approximation problem* is given by $S_d(f) = \mathrm{APP}_d(f) = f$ and G_d is a normed space, often chosen as $L_2(D)$.

4.1 Complexity in the Worst Case Setting

We assume that $S \colon F \to G$ is a mapping, where F is a subset of a normed space $\widetilde{F}$ and G is a normed space. The problem elements of $\widetilde{F}$ are usually functions defined on a set $D \subseteq \mathbb{R}^d$. We want to compute an approximation of $S(f)$ for $f \in F$. The set F is typically infinite dimensional and hence we cannot input $f \in F$ directly into the computer. We always work with the *real number model*, which implies that we can

only input finitely many real numbers $y_1, y_2, \ldots, y_n$ into the computer. Clearly, these numbers $y_i = y_i(f)$ should describe $f \in F$ as well as possible. Let

$$N(f) = [y_1(f), y_2(f), \ldots, y_n(f)] \in \mathbb{R}^n \tag{4.3}$$

be the *information* about f. We use algorithms of the form

$$A(f) = \varphi(N(f)), \tag{4.4}$$

where $\varphi, \varphi \colon N(F) \to G$, transforms $y = N(f)$ into an element of the target space G. The *worst case error* of A is given by

$$e^{\mathrm{wor}}(A) = \sup_{f \in F} \|S(f) - A(f)\|. \tag{4.5}$$

Remark 4.1. Here we use the *error* at f as

$$e(A, f) = \|S(f) - A(f)\|$$

that leads to the *worst case error* (4.5) of A by taking the supremum with respect to $f \in F$. In this book we primarily use this error but for some computational problems it is not appropriate, see e.g., [240]. Assume, for example, that F is a set of continuous functions on $[0, 1]$ with $f(0) < 0$ and $f(1) > 0$ and we want to approximate an arbitrary zero x of f, i.e., $f(x) = 0$. Here $G = \mathbb{R}$. Since f is continuous and changes sign at 0 and 1, such a number x exists, however, it is in general not unique. In any case, the set $f^{-1}(0) = \{x \in [0, 1] \mid f(x) = 0\}$ is nonempty for all $f \in F$. Then one might consider the *root criterion*, which is defined by

$$e(A, f) = \mathrm{dist}(f^{-1}(0), A(f)),$$

i.e., the error of A at f is now the distance of $A(f) \in \mathbb{R}$ to the nearest zero of f. Starting with this definition of the error at f we then define the worst case error of A by

$$e^{\mathrm{wor}}(A) = \sup_{f \in F} e(A, f).$$

This permits the definition of the error at f and the worst case error without specifying the solution operator S. □

We say that N is *partial* if N is many-to-one. This is generally the case in computational practice for continuous problems. If N is partial, then we are unable to identify $f \in F$ from the information $N(f)$ since there are many problem elements sharing the same information with f. Let

$$N^{-1}(y) = \{\tilde{f} \in F \mid N(\tilde{f}) = y\}$$

be the set of indistinguishable problem elements, and let

$$S(N^{-1}(y)) = \{S(\tilde{f}) \in G \mid \tilde{f} \in N^{-1}(y)\}$$

be the set of indistinguishable solution elements. For $M \subseteq G$ let

$$\mathrm{rad}(M) = \inf_{x \in G} \sup_{m \in M} \|x - m\|$$

be the *radius* of M. Roughly speaking, $\mathrm{rad}(M)$ is the smallest radius of all balls that contain M.

We shall denote the radius of $S(N^{-1}(y))$ by $r^{\mathrm{wor}}(N, y)$, which we call the *local radius of information N at y*, i.e.,

$$r^{\mathrm{wor}}(N, y) = \mathrm{rad}(S(N^{-1}(y))).$$

Observe that for $f \in N^{-1}(y)$ we have $A(f) = \varphi(y)$, so $A(f)$ is constant over $N^{-1}(y)$, and clearly the best we can do to minimize the worst case error of A given N is to take a center, if it exists, of the ball with the radius $r(N, y)$. Then the error is $r^{\mathrm{wor}}(N, y)$. If the center of $N^{-1}(y)$ does not exist then we can take $A(f)$ whose error is arbitrarily close to $r^{\mathrm{wor}}(N, y)$.

The (*global*) *radius of information N* is defined by

$$r^{\mathrm{wor}}(N) = \sup_{y \in N(F)} r^{\mathrm{wor}}(N, y).$$

Hence, the radius $r^{\mathrm{wor}}(N)$ of information N is roughly the smallest radius of balls that contain all sets $S(N^{-1}(y))$ of indistinguishable solution elements.

Hence, we minimize the error of A given N if we take a center of $S(N^{-1}(y))$ for each $y \in N(F)$, and then the error of A is $r^{\mathrm{wor}}(N)$. If the center of $S(N^{-1}(y))$ does not exist for some y then we can take A whose error is arbitrarily close to $r^{\mathrm{wor}}(N)$. This proves the following result, see [242, p. 50], which although quite simple, is nevertheless a basic result often used for the proof of lower bounds.

Theorem 4.2.

$$r^{\mathrm{wor}}(N) = \inf_{\varphi} \sup_{f \in F} \|S(f) - \varphi(N(f))\|. \tag{4.6}$$

Hence, the radius of information determines the worst case error of an optimal algorithm of the form (4.4). The radius of information plays a major role in information-based complexity. It measures the intrinsic uncertainty caused by the partial information N. Observe that $r^{\mathrm{wor}}(N)$ also depends on the mapping $S \colon F \to G$. It does *not*, however, depend on φ or A.

As a technical tool, we also use the *diameter of information N*, since it is often much easier to obtain than the radius. For $M \subseteq G$ let

$$\mathrm{diam}(M) = \sup_{m_1, m_2 \in M} \|m_1 - m_2\|.$$

Roughly speaking, $\mathrm{diam}(M)$ is the largest distance between two points of M. It is easy to prove that for each $M \subseteq G$ we have

$$\mathrm{rad}(M) \leq \mathrm{diam}(M) \leq 2\,\mathrm{rad}(M)$$

and these inequalities cannot be improved in general. That is, for any $\alpha \in [1, 2]$ there exist a normed space G and a set M such that $\mathrm{rad}(M) = \alpha \, \mathrm{diam}(M)$ and $\mathrm{diam}(M) \in (0, \infty)$.

The *local diameter* of information N is defined by

$$d(N, y) = \mathrm{diam}(S(N^{-1}(y))),$$

and the *(global) diameter* of information N by

$$d(N) = \sup_{y \in N(F)} d(N, y).$$

Combining these definitions we have

$$d(N) = \sup\{\|S(f_1) - S(f_2)\| \mid f_1, f_2 \in F, \ N(f_1) = N(f_2)\}.$$

Hence, the diameter $d(N)$ of information N is the largest distance between two solution elements which are indistinguishable with respect to N. Obviously the radius and diameter of information are related by the inequality

$$r(N) \le d(N) \le 2r(N), \tag{4.7}$$

and again these inequalities cannot be improved in general, see [242, p. 47].

Now we discuss the following questions:

- Which information N should be used?

- What is the cost of A?

- How are other errors of A defined?

- Which mappings φ should be used to define the algorithm $A = \varphi \circ N$ with reasonable cost?

- What is the ε-complexity (minimal cost) of approximating S?

4.1.1 Types of Information

To compute an approximation to $S(f)$ we need to know some information about f. Even if we would have full knowledge about f, we could only use partial information about f if F is infinite dimensional since we can only input finitely many numbers $y_1, y_2, \ldots, y_n$ into the computer. We mostly study the case where the numbers y_i are values of linear functionals, $y_i = L_i(f)$. Sometimes the form of L_i is restricted, as will be explained below.

In different parts of mathematics, the numbers y_i may be defined differently, for example by best n-term approximation or by the n largest coefficients of f in some basis. We comment on these choices in the second part of this subsection.

Information Given by Linear Functionals

In IBC we usually assume that we can compute certain linear functionals, and therefore

$$y_i = L_i(f) \quad \text{where } L_i \in \Lambda \subseteq \widetilde{F}^*.$$

Here $\widetilde{F}^*$ is the dual space of $\widetilde{F}$, i.e., it is the set of all continuous linear functionals $L \colon \widetilde{F} \to \mathbb{R}$. The class Λ is the class of admissible information operations given by certain functionals.

We do *not* discuss how the values $y_i = L_i(f)$ are computed, they are just given to us by an *oracle* or *subroutine* or even by some physical *measurement*.

We now discuss two classes of information N. The first one is the class of *non-adaptive* information, where the same information operations are computed for each $f \in F$. The information $N \colon F \to \mathbb{R}^n$ is called *non-adaptive* if there exist $L_1, L_2, \ldots, L_n \in \Lambda$ such that

$$N(f) = [L_1(f), L_2(f), \ldots, L_n(f)] \quad \text{for all } f \in F. \tag{4.8}$$

Observe that in this case the mapping $N \colon \widetilde{F} \to \mathbb{R}^n$ is linear. The number n of information operations, called the *cardinality* of N, is denoted by $\mathrm{card}(N)$. The computation of non-adaptive information can be done very efficiently in parallel and hence this information is sometimes called *parallel* information. The word *passive* information is also used.

The second class of information is called *adaptive*. Now the number and choice of information operations may vary with f, hence $N(f)$ is a finite or even infinite sequence of numbers. More precisely, the information N is called *adaptive* if it is of the form

$$N(f) = [L_1(f), L_2(f, y_1), \ldots, L_{n(f)}(f, y_1, y_2, \ldots, y_{n(f)-1})], \tag{4.9}$$

where $y_1 = L_1(f)$ and $y_i = L_i(f, y_1, y_2, \ldots, y_{i-1})$ for $i = 2, 3, \ldots, n(f)$. Here y_i is the value of the ith information operation and the choice of the ith operation L_i may depend on the previously computed values $y_1, y_2, \ldots, y_{i-1}$. Since only admissible information operations can be performed, we assume that

$$L_i(\cdot, y_1, y_2, \ldots, y_{i-1})$$

belongs to the class Λ for every fixed $y_1, y_2, \ldots, y_{i-1}$.

The number $n(f)$ denotes the total number of information operations on the problem element f and is called the *cardinality* of N at f. It is determined during the process of computing successive values y_i. More precisely, suppose that we have already computed $y_1 = L_1(f)$ and so on until $y_i = L_i(f, y_1, y_2, \ldots, y_{i-1})$. Based on the values $(y_1, y_2, \ldots, y_i)$ we decide whether another functional L_{i+1} is needed or not. If not then $n(f) = i$ and $N(f) = [y_1, y_2, \ldots, y_i] \in \mathbb{R}^i$. Otherwise we choose L_{i+1} and evaluate $y_{i+1} = L_{i+1}(f, y_1, y_2, \ldots, y_i)$. As mentioned above, the decision is made based on the available knowledge about f. More precisely, we have Boolean

functions $\text{ter}_i \colon \mathbb{R}^i \to \{0, 1\}$, called *termination functions*, and in the ith step our termination decision is "yes" if $\text{ter}_i(y_1, y_2, \ldots, y_i) = 1$. Thus, the cardinality $n(f)$ is equal to

$$n(f) = \min\{i \mid \text{ter}_i(y_1, y_2, \ldots, y_i) = 1\}, \qquad (4.10)$$

with the convention that $\min \emptyset = \infty$. Usually we choose termination functions to guarantee that $n(f)$ is finite for all $f \in F$.

Observe that adaptive information N is given by all the mappings L_i and the termination functions ter_i. In general, adaptive information requires sequential computation. That is, we have to wait until y_i is computed to decide whether another information operation is needed and, if so, what it should be. That is why adaptive information is sometimes also called *sequential* information. The word *active* information is also used.

So far, we did not specify the class Λ. In many cases F consists of functions that are defined on a set $D \subseteq \mathbb{R}^d$. In this case it is natural to assume that we have an oracle for function values, and hence Λ is the class of all functionals of the form

$$L(f) = f(x) \quad \text{for all } f \in F, \qquad (4.11)$$

for some $x \in D$. This information is called *standard information*, and is denoted by $\Lambda = \Lambda^{\text{std}}$.

One might assume that also other linear functionals can be computed, such as Fourier coefficients, wavelet coefficients, or other weighted integrals. For many theoretical studies it is convenient to allow *all* continuous linear functionals; we denote the class of all continuous linear functionals by $\Lambda = \Lambda^{\text{all}} = \widetilde{F}^*$.

Most results on tractability refer to the classes Λ^{std} or Λ^{all}. Nevertheless we stress that other types of linear functionals have been studied in the literature. We mention one example, see Bojanov [16] for a survey. Assume that a multiple integral $\int_\Omega f(x)\,dx$ with $\Omega \subseteq \mathbb{R}^d$ is approximately recovered from the given values of some integrals of f over manifolds of lower dimension (as, for example, in tomography). Hence, in this case Λ consists of certain integrals $f \mapsto \int_\Gamma f(x)\,dx$, where $\Gamma \subseteq \Omega$. In particular, such manifolds as planes and spheres are used in Babenko and Borodachov [4], [5].

Information Given by Nonlinear Functionals

It should be stressed that the functionals L_k in (4.9) cannot depend on f in an arbitrary way but only via the already computed values of

$$y_1 = L_1(f), y_2 = L_2(f, y_1), \ldots, y_{k-1} = L_{k-1}(f, y_1, y_2, \ldots, y_{k-2}).$$

This is because we are interested in feasible computations with small cost, including the cost of obtaining the information on f. Hence, we charge for operations that are needed to select the functional $L_k(\cdot, y_1, y_2, \ldots, y_{k-1})$ and we also charge for computing the value $L_k(f, y_1, y_2, \ldots, y_{k-1})$.

Sometimes it is interesting to even allow the L_k to depend on f in a more general way. Usually this is done without a cost analysis. We illustrate this point by two examples that we believe are typical for many other situations.

We begin with the classical problem of *uniform approximation*. For a given univariate real continuous function f defined on $[0, 1]$, we want to find the best approximation in the class P_n of polynomials of degree at most n, i.e.,

$$e_n^{\text{pol}}(f) = \inf_{p \in P_n} \|f - p_n\|, \quad \text{where } \|f - p_n\| = \max_{x \in [0,1]} |f(x) - p_n(x)|.$$

There is a rich and beautiful theory of this problem dating back to Bernstein, Chebyshev, Jackson, Weierstraß and many others. In particular we know that the best polynomial approximation

$$p_{n,f}(x) = L_0(f) + L_1(f)x + \cdots + L_n(f)x^n$$

exists, i.e., $e_n^{\text{pol}}(f) = \|f - p_{n,f}\|$, the error $e_n^{\text{pol}}(f)$ goes to 0 for an arbitrary continuous function with the rate of convergence depending on the smoothness of the function f, etc.

From our point of view, it is important to note that the coefficients $L_i(f)$ of the best polynomial approximation depend *nonlinearly* on f, and they form a special type of information about f. However, we know how to compute these coefficients only for very specific functions f. For a general continuous function f, we only know that they exist but do not know how to compute them. On the other hand, if we knew $L_1(f), L_2(f), \ldots, L_n(f)$ then we would replace f by the polynomial $p_{n,f}$ and many operations with $p_{n,f}$ are computationally easy.

If we want to approximate the coefficients $L_i(f)$ then this can be done by Remez's algorithm, which is a variant of Newton's algorithm, see, e.g., the book of Kowalski, Sikorski and Stenger [115]. However, the information used by Remez's algorithm is given by function values. If we want to do this for a special function f as a part of a precomputation, then the cost is not really an issue and this approach is completely acceptable. Otherwise, if we want to deal with functions from a general class F, then directly approximating f may be much faster than computing its best polynomial approximation.

This approach which does not address the cost of computing best approximants is also typical for other subareas of the theory of approximation and other norms. Examples include rational approximation, approximation by splines with free knots, and wavelet approximation.

Typically, only if the target space is a Hilbert space, it is possible to show that the best approximants depend linearly on f, or equivalently, that the best coefficients $L_i(f)$ are given by linear functionals. Only in the latter case, and only for the class Λ^{all}, it is clear how to compute the best approximants.

We now turn to our second example of *data compression*, see e.g., DeVore [39]. We explain this problem by assuming that $\widetilde{F}$ is a finite dimensional real space, thus $s = \dim(\widetilde{F}) < \infty$, although s is huge, say, at least in the millions. If the ψ_j form a

basis then $f = \sum_{j=1}^{s} c_j(f)\psi_j$ with real numbers $c_j(f)$ that depend linearly on f. Assume that all the $c_j(f)$ are known. In data compression, we want to replace the complete information given by these s numbers by only a few of them at the expense of losing the exact approximation of f. That is, we want to find indices $j_i \in \{1, 2, \ldots, s\}$ for which the partial information $N(f) = [j_1, c_{j_1}(f), j_2, c_{j_2}(f), \ldots, j_n, c_{j_n}(f)]$ will allow us to have a good approximation of f, hopefully, with n much less than s. In many cases, a good choice of c_{j_i} is to select the n largest coefficients from the set $\{|c_1(f)|, |c_2(f)|, \ldots, |c_s(f)|\}$. In this case, we have

$$L_i(f) = c_{j_i}(f) \quad \text{for } i = 1, 2, \ldots, n,$$

and the information is not linear in f. So the cost of computing $N(f)$ will be equal to the cost of selecting n largest elements out of s elements, which is linear in s even if we assume that the $|c_j(f)|$ are given for free. This is a reasonable assumption for data compression, since the only important factor of the cost is to transfer n largest coefficients and the rest of the cost can be neglected as a (usually very expensive) precomputing.

The situation is quite different if we assume that $\widetilde{F}$ is an infinite dimensional space and we still want to compute the n largest coefficients $c_j(f)$ from the representation of $f = \sum_{j=1}^{\infty} c_j(f)\psi_j$. So now we select the n largest elements from the set $\{|c_j(f)|\}$ of infinitely many elements. It is really not clear how we can do this computationally.

Consider the general problem of approximating $f \in F$ by an arbitrary expression of the form

$$g = \sum_{k=1}^{n} c_{j_k}(f)\psi_{j_k},$$

where $\{\psi_j\}_j$ is a given sequence of basis functions, and the coefficients c_{j_k} and the indices j_k may depend on f. A good n-term approximation g might be difficult if not impossible to compute, but if it is available then it can easily be transmitted and evaluated. Nonlinear approximation is strongly related to Bernstein and manifold widths, see DeVore [39], DeVore, Howard and Micchelli [40], DeVore, Kyriazis, Leviatan and Tikhomirov [41], Dung and Thanh [52], Dung [53], Oswald [183], Temlyakov [233], as well as [36], [37], [38] for many results and further references.

One can *assume* that a best n-term approximation (with respect to a given basis or with respect to an optimally chosen basis) is given and ask what kind of algorithms and error bounds can be proved under this assumption. The papers cited contain many results in this direction. Moreover many papers on adaptive wavelet algorithms or finite element algorithms assume this kind of information, see, e.g., the papers by Cohen, Dahmen and DeVore [31], Dahlke, Dahmen and DeVore [35] and Stevenson [225]. We will not further pursue this area.

4.1.2 Algorithms and Their Cost

We already said that all approximations of $S: F \to G$ are of the form $A(f) = \varphi(N(f))$, and $\varphi: N(F) \to G$ for some mapping φ that transforms the information $N(f)$ to an element of the target space G. The information N is built from admissible information given by linear functionals $L_i \in \Lambda$. An adaptive information operator N is of the form (4.9). Hence, in general, the functional L_i has to be determined such that $L_i(f)$ can be *computed*. In addition, also $A(f) = \varphi(y_1, y_2, \ldots, y_{n(f)})$ has to be computed. What kind of mappings or computations

$$(y_1, y_2, \ldots, y_{i-1}) \mapsto L_i \quad \text{and} \quad y \mapsto \varphi(y) \tag{4.12}$$

do we allow? Usually we have two different strategies depending on whether we deal with lower or upper bounds, respectively.

For lower bounds, we usually are very generous and allow *all* mappings of the form (4.12).[1] This makes our lower bounds stronger, since they are valid for an (unrealistically) large class of *idealized algorithms*. The reader may ask, however, whether one can prove reasonable lower bounds using only such a general assumption on the mappings. Surprisingly, the answer is often (not always) yes, since N is partial. We will see many examples where such lower bounds are quite close to upper bounds that are proved under much stronger conditions on the mappings in (4.12).

We now turn to the assumptions for the upper bounds. In the IBC model of computation, we assume that we can work with elements g_i of the space G and perform basic operations on them. That is, we can compute linear combinations

$$g = \sum_{k=1}^{m} c_k g_{i_k}$$

for $c_k \in \mathbb{R}$ and $g_{i_k} \in G$. More formally, we assume the real number model, which will be discussed later, and an *output* of the form

$$\text{out}(f) = [i_1, c_1, i_2, c_2, \ldots, i_m, c_m], \tag{4.13}$$

with $i_k \in \mathbb{N}$ and $c_k \in \mathbb{R}$, is identified with the element g of the target space G.

We comment on the choice of the sequence $\{g_i\}$. Usually, we are free to choose the sequence $\{g_i\}$. That is, we choose $\{g_i\}$ depending on the global parameters of the problem (such as S, F, G and the error parameter ε) but independently of f. It is desirable to choose $\{g_i\}$ such that the number m in $\text{out}(f)$ is minimized and the coefficients c_k and the indices i_k are easy to compute for all $f \in F$. This case is usually studied in IBC and tight complexity bounds are obtained for many problems.

Sometimes, we are *not* free to choose the g_i's. The sequence $\{g_i\}$ is given a priori as part of the formulation of the problem. For example, $\{g_i\}$ can be given as a sequence of B-splines and we want to use it to approximate solutions of partial differential

---

[1]Sometimes it is useful to assume that these mappings are measurable. Observe, however, that this is a very weak assumption.

equations. Furthermore, we may want to use the same sequence $\{g_i\}$ for different problems, i.e., for different operators S and/or different domains F. We do not discuss this second case further; we refer the reader to our paper [175].

The standard model of computation used in numerical analysis and scientific computing is the real number model. This is also the underlying model of this book. For a formal description see Blum, Shub and Smale [13], Blum, Cucker, Shub and Smale [14], Meer and Michaux [143], and [164], [174], [175], [176]. Here we only mention some properties of this model.

We assume that the reader is familiar with the concept of an algorithm over a ring as presented by Blum, Shub and Smale in [13], and Blum, Cucker, Shub and Smale in [14]. We sketch how computation is performed by such algorithms and we restrict ourselves to algorithms over the reals. Input and output consist of finitely many real numbers. We have arithmetic instructions, a jump instruction, and several copy instructions. We now describe these instructions. For simplicity, we restrict ourselves to real numbers, although it is obvious how to generalize everything to complex numbers.

The standard arithmetic operations are the following: addition of a constant, addition of two numbers, multiplication by a constant, multiplication of two numbers, and division of two numbers. Division by 0 is equivalent to a non-terminating computation. Other instructions, such as $x \mapsto \ln(x)$ or $x \mapsto \lfloor x \rfloor$ are also often allowed.

The algorithm is able to perform backward and forward jumps in the list of instructions. We also have the usual copy instructions including indirect addressing, see also [164]. It is clear that many problems of computational mathematics are computable by such algorithms. Examples include problems that are determined by finitely many parameters and whose solutions may be obtained by performing a finite number of arithmetic operations and comparisons. This holds for problems involving polynomials and matrices.

To deal also with problems that are defined for general spaces of functions, we need to have an *information* operation. Typically a *black box* computes $f(x)$ for any x or $L(f)$ for $L \in \Lambda$. This black box (or *oracle*) is an additional computational device. Observe that, in this case, the information about the function f is restricted to $f(x_i)$ or $L_i(f)$ for finitely many i. As already discussed, this information does not generally determine the function f uniquely; it is *partial*.

The cost of computation can be simply defined by counting how many operations of various types are performed. In computational mathematics, one usually counts the number of arithmetic operations, the number of comparisons, and the number of information operations (oracle calls). The costs of input, output as well as copy instructions are usually neglected. For simplicity, we also assume that input, and copy instructions cost 0, although it is obvious how to proceed without this assumption.

Assume that $\mathbf{c}_{ari}$ denotes the cost of performing one arithmetic operation, whereas $\mathbf{c}_{top}$ denotes the cost of one comparison, i.e., the cost of performing one comparison of two real numbers. Comparisons are related to topological complexity and that is why we use the notation $\mathbf{c}_{top}$ to denote the cost of such an operation, see Hertling [92], Smale [220], Vassiliev [252], and [174]. We assume that $\mathbf{c}_{info}$ is the cost of one information

operation. Suppose that the algorithm terminates on input $f \in F$. Let the computation require n_{ari} arithmetic operations, n_{top} comparisons, and n_{info} information operations. The cost of computing the output is

$$\text{cost}(\varphi, N, f) = n_{\text{ari}}\mathbf{c}_{\text{ari}} + n_{\text{top}}\mathbf{c}_{\text{top}} + n_{\text{info}}\mathbf{c}_{\text{info}}. \qquad (4.14)$$

Sometimes we call $n_{\text{info}}\mathbf{c}_{\text{info}}$ the *information cost* and $n_{\text{ari}}\mathbf{c}_{\text{ari}} + n_{\text{top}}\mathbf{c}_{\text{top}}$ the *combinatory or arithmetic cost*. The number $n_{\text{top}}\mathbf{c}_{\text{top}}$ is called *topological cost*, it is related to the *topological complexity* of a problem.

For simplicity, we assume that the cost of all arithmetic operations is the same. (Of course, this could be easily modified.) On the other hand, it is usually much more expensive to compute an information operation than any other operation. For some practical problems, computation of $f(x)$ or $L(f)$ may require billions of arithmetic operations and comparisons. That is, it can take hours even on a modern computer, whereas arithmetic operations or comparisons can be done in a tiny fraction of a second. That is why in many cases the cost is mainly given by the information cost. It is also the reason why we use different parameters for the cost of permissible operations. In general, we want to see how arithmetic, comparison, and information operations affect the complexity of a problem. The global cost of an algorithm $\varphi \circ N$ over F in the worst case setting is defined as

$$\text{cost}(\varphi, N, F) = \sup_{f \in F} \text{cost}(\varphi, N, f). \qquad (4.15)$$

Here we identify, for simplicity, the mapping $\varphi \circ N$ with the simplest algorithm for its computation, i.e., we first compute the information $N(f)$ and then apply φ to obtain $\varphi(N(f)) = (\varphi \circ N)(f)$.

4.1.3 Why Do We Use the Real Number Model?

The purpose of this section is to explain in a rather informal way why information-based complexity uses the real number model as its model of computation. We also explain why the results in the real number model are practically important for many, but not for all, computational problems.

First of all, we stress that the real number model is used in many areas of computation. It is almost universally used in scientific computing and numerical analysis. This model had already been used in algebraic complexity in the fifties in the work of Ostrowski [182] and Pan [185] for polynomial evaluations. It was also used in the work of Coppersmith and Winograd [32], Pan [186], Strassen [226], and others for the famous problem of matrix multiplication. For more general algebraic problems, the reader is referred to the books of Bini and Pan [12], and Bürgisser, Claussen and Shokrollahi [23]. The real number model is also used for computational geometry problems, see e.g., the book of Preparata and Shamos [199], and, in particular Jaromczyk and Wasilkowski [103], [104] for computation of convex hulls.

The formal definition of the real number model can be found in the paper of Blum, Shub and Smale [13], which had a great impact on the further study of this model and which presented the first NP-complete problem over the reals. For many IBC problems, as well as many problems in scientific computing and numerical analysis, we need to use the real number model with "oracles." The oracles are usually defined as subroutines (oracles, black boxes) which compute function values at a given point or the value of a more general linear functional. The formal extension of the real number model to include oracles can be found in Plaskota [195] and in [164]. More about the real number model can be found in Meer and Michaux [143], Weihrauch [274] and in [174], [175], [176], [290]. Pros and cons of the real number model versus the Turing number model can be found in Traub [239].

Before we go on, we pause for a moment and ask why so many people who solve scientific computing or numerical problems are using the real number model, and why there are so few complaints about this model. In fact, with a little exaggeration, one can say that only some theoreticians are unhappy with the real number model, whereas many practitioners do not experience any problem with it.

To be fair, we admit that today's (digital) computers do *not* use the real number model. In fact, we believe that tomorrow's computers won't use the real number model no matter how much progress we may observe in the future computer technology. Why? The reason is simple. In the real number model, we assume that we can store and perform arithmetic operations on real numbers, and that each such operation costs one unit. Since even one single real number may require an infinite number of bits to store, it is really hard to believe that this can be done by a physical device. Take for example the number π. In the real number model, we assume that any number can be a built-in parameter; of course, this also includes π as a special case. How can we really do this if we do not even know all the bits of π? In fact, the computation of the first n bits of π for large n is a very challenging problem, and it is studied in the bit model, see for instance the work of Brent [19], [20] and Salamin [206]. More information can be found in the book of J. M. Borwein and P. B. Borwein [17].

It is now natural to ask what is really used by today's computers. Almost universally today's computers use *floating point arithmetic* for scientific computation. In many cases, it is floating point arithmetic with *fixed* (*single*) precision. Sometimes double precision is used for part of the computation, and in rare cases varying precision is recommended. For the remainder of this section we assume that we are computing with fixed precision floating point arithmetic.

Most practitioners of scientific computing do not experience much difference between the results obtained in the real number model and floating point arithmetic (with fixed precision). Why is this the case? The reason may be that the real number model and floating point arithmetic are closely related. Indeed, they are. Of course, as always with mathematical theories, this is true modulo a couple of assumptions. These assumptions hold for many (but *not* for all) algorithms used in floating point arithmetic, and for many (but *not* for all) practical computational problems; this explains why the difference between the real number model and floating point arithmetic is rarely observed.

What are these assumptions that make the real number model practically indistinguishable from floating point arithmetic? They are:

- *The stability assumption*: stable algorithms exist and we use only *stable* algorithms, and

- *The error demand assumption*: the approximation error is larger than the product of the condition number, the roundoff unit of floating point arithmetic, and the accumulation constant of a stable algorithm.

We now discuss these assumptions in turn. Stability of an algorithm means that the computed result in floating point arithmetic is the exact solution for a slightly perturbed problem. Sometimes this property can be significantly weakened but we do not pursue this in our book, see for instance Kiełbasiński [111]. This is the best possible property since, in general, there is no way to know the exact data of the problem. The data are usually measured or observed with some deterministic or stochastic errors. Furthermore, when we input the data we also perturb them a little. So even assuming that we have an idealized computer with no rounding errors, at best we could only exactly solve the problem with slightly perturbed data. The essence of stability is that we demand this property even in the presence of rounding errors. There is a huge and beautiful, though in many cases tedious, theory of stability initiated by the fundamental work of Wilkinson [281], [282] in the sixties. The current state of art can be found in a recent monograph of Higham [96] with over one thousand references. As the result of this extensive study of many people over many years, we know stable algorithms for many computational problems. Still, plenty of open problems need to be solved and stability is far from being completely understood.

Many stable algorithms are careful implementations of algorithms analyzed over the reals. Sometimes a few things need to be changed, and a few precautions should be taken when we want to make the real number algorithm stable. Still, in most cases, the cost of the stable algorithm is basically the *same* in floating point arithmetic as in the real number model.

The cost is basically the same, since in both the real number model and floating point arithmetic the cost of a single arithmetic operation and comparison *does not* essentially depend on the size of the operands.[2] We only need to replace the unit cost of arithmetic operations and comparisons in the real number model by the actual cost of such operations in floating point arithmetic. The cost of an oracle may be treated similarly. In the real number model we usually assume that the cost of an oracle call is fixed; in floating point arithmetic this corresponds to the cost of one subroutine call that computes, for example, one function value.

Assume that we know an algorithm that enjoys some optimality properties in the real number model. For instance, it is often the case that we know an algorithm whose cost in the real number model is close to the complexity of the problem. If we know a

---

[2]The property of the real number model and floating point arithmetic that the cost of arithmetic operations does not depend on the size of operands is probably the most important difference between these models and the classical Turing machine model (the bit model).

stable implementation of such an algorithm then we have the same cost also in floating point arithmetic. Can we then claim that the complexity is the same in both the real number model and floating point arithmetic? Not yet, since the stable algorithm won't compute the same result as its real number counterpart.

This leads us to the error demand assumption. The first lesson in scientific computing is that there is really no way to compute the exact solution for most practical scientific problems.[3] As we have already mentioned, the reason is that the data of the problem are usually not given exactly. The good thing is, however, that we do *not* need to have the exact solution, and a reasonable approximation to the exact solution is good enough for most practical purposes. There is one more reason why a reasonable approximation is enough, and this reason is related to mathematical modeling. Usually, the computational problem is the solution of the mathematical problem which at best can only approximate the original problem by some modeling, and simplifications like linearization or discretization. So why should we insist on the exact solution of the computational problem when we already accepted the modeling error?

Hence, let us agree that we want to compute an approximation to the exact solution. Let ε be the approximation error that we can tolerate as the difference between the computed and the exact solution.[4] The actual value of ε depends on the particular application. We believe that for many problems ε needs not be very small and, with the proper scaling, ε from the interval $[10^{-8}, 10^{-2}]$ covers numerous applications.[5]

We assume that we know a stable algorithm for our problem. Since this algorithm computes the exact solution of a slightly perturbed problem, we need to estimate the error between the solutions of the exact and perturbed problems. This is measured by the *condition number* $\kappa(P)$ of the problem P, see e.g., Kiełbasiński [111], Wilkinson [282] or [285] for the precise definition. The condition number tells us about the sensitivity of the solution to small changes of the data. If the condition number is not too large, the problem is *well-conditioned*; otherwise, the problem is *ill-conditioned*. There is a deep theory of condition numbers, which is not restricted to scientific computing and numerical analysis. This theory is a form of sensitivity analysis. Nevertheless, for typical computational problems, the study of condition numbers is often part of the stability analysis, and can be found in the work of Wilkinson and many others.

We stress that the concepts of stability and conditioning are *not* related. Stability is a property of the algorithm whereas conditioning is a property of the problem. If the problem is ill-conditioned all stable algorithms may fail, since they will generally

---

[3]Notable exceptions are problems solved by symbolic computations and combinatorial problems. Such problems are not solved in floating point arithmetic.

[4]Hence, we now consider the absolute error. Obviously, it is also reasonable to consider the relative or normalized error, as will be done later.

[5]In many papers, including many IBC papers, results are obtained for the asymptotic case, i.e., for ε tending to 0. Such results are not necessarily practical for relatively large ε, say once more for $\varepsilon \in [10^{-8}, 10^{-2}]$. We believe that more emphasis should be devoted to the non-asymptotic case where ε does not have to be sufficiently small, and error and cost bounds are not presented in terms of $\mathcal{O}(h(\varepsilon))$ or $\Theta(h(\varepsilon))$ for some function h, but have an explicit dependence on ε. Obviously, obtaining such an explicit dependence on ε is much harder than obtaining asymptotic bounds, and requires much more work. Generalized tractability addresses these issues, see Chapter 8.

compute results with large errors. On the other hand, if the problem is well-conditioned then all stable algorithms will compute results with small errors.

To address the error of a stable algorithm more precisely, let ϱ be the roundoff unit of floating point arithmetic. That is, $\varrho = 2^{-t}$, where t is the number of mantissa bits in floating point arithmetic. For modern computers, $\varrho \in [10^{-16}, 10^{-8}]$. Furthermore, there is usually an option to significantly decrease ϱ by the use of double or multiple precision. We can now put the error demand assumption in a more technical way, namely

$$C\varrho(\kappa(P) + 1) \leq \varepsilon. \tag{4.16}$$

Here, C is the accumulation constant of rounding errors of a stable algorithm. Usually, C is a low degree polynomial in the number of inputs. Note that we added 1 to the condition number to have a reasonable lower bound on ε also for very well-conditioned problems for which $\kappa(P) \approx 0$.

Inequality (4.16) relates the approximation error to the quality of a stable algorithm given by its accumulation constant, to the quality of floating point arithmetic given by its roundoff unit and to the measure of sensitivity of the problem given by its condition number. We believe that (4.16) holds for many, but not for all, stable algorithms and computational problems. Indeed, if the problem is relatively well-conditioned, i.e., $\kappa(P)$ is not too large, then (4.16) holds in spades, since usually the allowed error ε is much larger than $C\varrho(\kappa(P) + 1)$. On the other hand, for ill-conditioned problems we may have troubles with (4.16). There is then the option of switching to double precision, which roughly corresponds to replacing ϱ by ϱ^2. This significantly extends the domain of applications for which the assumption (4.16) holds. In fact, (4.16) can be regarded as a guide for selecting an appropriate ϱ when solving a problem with a given condition number by a stable algorithm.

We briefly comment on well and ill-conditioned problems. Some people believe that ill-conditioned problems occur only for artificially-generated cases, and that with proper modeling, we should always end up with a relatively well-conditioned computational problem. We believe that this is *not* always the case, and that ill-conditioning is an inherent part of some practical problems. One notable example is given by ill-posed problems. However, a regularized ill-posed problem may be well or ill-conditioned. In any case, we prefer not to rule out ill-conditioned problems, and defer the discussion what to do if (4.16) does not hold to the concluding part of this section.

We are ready to discuss the error of a stable algorithm. Since a stable algorithm computes the exact solution of a slightly perturbed problem in floating point arithmetic, its error is bounded by $C\varrho\kappa(P)$. Due to (4.16), its error does not exceed ε, and the computed solution is an ε-approximation. Hence, we have solved our problem, although we have used floating point arithmetic instead of the real number model. Furthermore, as indicated earlier, we did so with basically the same cost as in the real number model. If the cost of our algorithm in the real number model is close to the complexity of the problem, we have achieved the same complexity in floating point arithmetic.

This shows that results in the real number model and floating point arithmetic are equivalent, as well as the practical importance of results from the real number model. We stress once more that the equivalence of results in the real number model and floating point arithmetic holds modulo the stability and error demand assumptions.

We now assume that at least one of the stability or error demand assumptions does not hold. Then we may lose the equivalence of results in the real number model and the floating point arithmetic. In this case, the real number model is no longer appropriate for computation. We should then use a different model, possibly yielding different complexity results and different optimal algorithms.

Let us first discuss the stability assumption. Imagine the following situation. For a given computational problem, we find good complexity bounds and optimal algorithms in the real number model. Then we take one of the optimal algorithms and try to find a stable implementation. We stress that for most computational problems, such stable algorithms have been found. Sometimes, however, the notion of stability had to be relaxed.

Still, there is always the possibility that for some problem, *no* optimal algorithm has a stable implementation. That is, there may be an intrinsic trade-off between complexity in the real number model and stability. Such a trade-off was shown by Miller [148] for matrix multiplication. Miller proved that any algorithm that multiplies two $n \times n$ real matrices with cost of order n^β for $\beta < 3$ cannot be stable, whereas the classical algorithm with cost of order n^3 is stable. Hence, the gain in cost must spoil stability for this problem.

A few comments are now in order. We do not want to give too many technical details, but let us only mention that the notion of stability used by Miller is quite strong. The trade-off between complexity and stability for matrix multiplication may disappear if a more relaxed notion of stability is used.

In any case, the real number model is potentially risky since we may be unable to find a (weakly) stable implementation of an optimal algorithm. In this case, the real number model results are useless for practical computations. The existence of such a practical problem has yet to be shown, but today we cannot rule it out.

One could try to resolve the potential lack of stable algorithms by restricting the study of complexity to stable algorithms. This is obviously a good idea. However, the technical difficulty of finding upper and lower complexity bounds for the class of stable algorithms is significantly higher than for the unrestricted class of algorithms. For example, for some classes of functions, we know that optimal quadrature formulas have positive weights, hence they are stable. Unfortunately, there are not many results along these lines.

We now turn to the case when the error demand assumption is not satisfied. Since the product $C\varrho$ is usually very tiny, this means that the condition number $\kappa(P)$ is too large, when compared to the approximation error ε.[6] Hence, the error demand assumption fails for ill-conditioned problems. As we already mentioned, one easy fix is to switch

---

[6]One can equivalently say that ε is too small, compared to the condition number. Since we assume that the choice of ε was indeed reasonable, i.e., not too small, we prefer to say that the condition number is too large.

to double or to multiple precision. Indeed, this is often done in computational practice. Then the cost of an operation depends logarithmically on the precision. If we use only single or double precision, the cost does not change much.

If the error demand assumption fails and fixed precision floating point arithmetic is too weak to solve our problem, we may choose a different model of computation. Then a natural choice is the *bit model*, in which all arithmetic and comparison operations are performed on numbers having a finite number of (binary) bits and in which the cost of all operations depends on the desired accuracy of the output and on the length of the input numbers. Oracles can also be used, but we must now compute the function value $f(x)$ to within a given accuracy δ, the cost being an increasing function of $1/\delta$. This model, usually without oracles, is studied in the work of Ko, Schönhage, Weihrauch and others, see for instance [108], [210], [274], [275]. There is also some IBC work relevant to the bit model, particularly the work of Plaskota, who studies noisy information with the cost depending on the noise level. Plaskota's monograph [195] covers this subject in depth.

In the bit model, we still want to compute an ε-approximation at minimal cost. The complexity study in the bit model is difficult since, in particular, we must determine precision needed for each operation used by an algorithm. Although the bit model may be used for all problems, we believe that its importance can be only seen for ill-conditioned problems, when the relative simplicity of floating point arithmetic is not available. A good model problem that should be also studied in the bit model is zero finding for polynomials or general functions. Clearly, the precision of computation must be increased as we have better approximations to a zero, see for example the paper of Neff and Reif [153].

4.1.4 Errors and Complexity

For each $f \in F$ we want to compute an approximation to $S(f)$. Let $A(f)$ be the computed approximation. The distance between $S(f)$ and $A(f)$ will be measured according to a given error criterion. The most basic error criterion in this book is

$$\text{the absolute error} \qquad \|S(f) - A(f)\|.$$

Examples of other error criteria include

$$\text{the relative error} \qquad \frac{\|S(f) - A(f)\|}{\|S(f)\|} \tag{4.17}$$

and

$$\text{the normalized (by } \|f\|) \text{ error} \qquad \frac{\|S(f) - A(f)\|}{\|f\|}, \tag{4.18}$$

with the interpretation of $0/0 = 0$. We stress that we mainly consider another normalization of the error. We first define the *initial error* of S by

$$e_0^{\text{wor}} := \inf_{g \in G} \sup_{f \in F} \|S(f) - g\|. \tag{4.19}$$

This is the minimal worst case error of a constant algorithm $A(f) = g$ for all $f \in F$. It is called the initial error since we can compute it by knowing only the formulation of the problem, without any information about any particular $f \in F$. For linear problems, where S is linear and F is the unit ball of the space $\widetilde{F}$, the initial error e_0^{wor} is just the norm of S since $g = 0$ is optimal. We believe that properly normalized problems should have the initial error of order 1. As already indicated in Chapter 3, this is *not* the case for some classical problems for which the initial error can be exponentially small or large in d for the d-variate problem.

We define

$$\text{\textit{the normalized (by the initial error) error}} \quad \frac{\|S(f) - A(f)\|}{e_0^{\mathrm{wor}}}. \tag{4.20}$$

For a moment we restrict ourselves to the absolute error criterion and recall that the *worst case error* of A is given by

$$e^{\mathrm{wor}}(A) = \sup_{f \in F} \|S(f) - A(f)\|. \tag{4.21}$$

The *normalized (by the initial error)* error of A is defined by $e^{\mathrm{wor}}(A)/e_0^{\mathrm{wor}}$. Often we study *normalized problems* with $e_0^{\mathrm{wor}} = 1$, for which there is no difference between these two error criteria.

For the absolute error, we want to find the smallest n for which the error is at most ε and define

$$n^{\mathrm{wor}}(\varepsilon, S, F) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{wor}}(A_n) \leq \varepsilon\}. \tag{4.22}$$

Here A_n is any algorithm of the form $A_n = \varphi \circ N$, where N uses at most n admissible information functionals from Λ. If we want to stress the role of Λ, then we write $n^{\mathrm{wor}}(\varepsilon, S, F, \Lambda)$.

For the normalized (by the initial error) error we use instead the definition

$$n^{\mathrm{wor}}(\varepsilon, S, F) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{wor}}(A_n) \leq \varepsilon \cdot e_0^{\mathrm{wor}}\}. \tag{4.23}$$

It should be always clear from the context whether we mean the absolute error, the normalized error, or another type of error. Sometimes it is more convenient to consider the numbers

$$e^{\mathrm{wor}}(n) = e^{\mathrm{wor}}(n, S, F) = \inf_{A_n} e^{\mathrm{wor}}(A_n) \tag{4.24}$$

or their normalized counterparts $e^{\mathrm{wor}}(n)/e_0^{\mathrm{wor}}$. Again, if we want to stress the role of Λ we write $e^{\mathrm{wor}}(n, \Lambda)$. Obviously, $n^{\mathrm{wor}}(\varepsilon, S, F)$ and $e^{\mathrm{wor}}(n)$ are inversely related.

The numbers $n = n^{\mathrm{wor}}(\varepsilon, S, F, \Lambda)$ describe the *information complexity* of the problem. That is, to solve the problem to within an error of ε, we need n information operations.

The *total complexity*

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, S, F, \Lambda) = \inf_{\substack{A = \varphi \circ N \\ e^{\mathrm{wor}}(A) \leq \varepsilon}} \sup_{f \in F} \mathrm{cost}(\varphi, N, f) \tag{4.25}$$

describes the minimal total cost (including not only the cost of the information but also the combinatory cost) needed to solve the problem to within ε.

As already indicated, for many problems the major part of the cost is the information cost, so that

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, S, F, \Lambda) \text{ is almost the same as } \mathbf{c}_{\mathrm{info}} n^{\mathrm{wor}}(\varepsilon, S, F, \Lambda). \qquad (4.26)$$

Hence, one gets sharp bounds on the total complexity by studying only the information complexity. Therefore in this book we mainly study the minimal numbers of information operations needed to solve the problem to within ε. In the worst case setting, this means that we study $n^{\mathrm{wor}}(\varepsilon, S, F, \Lambda)$.

We briefly mention why the name Information-Based Complexity is used. For problems satisfying (4.26), the total complexity is approximately equal to the information complexity. For other problems, the information complexity is a lower bound on the total complexity. In any case, for all problems we need to study the information complexity and this is usually the first step to establish bounds on the total complexity. To stress this part of analysis the field is called *information-based complexity*.

We end this section by a brief discussion of proof techniques for getting *lower bounds* on the complexity. Various lower bounds can be obtained depending on which part of algorithms and their cost is studied:

a) information cost and the notion of the radius of information are used to obtain lower bounds on the number of information operations needed;

b) combinatoric and/or arithmetic considerations are used to obtain lower bounds on the number of arithmetic operations;

c) topological considerations are used to obtain lower bounds on the number of branchings; the reader may consult the paper of Smale [220] where the concept of topological complexity was introduced, as well as the papers of Hertling [92], Vassiliev [252] and [174] for specific results concerning topological complexity.

Which lower bound is best depends on the particular problem. In general we should, of course, try to estimate the (suitably weighted) combination of all these costs to obtain good lower bounds for the total complexity. For problems studied in this book, (4.26) usually holds and that is why it is enough to have good estimates on the information complexity.

4.1.5 Information Complexity and Total Complexity

As we have already mentioned, the total complexity for many IBC problems is roughly equal to their information complexity. This is the case for most *linear* problems and for some *nonlinear* problems for which an almost optimal error algorithm can be implemented at cost proportional to the information complexity. There are, however, some counterexamples. That is, we know a few problems for which the information complexity is significantly smaller than the total complexity. There even exists a linear problem with finite, and even reasonably small information complexity, and

with infinite combinatory complexity, see [268]. This problem is a slight modification of the linear problem studied in [279] for which the worst case error of any linear algorithm is infinite whereas the worst case error of some nonlinear algorithms is finite and may be sufficiently small if proper information is used. There is also a result of Chu [28] which presents a problem whose total complexity is any increasing function of its information complexity. Hence, in full generality, the information and total complexities may be quite different. We must, however, admit that the problems cited in this subsection are quite artificial, and we do not know if there exists a practically important problem for which its combinatory complexity is provably much larger than its information complexity.

There is also a very interesting result of Papadimitriou and Tsitsiklis [187] who studied a nonlinear problem of decentralized control theory for the class of Lipschitz functions of four variables. Assuming that only function values can be computed, they proved that the information complexity is of order ε^{-4}. Using the Turing machine model of computation and assuming that the famous conjecture P $\neq$ NP holds, they got a surprising result that the total complexity cannot be polynomial in ε^{-1}. That is, it is very likely that the total number of bit operations needed to solve this problem to within ε is not polynomial in ε^{-1}, and therefore it is much larger than the information complexity. It is open what is the total complexity of this problem if the real number model of computation is used.

4.2 Basic Results for Linear Problems in the Worst Case Setting

We define *linear problems* by the following assumptions:

- $S \colon \widetilde{F} \to G$ is a *linear* operator with $\widetilde{F}$ and G being normed spaces;

- $F \subseteq \widetilde{F}$ is *non-empty*, *convex*, i.e., if $f_1, f_2 \in F$ then $t f_1 + (1 - t) f_2 \in F$ for all $t \in [0, 1]$, and *symmetric* (*balanced*), i.e., if $f \in F$ then also $-f \in F$;

- $\Lambda \subseteq \Lambda^{\mathrm{all}}$ is a class of *linear* functionals.

The reader may ask what convexity and symmetry of F have to do with linearity of the problem. It turns out, see [244, p. 32], that these two properties of being convex and symmetric[7] are equivalent to the existence of a *linear* operator $T \colon \widetilde{F} \to X$ for some normed space X such that

$$\{f \in \widetilde{F} \mid \|Tf\| < 1\} \subseteq F \subseteq \{f \in \widetilde{F} \mid \|Tf\| \le 1\}.$$

This means that, modulo the boundary of the set $\{f \in \widetilde{F} \mid \|Tf\| \le 1\}$, the set F is generated by a linear operator T. Furthermore, in all estimates it does not matter if we

---

[7]Formally we also need to assume that F is absorbing, i.e., for every $f \in \widetilde{F}$ there exists a positive c such that $cf \in F$. However, this property can be always guaranteed if we take $\widetilde{F} = \mathrm{span}(F)$.

take or do not take the boundary of the set $\{f \in \widetilde{F} \mid \|Tf\| \leq 1\}$. Hence, linearity of the problem is characterized by linearity of S and T and linearity of functionals from Λ. We stress that continuity of linear operators is *not* (yet) assumed.

We mainly restrict ourselves to the absolute error criterion with the worst case error given by

$$e^{\text{wor}}(A) = \sup_{f \in F} \|S(f) - A(f)\|. \tag{4.27}$$

Some of the following results are based on a useful formula for the diameter of non-adaptive information.

Lemma 4.3. *Consider a linear problem and linear non-adaptive information*

$$N^{\text{non}}(f) = [L_1(f), L_2(f), \ldots, L_n(f)].$$

Then

$$d(N^{\text{non}}) = d(N^{\text{non}}, 0) = 2 \sup_{h \in F, N^{\text{non}}(h)=0} \|S(h)\|. \tag{4.28}$$

Proof. Let $y \in N^{\text{non}}(F)$ and assume that $f_1, f_2 \in F$ with $N^{\text{non}}(f_1) = N^{\text{non}}(f_2) = y$. Then linearity of S and N^{non} yields $\|S(f_1) - S(f_2)\| = \|S(f)\|$, where $f = f_1 - f_2$ and $N^{\text{non}}(f) = 0$. Since F is symmetric, $-f_2 \in F$ and since F is convex $h := \frac{1}{2}f = \frac{1}{2}(f_1 - f_2) \in F$ and $N^{\text{non}}(h) = 0$. Thus,

$$\|S(f_1) - S(f_2)\| = 2\|S(h)\| \leq a := 2 \sup_{h \in F, N^{\text{non}}(h)=0} \|S(h)\|.$$

Hence, $d(N^{\text{non}}, y) \leq d(N^{\text{non}}) \leq a$. To prove that $d(N^{\text{non}}, 0) \geq a$, it is enough to take $f_1 = -f_2 = h \in F$ with $N^{\text{non}}(f_i) = 0$ to obtain

$$d(N^{\text{non}}, 0) \geq \|S(f_1) - S(f_2)\| = 2\|S(h)\|.$$

Taking the supremum with respect to h we get $d(N^{\text{non}}) \geq d(N^{\text{non}}, 0) \geq a$. $\square$

Lemma 4.3 states that for linear problems and non-adaptive information, as far as the diameter of information is concerned, the least informative data in the worst case setting is given by 0. This holds independently of the form of S, F and N^{non}. Furthermore, the lemma says that the diameter of information, which is (modulo a factor of two) a sharp lower bound on the worst case error of algorithms that use N^{non}, is given by twice the largest solution for problem elements with zero information. This means that zero information is (again modulo a factor of two) the least informative data for linear problems. Many information and complexity bounds have been obtained by using this lemma.

4.2.1 On the Power of Adaption

One of the more controversial issues in numerical analysis concerns adaption. By *adaptive/non-adaptive* algorithms we mean algorithms that use adaptive/non-adaptive

information. The use of adaptive algorithms is widespread and many people believe that well-chosen adaptive algorithms are much better than non-adaptive ones in most situations.[8]

Here we survey what is known theoretically on the power of adaption mostly for linear problems, see also [165], [242]. We will present some results which state that under natural assumptions adaptive algorithms are *not* much better than non-adaptive ones. There are also other results, however, saying that adaptive algorithms can be significantly superior to non-adaptive ones for some nonlinear problems. As we will see, the power of adaption critically depends on our a priori knowledge concerning the problem being studied; even a seemingly small change in the assumptions can lead to a different answer.

Let us begin with some well-known examples. The Gauss formula for numerical integration of univariate functions is non-adaptive since for all functions we compute function values at the same points. The bisection algorithm and the Newton algorithm for zero finding of a univariate function are adaptive, since for different functions we use different information operations. For some nonlinear problems, adaption can be exponentially better than non-adaption. For example, consider the zero finding problem for the class

$$F = \{f : [0, 1] \to \mathbb{R} \mid f \text{ is continuous and } f(0) < 0, \ f(1) > 0\}.$$

Then the minimal worst case error of algorithms using non-adaptive information of cardinality n is $1/(2n + 2)$ whereas the minimal worst case error of algorithms using adaptive information of cardinality n is $1/2^{n+1}$, see [244, p. 166–170]. Hence, adaption is indeed exponentially better than non-adaption for this problem. Adaption for other nonlinear problems, such as optimization, is surveyed in [165].

We consider algorithms of the form $A(f) = \varphi(N(f))$ for some mapping φ that transforms the information N to the target space G and N is of the form (4.9) and built from admissible information operations $L_i \in \Lambda$.

One might hope that it is possible to learn about f during the computation of $L_1(f), L_2(f, y_1), \ldots, L_{k-1}(f, y_1, y_2, \ldots, y_{k-2})$ in such a way that one can choose the next functional L_k to reduce the error more efficiently. Therefore one studies *adaptive* information, where the choice of L_k may depend on the previously computed values $L_1(f), L_2(f, y_1), \ldots, L_{k-1}(f, y_1, y_2, \ldots, y_{k-2})$. For instance, in the case of function values, $L_k(f, y_1, y_2, \ldots, y_{k-1}) = f(x_k)$ and the point x_k depends on the known function values via

$$x_k = \psi_k(f(x_1), f(x_2), \ldots, f(x_{k-1})),$$

where ψ_k may be an arbitrary function of $(k - 1)$ variables. In mathematical statistics, adaptive information is known as sequential design and non-adaptive information is known as non-sequential design.

---

[8]Unfortunately, different authors use the same word "adaption" for different things. What we discuss here are algorithms that use adaptive information, see Section 4.1.1. In the numerical analysis community also other numerical schemes are called adaptive, for example, if non-uniform meshes or nonlinear information (such as a best *n*-term approximation) are involved. For a discussion see also [37] and Remark 4.6.

Consider adaptive information N^{ada} of the form (4.9) defined on the set F. Since F is non-empty, convex and balanced, the zero element $f = 0$ belongs to F. We define the *non-adaptive* information N^{non} by

$$N^{\text{non}}(f) = [L_1(f), L_2(f, 0), \ldots, L_{n(0)}(f, 0, \ldots, 0)]. \tag{4.29}$$

Hence, N^{non} has cardinality $n = n(0)$, i.e., the same number of functionals are used as N^{ada} uses for the zero element $f = 0$. Moreover, we also use the same (fixed) functionals $L_1, L_2, \ldots, L_n$ for N^{non} as N^{ada} uses for $f = 0$.

The first general result about adaption is due to Bakhvalov [8], who assumed that S is a linear functional and the L_k are special linear functionals, for instance given by function values, $L_k(f) = f(x_k)$. In 1971 Bakhvalov proved that then adaption does *not* help.

The adaption problem is a little more complicated if we consider arbitrary linear operators S instead of linear functionals. It has been known since 1980 that non-adaptive algorithms are optimal up to a factor 2, as proved by Gal and Micchelli in [60] and in [244].

Theorem 4.4. *Consider a linear problem. For any adaptive information N^{ada} construct non-adaptive information N^{non} by (4.29). Then*

$$d(N^{\text{non}}) \leq d(N^{\text{ada}}) \quad \text{and} \quad r(N^{\text{non}}) \leq 2r(N^{\text{ada}}).$$

Proof. Lemma 4.3 yields

$$d(N^{\text{non}}) = d(N^{\text{non}}, 0) = d(N^{\text{ada}}, 0) \leq d(N^{\text{ada}}),$$

whereas (4.7) gives

$$r(N^{\text{non}}) \leq d(N^{\text{non}}) \leq d(N^{\text{ada}}) \leq 2r(N^{\text{ada}}). \qquad \square$$

Hence adaption can be better than non-adaption by a factor of at most 2. There are examples where adaption is (slightly) better than non-adaption, see [113], where it was shown that for some linear problem, adaption is 1.074 times better than non-adaption. Hence, the factor 2 in Theorem 4.4 cannot be replaced by 1 in general. But how much adaption can help for linear problems is the subject of our next open problem.

Open Problem 16.

- Determine the smallest number a for which

$$\inf_{N^{\text{non}}:\, \text{card}(N^{\text{non}}) \leq n} r(N^{\text{non}}) \leq a \inf_{N^{\text{ada}}:\, \text{card}(N^{\text{ada}}) \leq n} r(N^{\text{ada}})$$

  holds for arbitrary linear problem. We know that $a \in [1.074, 2]$.

In many cases, however, the factor 2 in Theorem 4.4 is not needed. This holds if the radius of a set is equal to the half of its diameter. The last property holds for

many spaces and then we actually have the sharper inequality $r(N^{\mathrm{non}}) \le r(N^{\mathrm{ada}})$. We already mentioned that this is the case for linear functionals, i.e., when the target space G is $\mathbb{R}$. More general conditions are given by the next result, with the following notation.

By $B(K)$ we mean the set of bounded functions on a set K with the norm $\|f\| = \sup_{x \in K} |f(x)|$, and by $C(K)$ we mean the set of continuous functions on a compact Hausdorff space K with the same norm. By $L_\infty(\mu)$ we denote the L_∞ space with a measure μ. For a proof of all cases of the next result, we refer to Creutzig and Wojtaszczyk [34].

Theorem 4.5. *Consider a linear problem. Assume additionally that*

- $G \in \{\mathbb{R}, L_\infty(\mu), B(K)\}$, *or*

- $\widetilde{F}$ *is a pre-Hilbert space, or*

- $S: \widetilde{F} \to G = C(K)$, *where S is compact.*

For any adaptive information N^{ada} construct non-adaptive information N^{non} by (4.29). *Then*

$$r(N^{\mathrm{non}}) \le r(N^{\mathrm{ada}}).$$

Proof. The proof is based on the fact that under the given assumptions, instead of (4.7) we have the equality $r(N^{\mathrm{non}}) = \frac{1}{2}d(N^{\mathrm{non}})$. $\qquad\square$

To understand intuitively Theorems 4.4 and 4.5 one might say that for zero information we do not have any chance to adjust the choice of the next functional L_k to decrease the error. The reader may want to use this argument for the integration example $S(f) = \int_0^1 f(x)\,\mathrm{d}x$ for Lipschitz functions,

$$F = \{f \in C([0,1]) \mid |f(x) - f(y)| \le |x - y|\}.$$

It turns out that the midpoint rule A_n,

$$A_n(f) = \frac{1}{n} \sum_{j=1}^{n} f\left(\frac{2j-1}{2n}\right)$$

has the minimal worst case error in the class of all adaptive algorithms using at most n function values for this particular class of functions, and

$$e^{\mathrm{wor}}(A_n) = \frac{1}{4n}.$$

Remark 4.6. a) The results so far have not shown any significant superiority of adaptive algorithms for linear problems. Nevertheless adaptive algorithms are often used

in practice. An important application of adaptive algorithms is computation with finite elements and wavelets, see Babuška and Rheinboldt [6], Cohen, Dahmen and DeVore [31], Dahlke, Dahmen and DeVore [35], Eriksson, Estep, Hansbo and Johnson [57] and Stevenson [225]. Remember, however, that the same word adaption is used with different meanings. A thorough discussion of the above results and their application to the solution of boundary value problems for elliptic partial differential equations is given in the book of Werschulz [277], see also [37].

b) Of course, it should be stressed that these results on adaption assume that we have a linear operator S and a convex and symmetric set F. The set F reflects the *a priori knowledge* concerning the problem; often it is known that f has a certain smoothness and this knowledge may be expressed by choosing an appropriate F. If our a priori knowledge about the problem leads to a set F that is either non-symmetric or non-convex (or both) then we certainly cannot apply Theorems 4.4 and 4.5 and it is possible that adaption is significantly better. In fact, adaption can help significantly for problems for which only one of the three linearity assumptions is violated and the other two still hold. Examples of such problems can be found in [242, p. 57–63]. In this sense, the assumptions of Theorems 4.4 and 4.5 are essential.

c) The idea behind Theorems 4.4 and 4.5 is that non-adaptive information that is good for the zero function $0 \in F$ is also good for any other $f \in F$. This is true for *any linear problem with any norm*. However, this result does *not* automatically lead to good non-adaptive algorithms. In particular, we do not claim that the optimal non-adaptive functionals L_i somehow correspond to a uniform mesh or grid. There are important examples where regular grid points are very bad and the optimal (non-adaptive) points are more complicated. We stress this fact because we noticed that some authors compare poor non-adaptive algorithms based, for example, on a regular grid with sophisticated adaptive algorithms and (wrongly) conclude that adaptive algorithms are superior.

We illustrate this point by an example related to the classical star-discrepancy, see Section 3.1.5 of Chapter 3 for the definition. For $k, d \in \mathbb{N}$, consider the regular grid

$$G_k^d = \left\{ x = [x_1, x_2, \ldots, x_d] \mid x_i \in \left\{ \tfrac{2j-1}{2k} \mid j = 1, 2, \ldots, k \right\} \right\}.$$

The star-discrepancy of this set is

$$\mathrm{disc}_\infty^*(G_k^d) = 1 - \left(\frac{k-1}{k} \right)^d,$$

see the book of Niederreiter [158]. Putting $k = d$ we obtain a set with d^d points and

$$\lim_{d \to \infty} \mathrm{disc}_\infty^*(G_d^d) = 1 - \frac{1}{e} \approx 0.63212.$$

This is very bad behavior showing that grid points are indeed a very poor choice. As we know from Section 3.1.5 of Chapter 3, much better (still non-adaptive) points exist since the star-discrepancy of d^d optimal points is $\mathcal{O}(d^{-(d-1)/2})$ with the factor in the big $\mathcal{O}$ notation independent of d.

d) So far, we studied the worst case error on symmetric convex sets F. This means that $f \in F$ is used as our a priori knowledge about the specific problem. One wants to use as much a priori information as possible. Often, one obtains a class F that is convex, but not symmetric. This is the case, for example, if we know that f or $S(f)$ is positive or monotone or convex. Some problems, such as ill-posed problems, can only be solved in a reasonable way if we use this extra information. Non-symmetric classes and their widths are not so often studied in approximation theory, see, however, the recent work of Gilewicz, Konovalov and Leviatan [63].   □

Hence, it is interesting to study how much adaption can help for linear problems defined on a convex but non-symmetric set F. For examples and results see [165]. The first result is due to Kiefer [110] and concerns the problem of numerical integration of monotone functions. For the problem of Kiefer, as well as for some other problems, adaption again does *not* help. For other problems adaption helps, in some cases quite dramatically. Such problems were studied by Korneichuk [114], Rote [201] and Sonnevend [223], see also [162]. Here we only give two examples and present an open problem.

Recovery Problem of Korneichuk [114]. Consider the recovery (approximation) problem $S = \mathrm{id} \colon F \to L_\infty([0, 1])$ for the class

$$F = \{f \colon [0, 1] \to [0, 1] \mid f \text{ monotone and } |f(x) - f(y)| \le |x - y|^\alpha\},$$

where $0 < \alpha < 1$ with the class Λ^{std}, i.e., we allow only function values, $L_i(f) = f(x_i)$. Then the minimal worst case error of non-adaptive algorithms that use n function values is $e_n^{\mathrm{wor}} = \Theta(n^{-\alpha})$, while the worst case error of an adaptive (bisection-like) algorithm is of order $n^{-1} \ln n$. Hence, especially for small α, adaption is much more powerful than non-adaption.

Power of Adaption for Convex Sets, see [163]. Assume that $S \colon \widetilde{F} \to G$ is a continuous linear operator and $F \subseteq \widetilde{F}$ is convex. We compare non-adaptive algorithms A_n^{non} and adaptive algorithms A_n^{ada} based on n information operations from Λ^{all}. Then

$$\inf_{A_n^{\mathrm{non}}} e^{\mathrm{wor}}(A_n^{\mathrm{non}}) \le 4(n + 1)^2 \inf_{A_n^{\mathrm{ada}}} e^{\mathrm{wor}}(A_n^{\mathrm{ada}}). \tag{4.30}$$

Hence, the error of adaptive algorithms can be at most $4(n + 1)^2$ times better than the error of non-adaptive algorithms.

Open Problem 17.

- We know that (4.30) holds for the class Λ^{all}. Is the same result also true for subclasses of Λ^{all} such as Λ^{std}?

- Verify if the factor $4(n+1)^2$ in (4.30) can be decreased to a factor linear in n. Note that a further decrease to $n^{-\beta}$ with $\beta < 1$ is impossible due to Korneichuk's result. This problem is related to estimates between Bernstein and Kolmogorov widths, see [163]. These widths are, however, usually studied only in the symmetric case.

It is also of interest to study the power of adaption for problems defined on a non-convex set F. Examples are classes of functions with certain singularities, see Huerta [101], Wasilkowski and Gao [265], and Werschulz [278]. For non-convex but symmetric sets F, the advantage of adaptive algorithms can be very large. This follows from Plaskota and Wasilkowski [196] who studied univariate integration of uniformly bounded r smooth functions with at most one point of singularity. They proved that the nth minimal error of non-adaptive information is of order n^{-1}, and of adaptive information is of order n^{-r}. Since r can be arbitrarily large, adaption can be arbitrarily better than non-adaption. It is interesting to add that if we have the class of uniformly bounded r smooth functions with two or more singularity points then non-adaptive and adaptive information are more or less of the same power since both the nth minimal errors are of order n^{-1}. However, adaption regains its power asymptotically even if we permit functions with arbitrarily (but finitely) many singular points.

Similar results are obtained for univariate approximation defined on more or less the same class of functions by Plaskota, Wasilkowski and Zhao [198]. The approximation error is measured in the L_p-norm for $p \in [1, \infty)$. Then for the class of functions with at most one singularity point, the nth minimal error of non-adaptive information is at best of order $n^{-1/p}$ and of adaptive information is of order n^{-r}. The authors also advocate that the L_∞-norm is not appropriate for studying singular functions, and that the *Skorohod* metric should be used instead. The notion of the Skorohod metric can be found also in, e.g., the books of Billingsley [11] and Partharasarathy [189].

4.2.2 Linear and Nonlinear Algorithms for Linear Problems

Smolyak's result states that linear algorithms are optimal in the class of all non-adaptive algorithms for linear functionals $S \colon \widetilde{F} \to \mathbb{R}$. Smolyak's result was part of his Ph.D thesis and was not published in a journal; it is generally known only through Bakhvalov's paper [8]. We formulate the results of Smolyak and Bakhvalov as follows.

Theorem 4.7. *Assume that $S \colon \widetilde{F} \to \mathbb{R}$ is a linear functional and F is a symmetric convex subset of $\widetilde{F}$. Assume that A is an arbitrary algorithm $A = \varphi \circ N^{\mathrm{ada}}$, where N^{ada} is of the form (4.9). We denote the non-adaptive information N^{non}, see (4.29), by*

$$N^{\mathrm{non}}(f) = [L_1(f), L_2(f), \ldots, L_n(f)].$$

Then there is a linear algorithm using nonadaptive information

$$A^{\mathrm{non}}(f) = \sum_{k=1}^{n} a_k L_k(f)$$

such that

$$e^{\mathrm{wor}}(A^{\mathrm{non}}) \le e^{\mathrm{wor}}(A).$$

Proof. Let

$$M = \{f \in F \mid N^{\mathrm{non}}(f) = N^{\mathrm{ada}}(f) = 0\}.$$

Since M is symmetric we have

$$|S(f)| \leq \tfrac{1}{2}|S(f) - A(0)| + \tfrac{1}{2}|S(-f) - A(0)|$$

and

$$\sup_{f \in M} |S(f)| \leq \sup_{f \in M} |S(f) - A(0)| \leq e^{\mathrm{wor}}(A).$$

As a consequence we obtain the important fact that

$$e^{\mathrm{wor}}(A) \geq \sup\{S(f) \mid f \in F, \ N^{\mathrm{non}}(f) = 0\} =: r.$$

Note that $r \geq 0$, due to the symmetry of F. Without loss of generality we may assume that $r < \infty$.

We prove that there is a linear algorithm $A^{\mathrm{non}} = \varphi^0 \circ N^{\mathrm{non}}$ with $e^{\mathrm{wor}}(A^{\mathrm{non}}) = r$. Define a convex set by

$$C = \{(S(f), L_1(f), L_2(f), \ldots, L_n(f)) \mid f \in F\} \subseteq \mathbb{R}^{n+1}$$

and consider a supporting hyperplane H through a boundary point y of C of the form

$$y = (r, 0, \ldots, 0).$$

We obtain $a_k \in \mathbb{R}$ such that

$$S(f) - \sum_{k=1}^{n} a_k L_k(f) \leq r \quad \text{for all } f \in F.$$

Since S is symmetric, we also obtain that the last sum is at least $-r$. Hence we have found that the linear algorithm

$$A^{\mathrm{non}}(f) = \sum_{k=1}^{n} a_k L_k(f)$$

satisfies $|S(f) - A^{\mathrm{non}}(f)| \leq r$ for all $f \in F$. Hence $e^{\mathrm{wor}}(A^{\mathrm{non}}) = r \leq e^{\mathrm{wor}}(A)$, where A^{non} uses the same non-adaptive information N^{non} that is used by A for the function $f = 0$. $\qquad\square$

We add in passing that if only the convexity of F is assumed, but not the symmetry, then for non-adaptive information an affine algorithm is optimal, see Sukharev [228]. This result was generalized to noisy information by Magaril-Ilyaev and Osipenko [139].

Many authors have studied the relation between general and linear algorithms for linear problems. In particular we want to mention the work of Creutzig and Wojtaszczyk [34], Mathé [140], Micchelli and Rivlin [146], Packel [184], Pietsch [191], Pinkus [193], as well as the works [242], [244], [279]. We should stress that some of

these authors use a different terminology and speak, for example, about s-numbers, n-widths, and the extension problem for linear operators. Nevertheless, the essence of their work addresses the relation mentioned above.

Actually, linear algorithms are optimal under the same conditions, see Theorem 4.5, under which adaption does not help, see again Creutzig and Wojtaszczyk [34].

Theorem 4.8. *Consider a linear problem. Assume additionally that*

- $G \in \{\mathbb{R}, L_\infty(\mu), B(K)\}$ *or*

- $\widetilde{F}$ *is a pre-Hilbert space or*

- $S \colon \widetilde{F} \to G = C(K)$, *where S is compact.*

Assume that $N \colon \widetilde{F} \to \mathbb{R}^n$ is a non-adaptive information. Then

$$r(N) = \inf_{\varphi \ linear} e^{\mathrm{wor}}(\varphi \circ N).$$

There are examples of linear problems for which linear algorithms are not optimal. Probably the first such example was found by Micchelli, see [242, p. 87]. It is of the form $S \colon \mathbb{R}^3 \to \mathbb{R}^2$ with an ℓ_4-norm in the target space. Other such examples are related to Sobolev embeddings and will be discussed in Section 4.2.4. See also the work of Donoho [49] and Candes and Tao [27] on compressed sensing.

An extreme example can be found in [279] or [242, p. 81–84]. Here, the error of *any* linear algorithm is infinite for any non-adaptive information N while the radius of a suitable non-adaptive information N is arbitrary small. This is *not* possible if the linear problem is *bounded* in the following sense:

- $\widetilde{F}$ and G are Banach spaces;

- S and N are bounded linear mappings;

- F is the unit ball of $\widetilde{F}$.

Under these conditions, Mathé [140] used an argument of Pietsch to prove the following result, see again Creutzig and Wojtaszczyk [34].

Theorem 4.9. *Consider a linear problem that is bounded with information $N \colon \widetilde{F} \to \mathbb{R}^n$. Then*

$$\inf_{\varphi \ linear} e^{\mathrm{wor}}(\varphi \circ N) \le (1 + \sqrt{n}) r(N). \qquad \square$$

Remark 4.10. The numbers $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{all}})$ and $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}})$ are inversely related to the information complexity of a problem, see Section 4.1.4. We know from the results above that linear algorithms are not always optimal. Nevertheless linear algorithms might be much easier to realize on a computer.

Therefore we also study the *linear widths* $e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{all}})$, sometimes also called *approximation numbers*, and the *linear sampling numbers* $e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{std}})$, where we only allow linear algorithms $\varphi \colon \mathbb{R}^n \to G$. $\qquad \square$

4.2.3 Linear Problems on Hilbert Spaces for Λ^{all}

In this section we assume that $H = \widetilde{F}$ is a Hilbert space and $S \colon H \to G$ is a bounded linear operator. We choose F as the unit ball of H. Then we know from Theorem 4.5 that adaption does not help and we know from Theorem 4.8 that linear algorithms are optimal. If we use the class Λ^{all} then there is a tight connection between minimal errors and *Gelfand numbers* or *Gelfand widths* that, in this case, are equal to the *approximation numbers* or *linear widths*. For more details about s-numbers and n-widths we refer to the books of Pietsch [191], [192] and Pinkus [193].

Theorem 4.11. *Let $S \colon H \to G$ be a bounded linear operator between a Hilbert space H with $\dim(H) \geq n$ and a normed space G. Then there exist orthonormal $e_1, e_2, \ldots, e_n$ in H such that the linear algorithm*

$$A_n(f) = \sum_{i=1}^{n} \langle f, e_i \rangle \, S(e_i)$$

is optimal, i.e.,

$$e^{\text{wor}}(n) = e^{\text{wor}}(A_n) = \sigma_{n+1}.$$

Here σ_{n+1} is the norm of S on the orthogonal complement of $\{e_1, e_2, \ldots, e_n\}$.

Proof. We know that there is an optimal subspace $U_n \subseteq H$ for the Gelfand widths of S, see the book of Pinkus [193, p. 16 and 31]. This follows from the fact that the unit ball of H or H^* is compact in the weak topology. It is also known that the optimal algorithm is linear and given by the so-called spline algorithm, see again, e.g., the book of Pinkus. $\qquad \square$

Spline algorithms and their properties are often described in the literature for the case when G is also a Hilbert space and $S \colon H \to G$ is compact, see [242], [244]. Compactness of S is equivalent to the fact that the problem is *solvable*, i.e., $\lim_{n \to \infty} e^{\text{wor}}(n) = 0$. Then there is a well known connection to the singular value decomposition of S, which we now describe. The adjoint operator $S^* \colon G \to H$ is defined by

$$\langle S(f), g \rangle_G = \langle f, S^*(g) \rangle_H$$

for all $f \in H$ and $g \in G$. Then the self-adjoint operator

$$W = S^*S \colon H \to H$$

is also compact. Let $\{\lambda_i, e_i\}$ be the eigenpairs of the operator W such that $W(e_i) = \lambda_i e_i$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and $\langle e_i, e_j \rangle = \delta_{ij}$. In this case, the optimal information is given by

$$N_n(f) = [\langle f, e_1 \rangle, \langle f, e_2 \rangle, \ldots, \langle f, e_n \rangle].$$

More precisely, we have the following result.

Corollary 4.12. *Let* $S: H \rightarrow G$ *be a bounded linear operator between a Hilbert space* H *with* $\dim(H) \geq n$ *and another Hilbert space* G. *Let* $\sigma_i = \sqrt{\lambda_i}$, *where* $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ *are the eigenvalues of* $W = S^*S: H \rightarrow H$ *with* $W(e_i) = \lambda_i e_i$ *and orthonormal* $\{e_i\}$. *Then the linear algorithm*

$$A_n(f) = \sum_{i=1}^{n} \langle f, e_i \rangle \, S(e_i)$$

is n *th optimal, i.e.,*

$$e^{\mathrm{wor}}(n) = e^{\mathrm{wor}}(A_n) = \sigma_{n+1}.$$

Note that

$$\langle S(e_i), S(e_j) \rangle_G = \langle e_i, (S^*S)(e_j) \rangle = \langle e_i, W(e_j) \rangle = \lambda_j \delta_{ij}.$$

Hence the elements $S(e_i)$ are orthogonal in G and $\langle S(e_i), S(e_i) \rangle_G = \lambda_i$.

4.2.4 Sobolev Embeddings

We discuss Sobolev embeddings, which correspond to approximation problems with solution operators being the identity (or embedding) from one Sobolev space to another (larger) Sobolev space. These embeddings are interesting for its own but also are basic for the understanding of operator equations and the optimal approximation of their solutions.

Sobolev spaces are just examples of more general Besov and Triebel–Lizorkin spaces. Nevertheless, to simplify the presentation, we only discuss the *classical Sobolev spaces* $W_p^k(\Omega)$ with an integer k, as well as the more general (*fractional*) *Sobolev spaces* $H_p^s(\Omega)$ with an arbitrary real s.

In this section we only study results concerning the optimal order of convergence for fixed spaces. These results do not change if we replace the norm of a space by an equivalent norm. Hence it is not important to distinguish between different equivalent norms. However, for the study of tractability, the choice of norm is crucial. We may have tractability for one norm, and intractability for an equivalent norm.

We need the *Fourier transform* $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ on the *Schwartz space* $\mathcal{S}$ and on the space $\mathcal{S}'$ of *tempered distributions*. We now briefly define these notions, see any book on distributions, e.g., Haroske and Triebel [77], for more detailed information.

For $d \in \mathbb{N}$, let

$$\mathcal{S}(\mathbb{R}^d) = \{f \in C^{\infty}(\mathbb{R}^d) \mid \|f\|_{k,\ell} < \infty \text{ for all } k, \ell \in \mathbb{N}_0\},$$

where

$$\|f\|_{k,\ell} = \sup_{x \in \mathbb{R}^d} (1 + \|x\|_2^2)^{k/2} \sum_{|\alpha| \leq \ell} |D^{\alpha} f(x)|.$$

Here D^α denotes a partial derivative, $\alpha \in \mathbb{N}_0^d$, and $|\alpha| = \sum_{i=1}^d \alpha_i$. That is,

$$D^\alpha f = \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} x_1 \partial^{\alpha_2} x_2 \ldots \partial^{\alpha_d} x_d} f.$$

The space $\mathcal{S}(\mathbb{R}^d)$ is often called the Schwartz space of rapidly decreasing infinitely differentiable functions in $\mathbb{R}^d$. A sequence $\{f_n\}$ in $\mathcal{S}(\mathbb{R}^d)$ is said to converge in $\mathcal{S}(\mathbb{R}^d)$ to $f \in \mathcal{S}(\mathbb{R}^d)$ if $\|f_n - f\|_{k,\ell} \to 0$ for all $k, \ell \in \mathbb{N}_0$.

For a smooth function $f \in \mathcal{S}(\mathbb{R}^d)$, the Fourier transform and its inverse are given by

$$(\mathcal{F} f)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-ixy} f(y) \, dy$$

and

$$(\mathcal{F}^{-1} f)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{ixy} f(y) \, dy$$

with $xy = \sum_{k=1}^d x_k y_k$. Both mappings $\mathcal{F}$ and $\mathcal{F}^{-1}$ are bijective on $\mathcal{S}(\mathbb{R}^d)$.

The space $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions is the set of all linear continuous functionals T over $\mathcal{S}(\mathbb{R}^d)$, i.e., T is linear and $T(f_n) \to T(f)$ whenever $\{f_n\}$ converges to f in $\mathcal{S}(\mathbb{R}^d)$. If $f \in L_p(\mathbb{R}^d)$ then T_f, given by

$$(T_f)(g) = \int_{\mathbb{R}^d} f(x) g(x) \, dx,$$

belongs to $\mathcal{S}'(\mathbb{R}^d)$ and is called a *regular distribution*. Then we define $\|T_f\|_{L_p(\mathbb{R}^d)} = \|f\|_{L_p(\mathbb{R}^d)}$ and we can identify f and T_f.

The Fourier transform $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ on $\mathcal{S}'(\mathbb{R}^d)$ are given by

$$(\mathcal{F} T)(f) = T(\mathcal{F} f) \quad \text{and} \quad (\mathcal{F}^{-1} T)(f) = T(\mathcal{F}^{-1} f),$$

where $f \in \mathcal{S}(\mathbb{R}^d)$. Then $\mathcal{F}$ as well as $\mathcal{F}^{-1}$ are bijective on $\mathcal{S}'(\mathbb{R}^d)$.

For $s \in \mathbb{R}$ and $1 < p < \infty$, the space $H_p^s(\mathbb{R}^d)$ can be defined by

$$H_p^s(\mathbb{R}^d) = \left\{ T \mid T \in \mathcal{S}'(\mathbb{R}^d),\ \|T\|_{H_p^s} = \left\| \mathcal{F}^{-1}\left((1 + \|\cdot\|^2)^{s/2} \mathcal{F} T \right) \right\|_{L_p(\mathbb{R}^d)} < \infty \right\}.$$

Hence $T \in H_p^s(\mathbb{R}^d)$ is a tempered distribution such that $\mathcal{F}^{-1}\left((1 + \|\cdot\|^2)^{s/2} \mathcal{F} T \right)$ is a regular distribution T_g with $g \in L_p(\mathbb{R}^d)$. The Sobolev embedding theorem says that $H_p^s(\mathbb{R}^d)$ is continuously embedded into the space $C_b(\mathbb{R}^d)$ of bounded continuous functions if $s > d/p$. Then we can identify $H_p^s(\mathbb{R}^d)$ with a space of continuous functions and function values are well defined.

In particular, if $s = k \in \mathbb{N}_0$, we obtain

$$H_p^s(\mathbb{R}^d) = W_p^k(\mathbb{R}^d) = \left\{ f \in L_p(\mathbb{R}^d) \mid \|f\|_{W_p^k} = \left(\sum_{|\alpha| \le k} \|D^\alpha f\|_{L_p(\mathbb{R}^d)}^p \right)^{1/p} < \infty \right\}.$$

Hence the norms of $H_p^s(\mathbb{R}^d)$ and $W_p^k(\mathbb{R}^d)$ are equivalent if $k = s \in \mathbb{N}_0$, see, e.g., the book of Triebel [249]. For $s = 0$ and $1 < p < \infty$, we obtain the L_p-spaces.

The spaces $C^k(\mathbb{R}^d)$ of k times continuously differentiable functions and the classical Sobolev spaces $W_p^k(\mathbb{R}^d)$ for $k \in \mathbb{N}_0$ and $p \in \{1, \infty\}$ do not fit nicely into this scale of functions but many results for $1 < p < \infty$ carry over to those spaces for $p \in \{1, \infty\}$.

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain (with a nonempty interior). Then the function spaces $H_p^s(\Omega)$ and $W_p^k(\Omega)$ are defined by restriction of functions from $H_p^s(\mathbb{R}^d)$. More precisely, $f \in H_p^s(\Omega)$ iff there exists $\tilde{f} \in H_p^s(\mathbb{R}^d)$ such that $\tilde{f}\big|_\Omega = f$. Then the norm of f is defined by

$$\|f\|_{H_p^s(\Omega)} = \inf\{\|\tilde{f}\|_{H_p^s} \mid \tilde{f}_{|\Omega} = f\}.$$

If $s = k \in \mathbb{N}_0$ then an equivalent norm can also be defined in the more traditional way in terms of derivatives without the need to extend f to $\mathbb{R}^d$ by the classical Sobolev norm

$$\|f\|_{W_p^k(\Omega)} = \Big(\sum_{|\alpha| \leq k} \|D^\alpha f\|_{L_p(\Omega)}^p \Big)^{1/p}.$$

Consider now the embedding, or the approximation problem, $\mathrm{App}(f) = f$, where

$$\mathrm{App} \colon H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega),$$

for a bounded Lipschitz domain $\Omega \subseteq \mathbb{R}^d$. This embedding is well defined and compact iff

$$s_1 - s_2 > d \left(\frac{1}{p_1} - \frac{1}{p_2} \right)_+, \tag{4.31}$$

see Haroske and Triebel [76]. From now on, we always assume that (4.31) holds. Otherwise there is no chance that the worst case errors of algorithms using n information operations can converge to 0 as n tends to infinity.

We want to approximate f from the unit ball F of the space $H_{p_1}^{s_1}(\Omega)$. We are interested in the numbers $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{all}})$ and $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}})$ for these embeddings, see Sections 4.1.1 and 4.1.4 for the definitions. The numbers $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{all}})$ are, up to a factor of two, the same as the *Gelfand widths* and they describe the error of the optimal (in general nonlinear) algorithm that is based on n arbitrary linear functionals. The respective error bounds for standard information, i.e., when we only allow algorithms that are based on function evaluations, are $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}})$ and are sometimes called (*nonlinear*) *sampling numbers*.

Remark 4.13. Function values can only be used if they are well defined. This is why the sampling numbers are usually only studied if

$$s_1 > d/p_1. \tag{4.32}$$

Then $H_{p_1}^{s_1}(\Omega)$ can be considered as a subset of $C(\bar{\Omega})$ and function values are well defined and $f \mapsto f(x)$ is a continuous functional on this space.

If the embedding inequality (4.32) into the space $C(\bar{\Omega})$ is not fulfilled then one can replace the space $H_{p_1}^{s_1}(\Omega)$ by the smaller space $H_{p_1}^{s_1}(\Omega) \cap C(\bar{\Omega})$. This will be convenient

later, when we also discuss randomized algorithms. We will do this replacement in the following without further mentioning it. That is, when $s_1 \leq d/p_1$ and we consider the class Λ^{std}, then the space $H_{p_1}^{s_1}(\Omega)$ is to be replaced by $H_{p_1}^{s_1}(\Omega) \cap C(\overline{\Omega})$. Observe that this last space is *not* a Banach space since we take the norm of $H_{p_1}^{s_1}(\Omega)$.   □

The numbers $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{all}})$ and $e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}})$ are inversely related to the information complexity of the embedding, see again Section 4.1.4. We know from Section 4.2.2 that linear algorithms are not always optimal. Nevertheless linear algorithms might be much easier to implement on a computer. Therefore we also study the *linear widths* or *approximation numbers* $e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{all}})$ and the *linear sampling numbers* $e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{std}})$ that are defined as the minimal worst case errors of linear algorithms using at most n information operations from the class Λ^{all} or Λ^{std}, respectively.

The case $p_2 \leq p_1$ is much simpler than $p_1 < p_2$ and we start with this simpler case.

Theorem 4.14. *Consider the embedding*

$$\mathrm{App}: H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega),$$

where $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain, $1 < p_2 \leq p_1 < \infty$, and (4.31) holds.

- *We have*

$$e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{all}}) = \Theta(e^{\mathrm{wor}}(n, \Lambda^{\mathrm{all}})) = \Theta(n^{-(s_1-s_2)/d}).$$

- *If $s_1 > d/p_1$ and $s_2 > 0$ then*

$$e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}}) = \Theta(n^{-(s_1-s_2)/d}).$$

  For $\Omega = [0,1]^d$, it is known that this order can be obtained by linear algorithms, i.e.,

$$e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{std}}) = \Theta(n^{-(s_1-s_2)/d}).$$

- *If $s_1 > d/p_1$ and $s_2 \leq 0$ then*

$$e^{\mathrm{wor\text{-}lin}}(n, \Lambda^{\mathrm{std}}) = \Theta(e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}})) = \Theta(n^{-s_1/d}).$$

- *If $s_1 \leq d/p_1$ then*

$$e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}}) = \Theta(1).$$

Remark 4.15. a) We stress that the constants in the Θ-notation are independent of n but may depend on all the parameters of the function spaces.

b) Analogous error bounds hold for the classical Sobolev spaces $W_p^k(\Omega)$ for $p = \infty$ (in this case, one may also consider the space $C^k(\overline{\Omega})$) and for $p = 1$ (and in this case the classical Sobolev space $W_1^k(\Omega)$ is embedded into $C(\overline{\Omega})$ for $d = k$).

c) In particular, nonlinear algorithms are not essentially better than linear algorithms if $p_2 \leq p_1$. Also the *Kolmogorov widths* (which are defined as the minimal error in linear n-dimensional subspaces) are of the same order as the linear widths in this case.   □

The results, and of course also the proofs, are much more complicated if $p_1 < p_2$. We start with the linear and the nonlinear sampling numbers.

Theorem 4.16. *Consider the embedding*

$$\text{App}\colon H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega),$$

where $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain and (4.31) *holds. We assume that $p_1 < p_2$.*

- *Let $s_1 > d/p_1$. If $s_2 > 0$ or $d/p_2 - d/p_1 < s_2 \le 0$ then*

$$e^{\text{wor}}(n, \Lambda^{\text{std}}) = \Theta(n^{-(s_1-s_2)/d+(1/p_1-1/p_2)}).$$

  The same order holds for the numbers $e^{\text{wor-lin}}(n, \Lambda^{\text{std}})$ if $d/p_2 - d/p_1 < s_2 \le 0$ or if $s_2 > 0$ and Ω is a cube.

- *Let $s_1 > d/p_1$. If $s_2 \le d/p_2 - d/p_1$ then*

$$e^{\text{wor-lin}}(n, \Lambda^{\text{std}}) = \Theta(e^{\text{wor}}(n, \Lambda^{\text{std}})) = \Theta(n^{-s_1/d}).$$

- *Let $s_1 \le d/p_1$. Then*
$$e^{\text{wor}}(n, \Lambda^{\text{std}}) = \Theta(1).$$

The results on sampling numbers presented in Theorems 4.14 and 4.16 are from Triebel [248] and Vybíral [253]. The case $s_1 \le d/p_1$ is much more interesting in the randomized setting, see Heinrich [86].

Open Problem 18. [9]
- We conjecture that

$$e^{\text{wor-lin}}(n, \Lambda^{\text{std}}) = \Theta(e^{\text{wor}}(n, \Lambda^{\text{std}}))$$

  holds for all bounded Lipschitz domains Ω, also if $s_2 > 0$. Prove (or disprove) this conjecture.

For the case $\Omega = [0, 1]^d$ we refer to Vybíral [253]. It would be of interest to have explicit constructions of linear sampling algorithms for arbitrary polyhedra.

Now we summarize the known results on the Gelfand numbers $e^{\text{wor}}(n, \Lambda^{\text{all}})$ and the approximation numbers $e^{\text{wor-lin}}(n, \Lambda^{\text{all}})$ in the remaining case $p_1 < p_2$. In both cases, the asymptotic behavior was studied in great detail and only few "limiting cases" are open. We start with linear algorithms, see Edmunds and Triebel [54], Caetano [24], and Triebel [249] for the final results.

---

[9]Recently solved by Heinrich [87].

Theorem 4.17. *Consider the embedding*

$$\text{App}: H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega),$$

where $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain and (4.31) *holds.*

- *If either*

$$p_1 \leq p_2 \leq 2 \quad or \quad 2 \leq p_1 \leq p_2 \quad or \quad p_2 \leq p_1$$

  then

$$e^{\text{wor-lin}}(n, \Lambda^{\text{all}}) = \Theta(n^{-(s_1-s_2)/d+(1/p_1-1/p_2)_+}).$$

- *If*

$$p_1 < 2 < p_2 \quad and \quad s_1 - s_2 > d \max(1 - 1/p_2, 1/p_1)$$

  then

$$e^{\text{wor-lin}}(n, \Lambda^{\text{all}}) = \Theta(n^{-\lambda}),$$

  where

$$\lambda = \frac{s_1 - s_2}{d} - \max\left(\frac{1}{2} - \frac{1}{p_2}, \frac{1}{p_1} - \frac{1}{2}\right) > \frac{1}{2}.$$

- *If*

$$p_1 < 2 < p_2 \quad and \quad s_1 - s_2 < d \max(1 - 1/p_2, 1/p_1)$$

  then

$$e^{\text{wor-lin}}(n, \Lambda^{\text{all}}) = \Theta(n^{-((s_1-s_2)/d-1/p_1+1/p_2)\cdot p_1'/2}),$$

  where $1/p_1 + 1/p_1' = 1$.

Remark 4.18. Consider

$$\text{App}: H_{p_1}^{s_1}(\Omega) \to L_{p_2}(\Omega),$$

with $s_1 > d/p_1$. Then the class Λ^{all} is significantly better than the class Λ^{std} for linear algorithms iff $p_1 < 2 < p_2$. In all other cases, we have

$$e^{\text{wor-lin}}(n, \Lambda^{\text{all}}) = \Theta(e^{\text{wor-lin}}(n, \Lambda^{\text{std}})),$$

see [173]. □

The Gelfand numbers $e^{\text{wor}}(n, \Lambda^{\text{all}})$ of Sobolev embeddings have been studied in great detail. The Gelfand numbers of diagonal operators were studied by Gluskin [65] and it is possible to apply his results to Sobolev embeddings, see Linde [137] and Tikhomirov [238]. These works contain details about the *Kolomogorov numbers* that might be more closely related to classical approximation theory – but not so much to the study of optimal algorithms. Results on the Gelfand numbers can also be found in the book of Lorentz, v. Golitschek and Makovoz [138]. Some basic ideas go back to Kashin [106], [107]. The following result, mainly a summary of known facts, is from Vybíral [254].

Theorem 4.19. *Consider the embedding*

$$\text{App: } H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega),$$

where $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain and (4.31) holds.

- *If either*

$$2 \leq p_1 \leq p_2 \quad or \quad p_2 \leq p_1$$

  then

$$e^{\text{wor}}(n, \Lambda^{\text{all}}) = \Theta(n^{-(s_1-s_2)/d + (1/p_1 - 1/p_2)_+}).$$

- *If*

$$1 \leq p_1 < p_2 \leq 2 \quad and \quad (s_1 - s_2)/d > \frac{1/p_1 - 1/p_2}{2/p_1 - 1}$$

  then

$$e^{\text{wor}}(n, \Lambda^{\text{all}}) = \Theta(n^{-(s_1-s_2)/d}).$$

- *If*

$$1 \leq p_1 < p_2 \leq 2 \quad and \quad (s_1 - s_2)/d < \frac{1/p_1 - 1/p_2}{2/p_1 - 1}$$

  or if

$$1 \leq p_1 \leq 2 < p_2 \leq \infty \quad and \quad (s_1 - s_2)/d < 1/p_2'$$

  then

$$e^{\text{wor}}(n, \Lambda^{\text{all}}) = \Theta(n^{(-(s_1-s_2)/d + 1/p_1 - 1/p_2) \cdot p_1'/2}).$$

- *If*

$$1 \leq p_1 \leq 2 < p_2 \quad and \quad (s_1 - s_2)/d > 1/p_2'$$

  then

$$e^{\text{wor}}(n, \Lambda^{\text{all}}) = \Theta(n^{-(s_1-s_2)/d + 1/2 - 1/p_2}),$$

where $1/p_i + 1/p_i' = 1$.

Remark 4.20. It is interesting that optimal nonlinear algorithms for the class Λ^{all} are sometimes significantly better than optimal linear algorithms. This can happen only if $p_1 < 2$ and $p_1 < p_2$. □

Remark 4.21. a) We defined the spaces H_p^s only for $1 < p < \infty$. However, the results above also hold for $p = 1$ and $p = \infty$, as well as for more general spaces. For $p \in \{1, \infty\}$ and $s \in \mathbb{N}$, we can also consider the classical Sobolev spaces. We recommend the recent papers of Vybíral [253], [254] for further results on sampling numbers and s-numbers (approximation numbers, Gelfand numbers, Kolmogorov numbers) of Sobolev embeddings.

b) The spaces $H_p^s = H_p^s(\Omega)$ with $1 < p < \infty$ and $s \in \mathbb{R}$ are called Sobolev spaces or fractional Sobolev spaces or Bessel potential spaces. They are special cases of the more general Triebel–Lizorkin spaces, $H_p^s = F_{p,2}^s$, see the recent book of Triebel [249]. For simplicity we discuss here only the spaces H_p^s although most results

are valid also for the spaces $F_{p,q}^s$ and the Besov spaces $B_{p,q}^s$; see again Triebel [249], where quasi-Banach spaces are also studied in great detail.

c) We do not even try to give historical remarks on the different widths and s-numbers and refer the reader to the books of Lorentz, v. Golitschek and Makovoz [138], Pietsch [191], [192], Pinkus [193], Tikhomirov [238], and Triebel [246].

However, we give a few remarks on sampling numbers. There is a vast literature for $\Omega = [0, 1]^d$ and also for the periodic case, i.e., when Ω is the torus. In these cases it is well known (but we do not know who proved this first) that the (linear as well as nonlinear) sampling numbers for the embedding $I : W_{p_1}^k \to L_{p_2}$ are of the form

$$\Theta\big(n^{-k/d+(1/p_1-1/p_2)_+}\big), \tag{4.33}$$

see, e.g., the book Ciarlet [29] for upper bounds and Heinrich [80] for upper and lower bounds. Special cases of this formula for $\Omega = [0, 1]^d$ are also contained in [160] and in the book of Temlyakov [232]. Besov spaces for $\Omega = [0, 1]^d$ are studied by Kudryavtsev [117].

More general domains were studied by Wendland [276], who basically studies the embedding from $C^k(\bar{\Omega})$ into $L_\infty(\Omega)$. A proof of (4.33) for bounded Lipschitz domains (and more general spaces) is given in [173].

As remarked in Theorem 2.14 of Triebel [247], assertions for approximation numbers with respect to bounded C^∞ domains in $\mathbb{R}^d$ remain valid for bounded Lipschitz domains. The reason is the following. If $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain then there is a (even universal, i.e., simultaneously for many function spaces) linear extension operator, see Rychkov [205] and the recent book of Triebel [249, Section 1.11 and 4.3] for details. Using this extension operator, one can use the results for a cube $\widetilde{\Omega}$ containing Ω to obtain algorithms for Ω. The same is true for other widths, such as the Gelfand numbers. For the sampling numbers, however, it is *not* clear whether the order of convergence depends on Ω or not. We believe that for most of the standard function spaces the order of convergence does not depend on Ω. But this conjecture has to be proved, since we can only use function values of the function $f : \Omega \to \mathbb{R}$ itself, not of a suitable extension $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ of f.

Triebel [250] studies sampling numbers for embeddings into L_{p_2} (with $0 < p_2 \le \infty$), where now $\Omega \subseteq \mathbb{R}^d$ is an arbitrary bounded domain. For suitably defined spaces it turns out that again the order of convergence is given by (4.33).   □

Remark 4.22. The results that are presented in this section can be applied to the optimal approximation of the solution of elliptic problems, see the book of Werschulz [277]. If we also want to include arbitrary (non-regular) elliptic problems then it is crucial to discuss the embedding

$$I : H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega)$$

with a negative s_2, see [36], [37], [38], where the sampling numbers and approximation numbers are also compared to best n-term approximation with respect to an optimal Riesz bases or with respect to optimal frames. Sampling numbers for the optimal approximation of elliptic PDEs are also studied by Vybíral [253].   □

Remark 4.23. We will discuss Sobolev embeddings in the randomized setting later in Remark 4.43. □

Remark 4.24. Let us discuss the problem of numerical integration,

$$S(f) = \int_\Omega f(x)\, \mathrm{d}x,$$

for the classes $H_p^s(\Omega)$ where $\Omega \subseteq \mathbb{R}^d$ is again a bounded Lipschitz domain.

For deterministic algorithms we have to assume that function evaluations are continuous, i.e., $s > d/p$, and then the optimal order of convergence is

$$e^{\mathrm{wor}}(n, \Lambda^{\mathrm{std}}) = \Theta(n^{-s/d}),$$

as is also the case for L_1-approximation. The upper bound follows from Theorem 4.14, where we have the same bound for the L_1-approximation. The lower bound is proved (again as the lower bound for L_1-approximation) via the familiar technique of bump functions.

With randomized algorithms, the order $n^{-s/d}$ can be improved for all $p > 1$ and now it is enough to assume that $H_p^s(\Omega)$ is embedded into $L_2(\Omega)$. Upper bounds can be proved via variance reduction and so one obtains the order of L_2-approximation plus 1/2. For $p \geq 2$ we obtain

$$e^{\mathrm{ran}}(n, \Lambda^{\mathrm{std}}) = \Theta(n^{-s/d-1/2}),$$

for $1 < p < 2$ we obtain

$$e^{\mathrm{ran}}(n, \Lambda^{\mathrm{std}}) = \Theta(n^{-s/d-1+1/p}).$$

All these results can be found in the literature, at least for special cases. Bakhvalov proved such results already in 1959, see Bakhvalov [7]. See also Heinrich [80], [81] and [160], [242]. For the proof of the upper bounds for arbitrary bounded Lipschitz domains the paper [173] can be used. There one can find also bounds for L_1-approximation (and hence upper bounds for integration) for more general function spaces. □

4.3 Some Results for Different Settings

4.3.1 Average Case Setting

In this section we describe some results for the average case setting. Many more details can be found in [242]. In the average case setting, the error and cost of algorithms are defined by the average performance of an algorithm with respect to a probability measure μ defined on Borel sets of F. Here F is a subset of a separable Banach space $\widetilde{F}$. Many problems treated in IBC are defined on subsets of infinite dimensional

spaces and we believe that a Gaussian measure or a truncated Gaussian measure may serve as a good candidate for the probability measure on the set F.

We mainly discuss results for linear problems $S \colon \widetilde{F} \to G$ on separable Banach spaces $\widetilde{F}$ and G. Recall that for $f \in F$ we want to approximate $S(f)$ knowing $y = N(f) \in \mathbb{R}^n$, where S is a solution operator and N is an information operator defined as in Section 4.1.1. Hence we assume that N is given by (4.9). In addition we always assume that S and N are measurable, which is a rather weak assumption. We assume that μ is a probability measure on F. Then the probability measure $\nu = \mu S^{-1}$ is the a priori measure on the set of solution elements $S(f) \in G$.

We also need the measure $\mu_1 = \mu N^{-1}$ on the Borel sets of $N(F)$. If $F \subseteq \widetilde{F}$ is closed, then there exists a unique (modulo sets of μ_1-measure zero) family of conditional probability measures $\mu_2(\cdot \mid y)$ defined on Borel sets of F such that

- $\mu_2(N^{-1}(y) \mid y) = 1$ for almost all $y \in N(F)$,

- $\mu_2(B \mid \cdot)$ is measurable for any Borel set B of F,

- $\mu(B) = \int_{N(F)} \mu_2(B \mid y)\mu_1(\mathrm{d}y)$ for any Borel set B of F.

Such a family is called a *regular conditional probability distribution*. It exists since F is a Polish space, see the book of Parthasarathy [189, p. 147]. Then for any measurable function $Q \colon F \to \mathbb{R}_+$ we have

$$\int_F Q(f)\mu(\mathrm{d}f) = \int_{N(F)} \left(\int_{N^{-1}(y)} Q(f)\mu_2(\mathrm{d}f \mid y) \right) \mu_1(\mathrm{d}y).$$

The essence of the last formula is that we can first integrate over the set $N^{-1}(y)$ of elements from F that are indistinguishable with respect to the information N, i.e., over all f such that $N(f) = y$. This integration is done with respect to the conditional measure $\mu_2(\cdot \mid y)$. Then we integrate over the set $N(F)$ of values of N with respect to the measure μ_1 that tells us about the distribution of information values.

Now we define the measure $\nu(\cdot \mid y)$ on the set $S(N^{-1}(y))$ by

$$\nu(B \mid y) = \mu_2(S^{-1}(B) \mid y) = \mu_2(\{f \in F \mid S(f) \in B\} \mid y),$$

where $B \subseteq G$ is measurable. Thus $\nu(\cdot \mid y)$ is a probability measure on $S(N^{-1}(y))$ and tells us about the distribution of solution elements $S(f)$ that are indistinguishable with respect to N.

The *local average radius* of the information N is defined by

$$r^{\mathrm{avg}}(N, y) = \inf_{x \in G} \sqrt{\int_{S(N^{-1}(y))} \|x - g\|^2 \nu(\mathrm{d}g \mid y)}.$$

Of course we have $r^{\mathrm{avg}}(N, y) \leq r^{\mathrm{wor}}(N, y)$, see Section 4.1. It can be checked that for a separable normed space G the mapping $y \mapsto r^{\mathrm{avg}}(N, y)$ is measurable, see

Wasilkowski [258] and also [242, p. 199]. Hence we also assume that G is separable, so that we can now define the *average radius of information N* by

$$r^{\mathrm{avg}}(N) = \sqrt{\int_{N(F)} r^{\mathrm{avg}}(N, y)^2 \mu_1(\mathrm{d}y)}.$$

Again it is easy to see that

$$r^{\mathrm{avg}}(N) \le r^{\mathrm{wor}}(N),$$

see Section 4.1.

Assume now that the algorithm $A = \varphi \circ N \colon F \to G$ is measurable. Then we define the *average case error* of A by

$$e^{\mathrm{avg}}(A) = \left(\int_F \|S(f) - \varphi(N(f))\|^2 \mu(\mathrm{d}f) \right)^{1/2}. \tag{4.34}$$

It is possible to modify this definition for non-measurable φ, see [242, p. 205]. Since measurability is only a weak assumption and, moreover, one can prove that measurable algorithms are optimal, see [169], we skip the details. As in the worst case, the average radius of information is equal to the average case error of an optimal algorithm. The following result is from Wasilkowski [258], it can be also found in [242].

Theorem 4.25. *Assume that $S \colon F \to G$ and N are measurable, where $\widetilde{F}$ and G are separable Banach spaces and $F \subseteq \widetilde{F}$ is closed. Then*

$$r^{\mathrm{avg}}(N) = \inf_{\varphi} \, e^{\mathrm{avg}}(\varphi, N).$$

We add that the infimum is attained by a unique (up to sets of measure zero) φ if $r^{\mathrm{avg}}(N) < \infty$ and if G is a strictly convex dual space, see [169].

Remark 4.26. As in Section 4.1 for the worst case error, we also generalize the definition (4.34) for some nonlinear problems such as zero finding. If the error $e(A, f) \ge 0$ is defined for an algorithm A at $f \in F$, for instance as in Remark 4.1, then, assuming that $e(A, \cdot)$ is measurable, we define the average error of A by

$$e^{\mathrm{avg}}(A) = \left(\int_F e(A, f)^2 \mu(\mathrm{d}f) \right)^{1/2}. \qquad \square$$

So far we have assumed that information N of the form of (4.9) is given and fixed. As in (4.10), let $n(f)$ denote the cardinality of N at f. We define the *average cardinality* of N by

$$\mathrm{card}^{\mathrm{avg}}(N) = \int_F n(f) \mu(\mathrm{d}f).$$

Observe that although $n(f)$ takes obviously integer values, the average cardinality of N does not have to be an integer. In fact, it is easy to see that $\mathrm{card}^{\mathrm{avg}}(N)$ can take any non-negative real value.

Since the cost of algorithms is often proportional to the cardinality of N, we take the average cardinality of N as a measure of the cost of an algorithm $A = \varphi \circ N$. This is justified if the arithmetic cost of N and φ is comparable or negligible to $\mathrm{card}^{\mathrm{avg}}(N)$.

We want to compare the efficiency of different algorithms with average cardinality at most n. Define

$$e^{\mathrm{avg}}(n) = \inf\{e^{\mathrm{avg}}(\varphi \circ N) \mid \mathrm{card}^{\mathrm{avg}}(N) \le n\}.$$

Formally, one can now consider n to be any positive real number. Without much loss of generality, we still assume, as in the worst case setting, that n is a non-negative integer.

We write $e^{\mathrm{avg}}(n, \Lambda)$ if we want to stress the set Λ of admissible linear functionals. For $n = 0$, $e^{\mathrm{avg}}(0)$ is the average error of the optimal constant algorithm which is also called the *initial error* in the average case setting.

For the absolute error criterion we define

$$n^{\mathrm{avg\text{-}abs}}(\varepsilon, S) = \min\{n \mid \text{there exists } A = \varphi \circ N \text{ with } \mathrm{card}^{\mathrm{avg}}(N) \le n, e^{\mathrm{avg}}(A) \le \varepsilon\},$$

and for the normalized error criterion we define

$$n^{\mathrm{avg\text{-}nor}}(\varepsilon, S) = \min\{n \mid \text{there exists } A = \varphi \circ N \text{ with} \\ \mathrm{card}^{\mathrm{avg}}(N) \le n, e^{\mathrm{avg}}(A) \le \varepsilon e^{\mathrm{avg}}(0)\}.$$

In the average case setting, we want to achieve the minimal average error with an average cardinality $\int_F n(f)\mu(\mathrm{d}f) \le n$. This means that $n(f)$ can be much larger than n but this can only happen with a small probability. It is an interesting problem to see how much we gain by using information having *varying cardinality* instead of fixed cardinality. That is, how much do we gain when we allow $n(f)$ to vary with f?

We will report later how much varying cardinality can help for linear problems and why it cannot change polynomial and weak tractability conditions. For nonlinear problems, the situation may be quite different. More precisely, there exist nonlinear problems where algorithms with average cardinality n are much better than algorithms with a fixed cardinality n. Such a problem is discussed in the following remark, see [171] for the proofs and more details.

Remark 4.27. We consider classes F_r of functions defined by

$$F_r = \{f \in C^r[0, 1] \mid f(0) < 0 < f(1)\},$$

where $r \in \mathbb{N}_0$. The class F_r is equipped with a conditional r-fold Wiener measure μ_r. Such a Gaussian measure is derived from the classical Wiener measure by r-fold integration and translation by suitable polynomials.

We want to approximate a zero of the nonlinear equations $f(x) = 0$ for $f \in F_r$. We use algorithms A that are based on function values $f(x_i)$ or derivatives $f^{(k_i)}(x_i)$ at adaptively chosen knots x_i. Of course, we assume that $k_i \le r$. The output of an algorithm A is a real number and we use the root criterion,

$$e(A, f) = \mathrm{dist}(f^{-1}(0), A(f)),$$

as our error criterion. We consider the worst case and average case errors of A defined as

$$e^{\text{wor}}(A) = \sup_{f \in F_r} e(A, f),$$

$$e^{\text{avg}}(A) = \int_{F_r} e(A, f)\mu_r(\mathrm{d}f).$$

In [171] we present an algorithm A_ε whose worst case error is at most ε and which is defined as follows. We set $[s_0, t_0] = [0, 1]$ as the initial interval with a zero of f. We compute $f(0)$ and $f(1)$. For $i = 1, 2, \ldots$, we compute $f(x_i)$ and the new enclosing interval $[s_i, t_i]$ of a zero of f. We stop if $t_i - s_i \leq 2\varepsilon$. In that case we return $(s_i + t_i)/2$. The algorithm uses steps of the regula falsi (R), the secant method (S), and the bisection method (B). In a bisection step we use $x_i = (s_{i-1} + t_{i-1})/2$, in a step of the regula falsi we define x_i as the zero of the line through $(s_{i-1}, f(s_{i-1}))$ and $(t_{i-1}, f(t_{i-1}))$, while the secant method uses the zero of the line through $(x_{i-1}, f(x_{i-1}))$ and $(x_{i-2}, f(x_{i-2}))$. If a secant step fails, i.e., it does not give a $x_i \in [s_{i-1}, t_{i-1}]$ then we do a bisection step, see [171] for the details and [166] for numerical results.

A typical pattern is R R S ... S B R R S B R R S S S S S S and so on, i.e., from one step on we use (with probability 1) only the secant method. We always have a length reduction $t_i - s_i \leq (t_{i-4} - s_{i-4})/2$ of the interval with a zero of f. The worst case error of the algorithm A_ε is bounded by ε for each $f \in F_0 \subseteq F_r$. The computational cost is proportional to the number of function evaluations. For $r \geq 2$, the average cardinality of function values of A_ε satisfies

$$\text{card}^{\text{avg}}(A_\varepsilon) \leq \frac{1}{\ln \frac{1 + \sqrt{5}}{2}} \cdot \ln \ln(1/\varepsilon) + c_r,$$

where c_r depends only on the measure μ_r, and can be computed numerically.

The algorithm A_ε enjoys a number of optimality properties. One can prove a lower bound for very general algorithms:

- instead of function evaluations we also allow the evaluation of derivatives at any points,

- instead of enclosing algorithms with a guaranteed error ε we consider arbitrary algorithms with average error ε.

Nevertheless, the average cardinality of any such algorithm cannot be much smaller than the cardinality of the algorithm A_ε. More precisely, if the average error of A is at most ε then

$$\text{card}^{\text{avg}}(A) \geq \frac{1}{\ln \alpha} \cdot \ln \ln(1/\varepsilon) + c_\alpha,$$

for any α with $\alpha > r + 1/2$ and $r \geq 2$.

The stopping rule "$t_i - s_i \leq 2\varepsilon$" in the algorithm A_ε is adaptive, since the number $n(f)$ of function evaluations depends on $f \in F_r$. Such an adaptive stopping rule is

crucial. It is known that with a non-adaptive stopping rule

$$n(f) = n \quad \text{for all } f \in F_r$$

one can only achieve a linear convergence, i.e., there exists a number $\beta \in (0, 1)$ such that for any algorithm using n information operations we have

$$e^{\text{avg}}(A) \geq \beta^n.$$

Hence to guarantee an average error ε we must have

$$n \geq \frac{\ln 1/\varepsilon}{\ln 1/\beta}.$$

This means that with fixed cardinality the minimal cost is of order $\ln \varepsilon^{-1}$, whereas with varying cardinality the minimal cost is of order $\ln \ln(1/\varepsilon)$. Note the exponential difference between the orders $\ln(1/\varepsilon)$ and $\ln \ln(1/\varepsilon)$. This difference is due to the fact that when we switch from the worst case cost to average cost, the difference between worst case errors and average case errors turns out to be insignificant for this particular nonlinear problem.

We add in passing that *global optimization* is another nonlinear problem where adaptive algorithms are much better than nonadaptive algorithms, see the recent paper of Calvin [26]. □

In this book we study the average case setting mainly for *linear problems with a Gaussian measure*. In Appendix B, we list major properties of Gaussian measures. We shall see that for such problems, algorithms with varying cardinality are *not* much better in terms of polynomial or weak tractability than algorithms with fixed cardinality. For simplicity, we restrict ourselves to spaces over the real field. In Appendix B it is mentioned how to deal with the complex field as well.

We now assume that $F = \widetilde{F}$ is a separable real Banach space that is equipped with a zero-mean Gaussian probability measure. Let $S \colon F \to G$ be a bounded linear operator into a separable real Hilbert space G. We always assume that the set Λ is a set of continuous linear functionals, hence $\Lambda \subseteq \Lambda^{\text{all}} = F^*$. Then the a priori measure $\nu = \mu S^{-1}$ of solution elements is also a Gaussian measure. Its mean is 0 and its correlation operator $C_\nu \colon G^* = G \to G$ is given by

$$C_\nu g = S(C_\mu(L_g S)) \quad \text{for all } g \in G,$$

where $L_g(h) = \langle g, h \rangle$ with the inner product in G. Note that $L_g S = L_g \circ S$ is an element of F^* and $(L_g S)(f) = \langle g, S(f) \rangle$. We have

$$\text{trace}(C_\nu) = \int_F \|S(f)\|^2 \mu(\mathrm{d}f) < \infty.$$

Consider first non-adaptive information $N = [L_1, L_2, \ldots, L_n]$, where $L_i \in F^*$. Without loss of generality we may assume that the L_i are orthonormal with respect to the scalar product given by μ, i.e.,

$$\langle L_i, L_j \rangle_\mu := L_i(C_\mu L_j) = \delta_{i,j} \quad \text{for all } i, j \in \{1, 2, \ldots, n\}.$$

Let $\mu_1 = \mu N^{-1}$ be the image measure on $N(F) = \mathbb{R}^n$. Then μ_1 is Gaussian with mean 0. The correlation operator of μ_1 is the identity, due to the μ-orthonormality of the L_i. Hence

$$\mu_1(A) = \frac{1}{(2\pi)^{n/2}} \int_A \exp\left(-\tfrac{1}{2} \sum_{j=1}^{n} t_j^2\right) dt_1\, dt_2 \dots dt_n.$$

Let $y = N(f) \in \mathbb{R}^n$. Then the conditional measure $\mu_2(\,\cdot \mid y)$ defined on the Borel sets of F and concentrated on $N^{-1}(y)$ is also Gaussian. Its mean is

$$m_{\mu,y} = \sum_{j=1}^{n} y_j C_\mu L_j$$

and its correlation operator $C_{\mu,N} : F^* \to F$ is given by

$$C_{\mu,N}(L) = C_\mu L - \sum_{j=1}^{n} \langle L, L_j\rangle_\mu\, C_\mu L_j \quad \text{for all } L \in F^*.$$

Observe that $C_{\mu,N}$ does *not* depend on y, and therefore the local average error of an optimal algorithm does not depend on y. As a consequence, we do not have "easy" or "difficult" information $y = N(f) \in \mathbb{R}^n$ in the average case setting. This should be contrasted with the worst case setting in which zero information was the most difficult and the local errors and local radii depend on $y = N(f)$,

Observe that $C_{\mu,N}(L_i) = 0$ for $i = 1, 2, \dots, n$ and $C_{\mu,N}(L) = C_\mu L$ for all functionals L from F^* which are μ-orthogonal to the L_i, i.e., for which

$$\langle L, L_i\rangle_\mu = L(C_\mu L_i) = 0 \quad \text{for all } i = 1, 2, \dots, n.$$

This means that the information $y = N(f)$ changes the a priori measure μ by shifting the mean element from 0 to $m_{\mu,y}$ and by annihilating the correlation operator C_μ in the linear subspace spanned by $L_1, L_2, \dots, L_n$.

The a posteriori measure $\nu(\,\cdot \mid y) = \mu_2(S^{-1} \cdot \mid y)$ is defined on the Borel sets of G. It is also Gaussian with mean

$$m(y) = S m_{\mu,y} = \sum_{j=1}^{n} y_j S(C_\mu L_j).$$

Its correlation operator does not depend on y and is given by

$$C_{\nu,N}(g) = S(C_{\mu,N}(L_g S)) = C_\nu g - \sum_{j=1}^{n} \langle g, S C_\mu L_j\rangle\, S C_\mu L_j$$

for all $g \in G$.

To minimize the local average error we use the *central algorithm* φ^{cen} that is the mean of the conditional measure $\nu(\cdot \mid y)$,

$$\varphi^{\text{cen}}(y) = m(y) = S m_{\mu,y} = \sum_{j=1}^{n} y_j S(C_\mu L_j).$$

The local average radius $r^{\text{avg}}(N, y)$ does not depend on y and is equal to

$$r^{\text{avg}}(N, y)^2 = \int_G \|g\|^2 \nu(\mathrm{d}g \mid 0) = \text{trace}(C_{\nu,N}) = \text{trace}(C_\nu) - \sum_{j=1}^{n} \|S(C_\mu L_j)\|^2.$$

We summarize this in the following theorem.

Theorem 4.28. *Let F be a separable real Banach space equipped with a zero-mean Gaussian measure and a correlation operator C_μ. Let $S: F \to G$ be a bounded linear operator into a separable real Hilbert space G and $\nu = \mu S^{-1}$ be a Gaussian measure on solution elements. Consider non-adaptive information $N = [L_1, L_2, \ldots, L_n]$ with $L_i \in F^*$ and $L_i(C_\mu L_j) = \delta_{i,j}$. Then the algorithm*

$$\varphi^{\text{cen}}(y) = \sum_{j=1}^{n} y_j S(C_\mu L_j)$$

is central and optimal, with average case error

$$r^{\text{avg}}(N) = \left(\text{trace}(C_\nu) - \sum_{j=1}^{n} \|S(C_\mu L_j)\|^2 \right)^{1/2}.$$

We give the following interpretation of this result, see again [242] for more details. Before the computation, we only know the measure μ on F and the image measure $\nu = \mu S^{-1}$ on G. The best approximation to solution elements $g = S(f)$ is 0 as the mean of the a priori measure ν. The average radius of the zero information is then

$$r^{\text{avg}}(0) = e^{\text{avg}}(0) = \sqrt{\text{trace}(C_\nu)}.$$

This is the *initial error* in the average case setting.

After computing $y = N(f)$, the best approximation is the mean of the a posteriori measure $\nu(\cdot \mid y)$ and its error is given by the trace of $C_{\nu,N}$.

Assume now that $S: F \to \mathbb{R}$ is a linear functional. Then ν and $\nu(\cdot \mid y)$ are 1-dimensional Gaussian measures and $r^{\text{avg}}(N) = \sqrt{\text{trace}(C_{\nu,N})}$ is the standard deviation of the conditional measure $\nu(\cdot \mid y)$. The initial error can be written as

$$e^{\text{avg}}(0) = \sqrt{\text{trace}(C_\nu)} = \left(\int_F |S(f)|^2 \mu(\mathrm{d}f) \right)^{1/2} = \sqrt{S(C_\mu S)} = \|S\|_\mu.$$

Since $\langle L_i, L_j \rangle = \delta_{i,j}$, we can rewrite the average radius of information as

$$r^{\mathrm{avg}}(N) = \sqrt{\|S\|_\mu^2 - \sum_{i=1}^n \langle S, L_i \rangle_\mu^2} = \left\| S - \sum_{i=1}^n \langle S, L_i \rangle_\mu L_i \right\|_\mu$$

$$= \inf_{g_i \in G} \left\| S - \sum_{i=1}^n g_i L_i \right\|_\mu.$$

The same is true in the general case when G is an arbitrary separable Hilbert space. For a bounded linear operator $Q \colon F \to G$ we define

$$\|Q\|_\mu := \left(\int_F \|Q(f)\|^2 \mu(\mathrm{d}f) \right)^{1/2}.$$

Then the initial error can be expressed as

$$e^{\mathrm{avg}}(0) = \left(\int_F \|S(f)\|^2 \mu(\mathrm{d}f) \right)^{1/2} = \|S\|_\mu.$$

We denote by $g_i L_i$ the operator that is defined by $(g_i L_i)(f) = L_i(f) g_i$ and obtain the following corollary.

Corollary 4.29. *The average radius of a non-adaptive information N equals the error of the best approximation of S by linear combinations of the L_i in the norm depending on the measure μ, i.e.,*

$$r^{\mathrm{avg}}(N) = \inf_{g_i \in G} \left\| S - \sum_{i=1}^n g_i L_i \right\|_\mu.$$

We now present the optimal non-adaptive information for $\Lambda = \Lambda^{\mathrm{all}}$. Let

$$N = [L_1, L_2, \dots, L_n]$$

with $L_i \in \Lambda^{\mathrm{all}}$ and assume as before that $L_i(C_\mu L_j) = \delta_{i,j}$. Then Theorem 4.28 implies that we should maximize $\sum_{j=1}^n \|S(C_\mu L_j)\|^2$.

Let $\{\eta_i^*\}$ be the orthonormal eigenelements of the correlation operator $C_\nu \colon G \to G$, where $\nu = \mu S^{-1}$ is the a priori measure of solution elements, i.e.,

$$C_\nu \eta_i^* = \lambda_i^* \eta_i^*,$$

where $\lambda_1^* \geq \lambda_2^* \geq \cdots \geq 0$. Without loss of generality we assume that $\lambda_i^* > 0$. Define L_i^* by

$$L_i^*(f) = (\lambda_i^*)^{-1/2} \langle S(f), \eta_i^* \rangle.$$

Then the information

$$N_n^*(f) = [L_1^*(f), L_2^*(f), \dots, L_n^*(f)]$$

is an nth optimal non-adaptive information for the class Λ^{all} with radius

$$r^{\mathrm{avg}}(N_n^*) = \sqrt{\sum_{j=n+1}^{\infty} \lambda_i^*}.$$

Moreover, the L_i are μ-orthonormal, $L_i^*(C_\mu L_j^*) = \delta_{i,j}$.

It was proved by Wasilkowski [260] that adaptive information with fixed cardinality does not help and that adaptive information N with varying cardinality $\mathrm{card}^{\mathrm{avg}}(N) \leq n$ is not better than the nth optimal non-adaptive information for the class Λ^{all}. We summarize these results in the following theorem, see [242] for the details.

Theorem 4.30. *Let the assumptions of Theorem 4.28 hold. Then*

$$e^{\mathrm{avg}}(n, \Lambda^{\mathrm{all}}) = \sqrt{\sum_{j=n+1}^{\infty} \lambda_i^*}.$$

The nth optimal information in the class Λ^{all} is non-adaptive, and is given by

$$N_n^*(f) = [L_1^*(f), L_2^*(f), \dots, L_n^*(f)],$$

with

$$L_i^*(f) = (\lambda_i^*)^{-1/2} \langle S(f), \eta_i^* \rangle \quad \text{for } i = 1, 2, \dots, n. \qquad \square$$

We briefly discuss information classes Λ that are proper subsets of F^*. We can apply Corollary 4.29 to obtain the following result.

Corollary 4.31. *The information $N_n^* = [L_1^*, L_2^*, \dots, L_n^*]$ is nth optimal in the class Λ iff $L_i^* \in \Lambda$ and*

$$\inf_{g_i \in G} \left\| S - \sum_{i=1}^n g_i L_i^* \right\|_\mu = \inf_{L_i \in \Lambda} \inf_{g_i \in G} \left\| S - \sum_{i=1}^n g_i L_i \right\|_\mu.$$

In general, it is difficult to find L_i^* satisfying the last corollary. For the class $\Lambda = \Lambda^{\mathrm{std}}$, the functionals L_i^* are known only in rare cases. However, we often know linear functionals from Λ^{std} with the same order of convergence as L_i^*.

Consider now adaptive information with linear functionals from Λ. We can distinguish between information with fixed cardinality n and information with varying cardinality $\mathrm{card}^{\mathrm{avg}}(N) \leq n$. We summarize some results of Wasilkowski [260] on adaption, see [242] for the details.

Theorem 4.32. *Let the assumptions of Theorem 4.28 hold.*

a) *Adaptive information with fixed cardinality n is not better than the nth non-adaptive information.*

b) *Information with varying cardinality can be better than information with fixed cardinality. If, however, the squared errors $e^{\mathrm{avg}}(\varphi_n, N_n)^2$ of optimal linear algorithms $\varphi_n \circ N_n$ using the nth optimal non-adaptive information form a convex sequence, then adaptive information of $\mathrm{card}^{\mathrm{avg}}(N) \leq n$ is not better than the nth optimal non-adaptive information.*

Hence, the only case for which adaptive information with varying cardinality can help for the class $\Lambda \subseteq \Lambda^{\mathrm{all}}$ is when the squares of the nth minimal radii $r^{\mathrm{avg\text{-}non}}(n)$ of non-adaptive information form a non-convex sequence. In this case, Plaskota [194] showed there are linear problems and classes Λ for which adaption with varying cardinality can significantly help for some ε. More precisely, let $n^{\mathrm{avg\text{-}non}}(\varepsilon)$ denote the minimal number of nonadaptive information operations to guarantee that the average case error is at most ε, and $n^{\mathrm{avg\text{-}ada}}(\varepsilon)$ denote the minimal expected number of varying cardinality adaptive information operations to guarantee that the average case error is at most ε. Then there exists a linear problem and a class Λ such that for any arbitrarily small positive number c there exists a positive ε such that

$$n^{\mathrm{avg\text{-}ada}}(\varepsilon) \leq c n^{\mathrm{avg\text{-}non}}(\varepsilon).$$

On the other hand, this gain for a specific ε does not really help for tractability. Due to Theorem 7.7.2 from [242] we know that for all positive ε we have

$$\sup_{x>1} \min \left(n^{\mathrm{avg\text{-}non}}(x\varepsilon), \frac{x^2-1}{x^2} n^{\mathrm{avg\text{-}non}}(\varepsilon) \right) \leq n^{\mathrm{avg\text{-}ada}}(\varepsilon) \leq n^{\mathrm{avg\text{-}non}}(\varepsilon).$$

In particular, take $x = 2$. Then

$$\tfrac{3}{4} n^{\mathrm{avg\text{-}non}}(2\varepsilon) \leq n^{\mathrm{avg\text{-}ada}}(\varepsilon) \leq n^{\mathrm{avg\text{-}non}}(\varepsilon) \quad \text{for all } \varepsilon > 0.$$

This implies essentially the same bounds for polynomial tractability if we use non-adaptive information or adaptive information with varying cardinality. Clearly, weak tractability is also equivalent for these two cases.

4.3.2 Probabilistic Setting

In this section we describe some results for the probabilistic setting; more details can be found in [242]. We discuss results for bounded linear problems $S\colon F \to G$ on a separable real Banach space F that is equipped with a zero-mean Gaussian measure μ, and G is a separable real Hilbert space.

Recall that we want to approximate $S(f)$ knowing $y = N(f) \in \mathbb{R}^n$, where N is a measurable information operator defined as in Section 4.1.1. Hence we assume that N is given by (4.9). As discussed before, the probability measure $\nu = \mu S^{-1}$ is the a priori measure of solution elements $S(f) \in G$ and is also Gaussian.

For a measurable approximation $\varphi \circ N \colon F \to G$, we define the error in the probabilistic setting by

$$e^{\mathrm{prob}}(\varphi, N, \delta) = \inf_{M:\mu(M)\leq\delta} \sup_{f\in F\setminus M} \|S(f) - \varphi(N(f))\|.$$

Here $\delta \in [0, 1)$ is a parameter that controls the measure of a set for which we ignore the size of the error $\|S(f) - \varphi(N(f))\|$. Clearly, the error in the probabilistic case is a non-increasing function of δ.

Let N be a non-adaptive information, $N = [L_1, L_2, \ldots, L_n]$, with μ-orthogonal functionals L_j. We know from Section 4.3.1 that $\nu(\cdot \mid y)$ is Gaussian with mean $m(y) = \sum_{j=1}^{n} S(C_\mu L_j)$ and correlation operator

$$C_{\nu,N} = C_\nu - \sum_{j=1}^{n} \langle \cdot, S(C_\mu L_j) \rangle S(C_\mu L_j),$$

where C_ν is the correlation operator of ν and $\langle \cdot, \cdot \rangle$ stands for the inner product of G. It turns out that the linear algorithm

$$\varphi^*(y) = \sum_{j=1}^{n} y_j S(C_\mu L_j),$$

which is optimal in the average case setting, is also optimal in the probabilistic setting. Its error is the same as the radius of information N in the probabilistic setting, with

$$r^{\text{prob}}(N, \delta) = e^{\text{prob}}(\varphi^*, N, \delta) = \inf\{x \mid \nu_N(B(0, x)) \geq 1 - \delta\}. \tag{4.35}$$

Here ν_N is Gaussian with mean 0 and correlation operator $C_{\nu,N}$, and $B(0, x)$ denotes the ball with center 0 and radius x of the space G, i.e., $B(0, x) = \{g \in G \mid \|g\| \leq x\}$. Hence, the radius of information in the probabilistic setting is the smallest radius of the ball whose measure is at least $1 - \delta$.

Before we start any computation, which formally corresponds to zero information $N = 0$, the best approximation to solution elements $g = S(f)$ is 0 as the mean of $\nu = \mu S^{-1}$. Hence the *initial error in the probabilistic setting* is

$$r^{\text{prob}}(0, \delta) = e^{\text{prob}}(0) = \inf\{x \mid \nu(B(0, x)) \geq 1 - \delta\}. \tag{4.36}$$

After the computation of $y = N(f)$, the measure ν has changed to the conditional measure $\nu(\cdot \mid y)$ and we obtain the new radius (4.35). Observe that both ν and ν_N are zero-mean Gaussian measures. They differ by their correlation operators and most importantly, by their traces. We have

$$\text{trace}(C_{\nu,N}) = \text{trace}(C_\mu) - \sum_{j=1}^{n} \|S(C_\mu L_j)\|^2.$$

We explain in Appendix B that the Gaussian measure of the ball of radius $x > 0$ goes to 1 if the trace of the Gaussian measure goes to 0. This means that the radius in the probabilistic setting also goes to 0 as the trace approaches 0.

Assume now that $S \colon F \to \mathbb{R}$ is a bounded linear functional. Then the measure ν_N is 1-dimensional Gaussian with mean 0 and variance

$$\sigma(N) = \inf_{g_j \in \mathbb{R}} \left\| S - \sum_{j=1}^{n} g_j L_j \right\|_\mu^2.$$

As we know from Section 4.3.1, the average radius of information N with the absolute error criterion is given by $r^{\mathrm{avg}}(N) = \sqrt{\sigma(N)}$. Note that

$$\nu_N(B(0, x)) = \sqrt{\frac{2}{\pi \sigma(N)}} \int_0^x \exp\left(\frac{-t^2}{2\sigma(N)}\right) dt = \psi\left(\frac{x}{r^{\mathrm{avg}}(N)}\right),$$

where

$$\psi(z) = \sqrt{\frac{2}{\pi}} \int_0^x \exp(-t^2/2)\, dt$$

is the probability integral. This yields

Corollary 4.33. *Let* $S\colon F \to \mathbb{R}$ *be a bounded linear functional and* $N\colon F \to \mathbb{R}^n$ *be a non-adaptive information. Then*

$$r^{\mathrm{prob}}(N, \delta) = \psi^{-1}(1 - \delta)\, r^{\mathrm{avg}}(N).$$

For $\delta \to 0$ *it follows that*

$$r^{\mathrm{prob}}(N, \delta) = \sqrt{2 \ln(1/\delta)}\, r^{\mathrm{avg}}(N)\, (1 + o(1)).$$

We return to S being a bounded linear operator. There is a tight link between the average case setting and the probabilistic setting for linear problems $S\colon F \to G$, since the same non-adaptive information $N^* = [L_1^*, L_2^*, \ldots, L_n^*]$ is optimal. In the probabilistic setting adaption *never* helps, while in the average case setting adaption may help through varying cardinality.

For the absolute error criterion, we want to find the minimal n for which we solve the problem to within ε. That is, this minimal number is now equal to

$$n^{\mathrm{prob\text{-}abs}}(\varepsilon, \delta, S) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{prob}}(A_n; \delta) \le \varepsilon\}.$$

Let

$$n^{\mathrm{avg\text{-}abs}}(\varepsilon, S) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{avg}}(A_n) \le \varepsilon\} \qquad (4.37)$$

be the corresponding minimal cardinality of information needed to solve the same problem to within ε with the absolute error criterion in the average case setting. In both these definitions we consider information with fixed cardinality n. From the last corollary we obtain the following one.

Corollary 4.34. *Let* $S\colon F \to \mathbb{R}$ *be a bounded linear functional. Then*

$$n^{\mathrm{prob\text{-}abs}}(\varepsilon, \delta, S) = n^{\mathrm{avg\text{-}abs}}(\varepsilon/\psi^{-1}(1 - \delta), S).$$

Thus, the minimal cardinality in the probabilistic setting is the same as the minimal cardinality in the average case setting if we replace ε by $\varepsilon/\psi^{-1}(1 - \delta)$. Note that $\psi(z) = 1 - \sqrt{2/\pi}z^{-1}\exp(-z^2/2)(1 + o(1))$ as z goes to infinity. This implies that

$$\psi^{-1}(1 - \delta) = \sqrt{2 \ln \delta^{-1}}\, (1 + o(1)) \quad \text{as } \delta \to 0.$$

Hence, for small δ, we have

$$\varepsilon/\psi^{-1}(1-\delta) \approx \varepsilon/\sqrt{2\ln\delta^{-1}}.$$

Assuming again that S is a bounded linear operator, we turn to the normalized error criterion. We now want to find the minimal number n for which we reduce the initial error by a factor ε, so that

$$n^{\text{prob-nor}}(\varepsilon, \delta, S) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\text{prob}}(A_n; \delta) \le \varepsilon e^{\text{prob}}(0; \delta)\}.$$

Obviously,

$$n^{\text{prob-nor}}(\varepsilon, \delta, S) = n^{\text{prob-abs}}(\varepsilon e^{\text{prob}}(0; \delta), \delta, S).$$

The essence of Corollary 4.34 is that for bounded linear functionals, the probabilistic setting can be analyzed fully by the average case setting. In this volume we will study only the class Λ^{all}, the study of the class Λ^{std} is deferred to Volumes II and III. Obviously, for the class Λ^{all}, bounded linear functionals are trivial and can be solved exactly by using at most one information operation. This means, in particular, that the study of the probabilistic setting for the class Λ^{all} makes sense only for bounded linear operators. Not much is known about tractability in this setting. We are only aware of the papers of Lifshits and Tulyakova [135] and Lifshits and Zani [136] on this subject. That is why the reader will not find a tractability study in the probabilistic setting for bounded linear operators and for the class Λ^{all} in Volume II.

4.3.3 Randomized Algorithms

Many problems are difficult to solve with deterministic algorithms. One may hope that randomized algorithms make many problems much easier. It is natural to ask for which problems does randomization help. This question has been extensively studied and there exist many positive as well as negative answers. The most classical result is probably due to Bakhvalov [7]. The optimal rate of convergence for the integration problem for C^k-functions on $[0, 1]^d$ is $n^{-k/d}$ for deterministic algorithms in the worst case setting and $n^{-k/d-1/2}$ for randomized algorithms, see NR 4.3.3:1 for more references.

In this section we consider the approximation of $S: F \to G$ by randomized algorithms. Again we assume that F is a subset of a normed space $\tilde{F}$ and also G is a normed space.

The model of computation can be formalized in slightly different ways, see the work of Heinrich [79], [80], [81], Mathé [141], Nemirovsky and Yudin [155], Wasilkowski [262] and [160], [164], [242]. We do not explain all the technical details of these models and only give a reason why it makes sense to study different models for upper and lower bounds, respectively.

- Assume that we want to construct and to analyze concrete algorithms that yield upper bounds for the (total) complexity of given problems including the arithmetic cost and the cost of the random number generator. Then it is reasonable

to consider a more restrictive model of computation where, for example, only the standard arithmetic operations are allowed. One may also restrict the use of random numbers and study so-called *restricted Monte Carlo methods*, where only random bits are allowed, see the Ph.D. thesis of H. Pfeiffer [190], Gao, Ye and Wang [61], as well as see [90], [164], [168].

- For the proof of lower bounds we take the opposite view and allow general randomized mappings of the form (4.12) and a very general kind of randomness. This makes the lower bounds stronger.

It turns out that the results are often very robust with respect to changes of the computational model. For the purpose of this book, we define *randomized algorithms* as follows.

Definition 4.35. A *randomized algorithm A* is a pair consisting of a probability space (Ω, Σ, μ) and a family $(N_\omega, \varphi_\omega)_{\omega \in \Omega}$ of mappings such that the following holds:

1. For each fixed ω, the mapping $A_\omega = \varphi_\omega \circ N_\omega$ is a deterministic algorithm as in Section 4.1, see (4.3), (4.4) and (4.9), based on adaptive information from the class Λ.

2. Let $n(f, \omega)$ be the cardinality of the information N_ω for $f \in F$. Then the function $\omega \mapsto n(f, \omega)$ is measurable for each fixed f.

Let A be a randomized algorithm. Then

$$n(A) = \sup_{f \in F} \int_\Omega n(f, \omega)\, \mu(\mathrm{d}\omega)$$

is called the *cardinality of A* and

$$e^{\mathrm{ran}}(A) = \sup_{f \in F} \left(\int_\Omega^* \| S(f) - \varphi_\omega(N_\omega(f)) \|^2\, \mu(\mathrm{d}\omega) \right)^{1/2}$$

is called the *error* of A. By $\int^*$ we denote the upper integral. For $n \in \mathbb{N}$, define

$$e^{\mathrm{ran}}(n) = \inf\{e^{\mathrm{ran}}(A) \mid n(A) \leq n\}.$$

Remark 4.36. If $A \colon F \to G$ is a deterministic algorithm then A can also be treated as a randomized algorithm with respect to a Dirac (atomic) measure μ. In this sense we can say that deterministic algorithms are special randomized algorithms. Hence the inequality

$$e^{\mathrm{ran}}(n) \leq e^{\mathrm{wor}}(n) \tag{4.38}$$

is trivial.

The number $e^{\mathrm{ran}}(0)$ is called the *initial error in the randomized setting*. For $n = 0$, we do not sample f, and φ_ω is independent of f, but may depend on ω. It is easy to check that for a linear S, the best we can do is to take $\varphi_\omega \equiv 0$ and then

$$r^{\mathrm{ran}}(0) = e^{\mathrm{wor}}(0).$$

This means that for linear problems the initial errors are the same in the worst case and randomized settings.

For problems that are *not* linear, it may, however, happen that the initial error in the randomized setting is smaller than in the worst case setting. For such problems we can take F and G as not necessarily normed spaces. For instance, take $F = G = \{-\frac{1}{2}, \frac{1}{2}\} \subseteq \mathbb{R}$ as the two element set, and $Sf = f$. Since the values of φ_ω must be in G, we can only have $\varphi_\omega = -\frac{1}{2}$ or $\varphi_\omega = \frac{1}{2}$. This easily allows us to check that the initial error in the worst case setting is just 1, whereas the initial error in the randomized setting is $2^{-1/2}$ and is achieved by taking the uniform distribution, i.e., $\Omega = G$ and both $-\frac{1}{2}$ and $\frac{1}{2}$ occur with probability $\frac{1}{2}$.

It is also possible to make the initial error in the randomized setting arbitrarily small whereas the initial error in the worst case is still 1. This holds if we agree to measure the distance between different elements not by a norm. We give an artificial example. Let $F = G = \{f_1, f_2, \ldots, f_m\}$ for distinct f_i, and $Sf = f$. Define the local error of an algorithm $A\colon F \to F$ by $e(A(f_i), f_i) = 1$ if $A(f_i) = f_i$ and $e(A(f_i), f_i) = 0$ otherwise. This corresponds to the problem where we want to compute an f_j different from the input f_i. Then any constant deterministic algorithm has error 1, while the initial error in the randomized setting is just $m^{-1/2}$, and it is achieved for the uniform distribution on F. Since m can be arbitrarily large, the initial error in the randomized setting can be indeed arbitrarily small. $\qquad\square$

To prove lower bounds for randomized algorithms, there basically exists only one proof technique which goes back to Bakhvalov [7]. The main point is to observe that the errors in the randomized setting cannot be smaller than the errors in the average case setting. For this technique one has to prove lower bounds on the average case error (for deterministic algorithms) and for a "bad" probability measure ϱ that leads to large error bounds.

More specific, let ϱ be a probability (Borel-) measure on F and let $A = \varphi \circ N$ be a deterministic algorithm. Then we define

$$e^{\mathrm{avg}}(A, \varrho) = \left(\int_F^* \|S(f) - A(f)\|^2\, \varrho(\mathrm{d}f) \right)^{1/2}$$

and

$$e^{\mathrm{avg}}(n, \varrho) = \inf_{A = \varphi \circ N} e^{\mathrm{avg}}(A, \varrho),$$

where the infimum runs over all (deterministic) algorithms using information from the given class Λ with

$$n^{\mathrm{avg}}(A, \varrho) = \int_F^* n(f)\, \varrho(\mathrm{d}f) \le n.$$

We assume that ϱ is atomic in the sense that

$$\varrho(M) = \sum_{i=1}^{k} c_i 1_M(f_i), \tag{4.39}$$

where $f_1, f_2, \ldots, f_k \in F$ and $c_i > 0$ with $\sum_{i=1}^{k} c_i = 1$, for some k depending on n. Here, 1_M is the indicator (characteristic) function of M. Then one can use the following lemma.

Lemma 4.37. *Assume that $S \colon F \to G$ is measurable and ϱ is an atomic probability measure on F as in* (4.39)*. Then*

$$e^{\mathrm{ran}}(n) \geq \frac{\sqrt{2}}{2} e^{\mathrm{avg}}(2n, \varrho). \tag{4.40}$$

Proof. Let A be a randomized algorithm with $n(A) \leq n$. Hence

$$n \geq \sup_{f \in F} \int_{\Omega} n(f, w) \, \mu(\mathrm{d}\omega) \geq \int_{F} \int_{\Omega}^{*} n(f, \omega) \mu(\mathrm{d}\omega) \, \varrho(\mathrm{d}f)$$

$$\geq \int_{\Omega}^{*} \int_{F} n(f, \omega) \, \varrho(\mathrm{d}f) \, \mu(\mathrm{d}\omega) = \int_{\Omega}^{*} n^{\mathrm{avg}}(N_{\omega}, \varrho) \, \mu(\mathrm{d}\omega).$$

We set $\Omega_0 = \{\omega \in \Omega \mid n^{\mathrm{avg}}(N_{\omega}, \varrho) \leq 2n\}$ and conclude by Chebyshev's inequality that $\mu(\Omega_0) \geq \frac{1}{2}$. Now we have

$$e(A)^2 = \sup_{f \in F} \int_{\Omega}^{*} \|S(f) - \varphi_{\omega}(N_{\omega}(f))\|^2 \, \mu(\mathrm{d}\omega)$$

$$\geq \int_{F} \int_{\Omega}^{*} \|S(f) - \varphi_{\omega}(N_{\omega}(f))\|^2 \, \mu(\mathrm{d}\omega) \, \varrho(\mathrm{d}f)$$

$$\geq \int_{\Omega}^{*} \int_{F} \|S(f) - \varphi_{\omega}(N_{\omega}(f))\|^2 \, \varrho(\mathrm{d}f) \, \mu(\mathrm{d}\omega)$$

$$= \int_{\Omega} e^{\mathrm{avg}}(A_{\omega}, \varrho)^2 \, \mu(\mathrm{d}\omega) \geq \mu(\Omega_0) \cdot \inf_{\omega \in \Omega_0} e^{\mathrm{avg}}(A_{\omega}, \varrho)^2 \geq \frac{1}{2} e^{\mathrm{avg}}(2n, \varrho)^2.$$

This proves the lemma. $\qquad\qquad\square$

We want to prove lower bounds for $e^{\mathrm{ran}}(n)$ for the case where $S \colon \widetilde{F} \to G$ is a bounded linear operator between Hilbert spaces and $\Lambda = \Lambda^{\mathrm{all}}$. Unfortunately, Lemma 4.37 in its present form cannot be used, since it would only yield the trivial lower bound $e^{\mathrm{ran}}(n) \geq 0$. For technical reasons we define *measurable* randomized algorithms as follows. Here we assume that also $S \colon \widetilde{F} \to G$ is measurable.

Definition 4.38. A *measurable randomized algorithm A* is a pair consisting of a probability space (Ω, Σ, μ) and a family $(N_{\omega}, \varphi_{\omega})_{\omega \in \Omega}$ of mappings such that the following hold:

1. For each fixed ω, the mapping $A_\omega = \varphi_\omega \circ N_\omega$ is a deterministic algorithm as in Section 4.1, see (4.3), (4.4) and (4.9), based on adaptive information from the class Λ.

2. Let $n(f, \omega)$ be the cardinality of the information N_ω for $f \in F$. Then the function n is measurable.

3. The mapping $(f, \omega) \mapsto \varphi_\omega(N_\omega(f)) \in G$ is measurable.

Remark 4.39. If μ is a Dirac measure then we obtain a deterministic algorithm with two properties:

1. The function $n \colon F \to \mathbb{N} \cup \{\infty\}$ is measurable.

2. The mapping $A = \varphi \circ N$ is measurable.

If, on the other hand, $A \colon F \to G$ is a deterministic algorithm with these two properties then A can also be considered as a measurable randomized algorithm. This leads to an open problem, although this is a problem concerning only deterministic algorithms. $\qquad\square$

Open Problem 19.

- Assume that $S \colon F \to G$ is measurable and Λ is a class of admissible information functionals. Define

$$e^{\mathrm{wor}}(n) = \inf_{A_n} e^{\mathrm{wor}}(A_n) \tag{4.41}$$

as in Section 4.1.4. Under which assumptions can we replace the set of all (also non-measurable) algorithms A_n by the set of Borel measurable algorithms without changing the infimum?

To solve the last open problem one can probably apply measurable selection theorems, see the survey of Wagner [255] and [169]. A partial solution to Open Problem 19 can be obtained if $S \colon \widetilde{F} \to G$ is linear and continuous and F is the unit ball of $\widetilde{F}$, both $\widetilde{F}$ and G are Banach spaces. The quantity $e^{\mathrm{wor}}(n)$ increases at most by a factor of 2 if we only allow non-adaptive information N_n and by another factor of 2 if we only allow continuous algorithms φ_n. The last relation is a result of Mathé [140, Theorem 11 (v)]. Hence we obtain at least the inequality

$$e^{\mathrm{wor}}(n) \geq \frac{1}{4} \inf_{A_n \text{ continuous}} e^{\mathrm{wor}}(A_n).$$

Hence (4.41) holds modulo a factor of $1/4$. We believe that this factor is *not* needed and measurable algorithms are always optimal, at least if $\widetilde{F}$ and G are separable.

Definition 4.40. We define

$$e^{\mathrm{ran\text{-}meas}}(n) = \inf\{e^{\mathrm{ran}}(A) \mid n(A) \leq n\},$$

where A is a *measurable* randomized algorithm.

Remark 4.41. If we restrict ourselves to measurable randomized algorithms, then it is easy to extend the inequality of Lemma 4.37, i.e.,

$$e^{\text{ran-meas}}(n) \geq \frac{\sqrt{2}}{2} e^{\text{avg}}(2n, \varrho), \tag{4.42}$$

for arbitrary (Borel) probability measures on F. The proof is the same as the proof of Lemma 4.37. □

Now we assume in addition that $S: \widetilde{F} \to G$ is a bounded linear operator and F is the unit ball of $\widetilde{F}$. We also assume that $\Lambda = \Lambda^{\text{all}}$. Then Heinrich [79] proved the lower bound

$$e^{\text{ran-meas}}(n) \geq c \cdot \frac{a_{2n}(S, \varrho)}{\int_F \|x\| \, \varrho(\mathrm{d}\,x)}.$$

Here $c > 0$ is an absolute constant and ϱ is a (centered) Gaussian measure on $\widetilde{F}$, while $a_{2n}(S, \varrho)$ is the average case error of an optimal linear algorithm with cardinality $2n$.

We can say a little more when $\widetilde{F}$ and G are Hilbert spaces, see [161], where only the case with $n(f, \omega) = n \in \mathbb{N}$ was studied. For the sake of simplicity we assume that the problem is solvable, i.e., $\lim_{n \to \infty} e^{\text{wor}}(n) = 0$. Solvability holds iff S is compact.

Theorem 4.42. *Assume that $S: \widetilde{F} \to G$ is a compact linear operator between Hilbert spaces $\widetilde{F}$ and G with $\Lambda = \Lambda^{\text{all}}$. Then*

$$e^{\text{ran-meas}}(n) \geq \tfrac{1}{2} e^{\text{wor}}(4n - 1). \tag{4.43}$$

Proof. We know that $S(e_i) = \sigma_i \tilde{e}_i$ with orthonormal $\{e_i\}$ in $\widetilde{F}$ and $\{\tilde{e}_i\}$ in G, see Section 4.2.3. From Theorem 4.11 we also know that $e^{\text{wor}}(n) = \sigma_{n+1}$.

For $m > n$, consider the normed $(m - 1)$-dimensional Lebesgue measure ϱ_m on the unit sphere $E_m = \left\{ \sum_{i=1}^m \alpha_i e_i \mid \alpha_i \in \mathbb{R}, \ \sum_{i=1}^m \alpha_i^2 = 1 \right\}$. Then

$$S_n^* \left(\sum_{i=1}^\infty \alpha_i e_i \right) = \sum_{i=1}^n \sigma_i \alpha_i \tilde{e}_i$$

is the optimal algorithm using linear information of cardinality n. This is true for the worst case, with error σ_{n+1}, as well as for the average error with respect to ϱ_m, as follows from the results of [241], [266]. Hence

$$e^{\text{avg}}(n, \varrho_m)^2 = \int_{E_m} \sum_{i=n+1}^m \sigma_i^2 \alpha_i^2 \, \varrho_m(\mathrm{d}\alpha).$$

Since $\int_{E_m} \alpha_i^2 \, \varrho_m(\mathrm{d}\alpha) = 1/m$ we obtain

$$e^{\text{avg}}(n, \varrho_m)^2 = \frac{1}{m} \sum_{i=n+1}^m \sigma_i^2. \tag{4.44}$$

If we put $m = 2n$ then we obtain

$$e^{\mathrm{avg}}(n, \varrho_{2n}) \geq \tfrac{1}{2}\sqrt{2}\sigma_{2n}.$$

Together with (4.42), we obtain

$$e^{\mathrm{ran\text{-}meas}}(n) \geq \tfrac{1}{2}\sqrt{2}e^{\mathrm{avg}}(2n, \varrho_{4n}) \geq \tfrac{1}{2}\sigma_{4n} = \tfrac{1}{2}e^{\mathrm{wor}}(4n - 1). \qquad \square$$

Remark 4.43. Here we discuss the approximation of Sobolev embeddings in the randomized setting, see Section 4.2.4 for the worst case setting. So far, mainly the case

$$I : W_{p_1}^k([0, 1]^d) \to L_{p_2}([0, 1]^d) \tag{4.45}$$

has been studied. More general Lipschitz domains are not a problem for the class Λ^{all}. For Λ^{std} it is not clear whether the order of convergence depends on Ω or not, see the remarks in Section 4.2.4.

We start with the class Λ^{std}. If $W_{p_1}^k([0, 1]^d)$ is embedded into $C([0, 1]^d)$, then the optimal rate is the same as for deterministic algorithms, namely

$$n^{-k/d + (1/p_1 - 1/p_2)_+}. \tag{4.46}$$

This holds for linear and nonlinear algorithms. This result is due to Mathé [141], partial results were proved earlier, see [160], [242].

If $W_{p_1}^k([0, 1]^d)$ is *not* embedded into $C([0, 1]^d)$, but still into $L_{p_2}([0, 1]^d)$, then deterministic algorithms with Λ^{std} cannot converge. Randomized algorithms still converge and optimal error bounds were recently proved by Heinrich [86].

Now we turn to the case Λ^{all}. The main results of Mathé [141] and Heinrich [79] can be summarized as follows. Let $1 \leq p_1, p_2 \leq \infty$ and $k, d \in \mathbb{N}$ with $k > d$. If $\max(p_1, p_2) < \infty$ or if $p_1 = p_2 = \infty$ then the optimal rate is $n^{-k/d}$. If $1 \leq p_1 < \infty$ and $p_2 = \infty$ then the rate is $n^{-k/d}(\ln n)^{1/2}$. Finally, if $p_1 = \infty$ and $1 \leq p_2 < \infty$ then the rate is between $n^{-k/d}(\ln n)^{-1/2}$ and $n^{-k/d}$. Observe that the optimal rate for $k \leq d$ is unknown. $\qquad \square$

We end this section with some open problems. These problems show that certain fundamental questions, all settled for a long time for deterministic algorithms, are still open for randomized algorithms.

Open Problem 20.

- Due to Theorem 4.4, we know that adaption can only help by a factor of at most 2 in the worst case setting for linear problems. Prove or disprove a similar result for randomized algorithms.

    It would be also very interesting to study the power of adaption for linear operators on convex sets, see (4.30) and Open Problem 17.

Open Problem 21.

- In Theorem 4.8 we state that linear algorithms are optimal for linear problems in the worst case setting. It is tempting to suggest that under the same conditions general randomized algorithms are only slightly better than optimal linear randomized algorithms of the form

$$A_n^\omega(f) = \sum_{i=1}^n L_i^\omega(f) g_i^\omega,$$

where the random variables L_i and g_i do not depend on f. Give a precise form of this conjecture and prove or disprove it.

Open Problem 22.

- Is there a form of Theorem 4.9 for randomized algorithms?

Open Problem 23.

- We know that if the assumptions of Theorem 4.42 hold, then randomized algorithms are only slightly better than deterministic algorithms. Nevertheless it would be interesting to characterize optimal randomized algorithms and to compute the values of $e^{\mathrm{ran}}(n)$ exactly. If we restrict ourselves to linear algorithms then the nth minimal error is given by

$$\max_{m>n} \sqrt{\frac{m-n}{\sum_{i=1}^m \sigma_i^{-2}}}, \tag{4.47}$$

where $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots$ are the singular values of $S\colon \widetilde{F} \to G$, see [161]. It is worthwhile to mention that the numbers in (4.47) appear in different extremal problems for different computational settings, see the work of Fang and Qian [58].

Open Problem 24.

- Assume that $S\colon \widetilde{F} \to G$ is a compact linear operator between Hilbert spaces $\widetilde{F}$ and G with $\Lambda = \Lambda^{\mathrm{all}}$. We believe that

$$e^{\mathrm{ran}}(n) \geq \tfrac{1}{2} e^{\mathrm{wor}}(4n-1), \tag{4.48}$$

but could only prove such an inequality for measurable randomized algorithms, see Theorem 4.42. Prove (4.48) in full generality.

More generally, find conditions under which $e^{\mathrm{ran}}(n) = e^{\mathrm{ran\text{-}meas}}(n)$.

Open Problem 25.[10]

- Study Sobolev embeddings

$$I\colon H_{p_1}^{s_1}(\Omega) \to H_{p_2}^{s_2}(\Omega)$$

---

[10]Partly solved recently by Heinrich [87], [88].

in the randomized setting. As Remark 4.43 indicates, the known results mainly concern the case $I : W_p^k([0, 1]^d) \to L_q([0, 1]^d)$; other domains and other smoothness classes were not much studied.

For the study of elliptic boundary value problems, the case $s_2 < 0$ and $p_2 = 2$ is especially important, see Remark 4.22 for the worst case setting.

4.4 Multivariate Problems and Tractability

In this section we assume that a whole sequence of problems

$$S_d : F_d \to G_d, \quad d \in \mathbb{N},$$

is given. The sets F_d are subsets of normed spaces $\widetilde{F}_d$, such as the unit ball, and usually the problem elements of $\widetilde{F}_d$ are functions defined on a set $D_d \subseteq \mathbb{R}^d$. The set G_d is a normed space. Similarly, $\Lambda_d \subseteq \widetilde{F}^*$ is a set of admissible information functionals.

As in Section 4.1.4 we define certain optimal error bounds and information complexities and, of course, these numbers now also depend on d. We start with the absolute error in the worst case setting. We want to find the smallest n for which the error is at most ε and define

$$n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{wor}}(A_n) \leq \varepsilon\}. \tag{4.49}$$

Here A_n is any algorithm that approximates S_d and is of the form $A_n = \varphi \circ N$, where N uses at most n functionals from Λ_d. For the normalized (by the initial error) error we obtain instead the definition

$$n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{wor}}(A_n) \leq \varepsilon \cdot e_0^{\mathrm{wor}}\}. \tag{4.50}$$

It should be always clear from the context whether we mean the absolute error, the normalized error, or another type of error. The numbers $n = n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda_d)$ describe the *information complexity* of the problem and in many cases also the complete complexity of the problem. Sometimes it is more convenient to consider the numbers

$$e^{\mathrm{wor}}(n, S_d) = e^{\mathrm{wor}}(n, S_d, F_d) = \inf_{A_n} e^{\mathrm{wor}}(A_n) \tag{4.51}$$

or their normalized counterparts $e^{\mathrm{wor}}(n, S_d)/e^{\mathrm{wor}}(0, S_d)$. Again, if we want to stress the role of Λ_d we write $e^{\mathrm{wor}}(n, S_d, \Lambda_d)$. It is obvious that $n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda_d)$ and $e^{\mathrm{wor}}(n, S_d, \Lambda_d)$ are inversely related.

In other settings, we proceed analogously. For example, in the average case setting, we have for the absolute error,

$$n^{\mathrm{avg}}(\varepsilon, S_d, \Lambda_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{avg}}(A_n) \leq \varepsilon\}, \tag{4.52}$$

and for the normalized error,

$$n^{\mathrm{avg}}(\varepsilon, S_d, \Lambda_d) = \min\{n \mid \text{there exists } A_n \text{ with } e^{\mathrm{avg}}(A_n) \leq \varepsilon \cdot e_0^{\mathrm{avg}}\}. \tag{4.53}$$

In the probabilistic and randomized settings we have similar definitions

A traditional way to measure the complexity of multivariate problems is to study the optimal order of convergence and one might believe that a problem is "well behaved" if the order of convergence does not depend on d. We have seen in Chapter 3 that this intuition is quite often misleading and therefore we need to study more carefully which problems are tractable and which are not.

4.4.1 Polynomial Tractability

Most papers on tractability of continuous multivariate problems study what we call in this book polynomial tractability. We already introduced this notion in Chapter 3 for a number of examples. We now give the formal definition of polynomial tractability. Let $n(\varepsilon, S_d, \Lambda_d)$ be the information complexity in the worst case, average case or randomized setting for the absolute or normalized error for approximating S_d.[11]

Definition 4.44. A problem $S = \{S_d\}$ is *polynomially tractable* (for the class Λ_d), if there exist non-negative numbers C, q, and p such that

$$n(\varepsilon, S_d, \Lambda_d) \leq C d^q \varepsilon^{-p} \quad \text{for all } d \in \mathbb{N}, \varepsilon \in (0, 1). \tag{4.54}$$

The problem is *strongly polynomially tractable* if (4.54) holds with $q = 0$. In this case, the exponent of strong polynomial tractability is defined as the infimum of all p satisfying (4.54) with $q = 0$.

Finally, if (4.54) does *not* hold then S is said to be *polynomially intractable*.

Let $n(\varepsilon, d) = n(\varepsilon, S_d, \Lambda_d)$. Assume, for example, that

$$n(\varepsilon, d) \approx 2^d \varepsilon^{-p}.$$

Then the order of the error of optimal algorithms A_n is $n^{-1/p}$ independently of the dimension d but the problem is *not* polynomially tractable since it is impossible to have (4.54) with C independent of d.

Of course, intuitively we may think that polynomially tractable problems are "easy", and that polynomially intractable problems are "difficult". This is not always correct. Assume, for example, that for a specific problem we have

$$n(\varepsilon, d) \approx 10 d^{10} \varepsilon^{-10}.$$

Then the problem is polynomially tractable although $n(\varepsilon, d)$ might be too large for d and ε arising in computational practice. For example, take $\varepsilon = 10^{-1}$ and $d = 10$. Then $n(\varepsilon, d) \approx 10^{21}$ is huge. On the other hand, take another problem for which

$$n(\varepsilon, d) \approx 1.01^d \varepsilon^{-1}.$$

---

[11] For simplicity, we omit here the probabilistic setting which requires one more parameter δ. In Volume II we provide a proper generalization of polynomial tractability in the probabilistic setting.

Then the problem is polynomially intractable but for the same $\varepsilon = 10^{-1}$ and $d = 100$, we obtain $n(\varepsilon, d) \approx 10e$, which is not large.

Hence (4.54) is a theoretical condition that is often (but not always) useful for applications. Obviously, the smaller C, q and p in (4.54) the more useful is the notion of polynomial tractability. This is very similar to the study of *polynomial complexity* for discrete problems, see, e.g., the discussion in Arora and Barak [3].

Assume that two sequences

$$S_{d,i} \colon F_d \to G_d, \quad d \in \mathbb{N},$$

of multivariate continuous problems are given, $i \in \{1, 2\}$. If both problems $\{S_{d,1}\}$ and $\{S_{d,2}\}$ are polynomially tractable then intuitively the sum $\{S_{d,1} + S_{d,2}\}$ should also be polynomially tractable. This is true, and easy to prove, for the absolute error. It is *not* true, however, for the normalized error which is shown in the following example from [287].

Example 4.45. Assume that $F_d = G_d$ are Hilbert spaces and all the operators $S_{d,i}$ are compact and self-adjoint. We assume that $S_{d,1}$ and $S_{d,2}$ have the same eigenvectors with corresponding eigenvalues are $\lambda_{j,d} = 1/j$ for all $j \in \mathbb{N}$ for $S_{d,1}$ and $\beta_{j,d} = -1/j$ for $j \leq 2^d - 1$ and $\beta_{j,d} = -1/(2j)$ for $j \geq 2^d$ for $S_{d,2}$.

Then clearly both problems are properly normalized, $\|S_{d,i}\| = 1$. They are strongly polynomially tractable with exponent one. However, their sum $S_{d,1} + S_{d,2}$ has non-zero eigenvalues $\beta_j = 1/(2j)$ for $j \geq 2^d$ and norm 2^{-d-1}. The normalized operator

$$\frac{S_{d,1} + S_{d,2}}{\|S_{d,1} + S_{d,2}\|}$$

has non-zero eigenvalues $2^d/j$ for $j \geq 2^d$. Then for the normalized error we have

$$n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda^{\mathrm{all}}) = \left\lceil 2^d (\varepsilon^{-1} - 1) \right\rceil.$$

Clearly, this problem is intractable. The normalization introduced an exponential factor in d which caused intractability.

We add in passing that we obtain a somewhat similar result if we compare the L_2-discrepancy with the L_∞-discrepancy, see Section 3.1.5 of Chapter 3. The numbers $e^{\mathrm{wor}}(n, d)$ are smaller for the L_2-discrepancy, so that one might think that this is the "easier" problem. However, this problem is intractable with respect to the normalized error, while the "more difficult" problem of L_∞-discrepancy is tractable. Again this is a consequence of normalization.

4.4.2 Weak Tractability

The essence of tractability is that the minimal number of information operations $n(\varepsilon, S_d, \Lambda_d)$ needed to solve the problem to within ε is *not* exponential in ε^{-1} and d for the absolute or normalized error criterion.

Hence, we want to guarantee that $n(\varepsilon, d)$ is asymptotically much smaller than $a^{\varepsilon^{-1}+d}$ for any $a > 1$. This means that a *necessary* condition on tractability is

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0. \tag{4.55}$$

If condition (4.55) holds then we say that the problem $S = \{S_d\}$ is *weakly tractable*, whereas if (4.55) is *not* satisfied then we say that the problem $S = \{S_d\}$ is *intractable*.

4.4.3 Generalized Tractability

Generalized tractability may differ from polynomial tractability in two ways. The first is the domain of (ε, d). For polynomial tractability, ε and d are independent, and $(\varepsilon^{-1}, d) \in [1, \infty) \times \mathbb{N}$. For some applications, as in mathematical finance, d is huge but we are only interested in a rough approximation, so that ε is not too small. There may be also problems for which d is relatively small and we are interested in a very accurate approximation which corresponds to a very small ε. For generalized tractability, we assume that $(\varepsilon^{-1}, d) \in \Omega$, where

$$[1, \infty) \times \{1, 2, \ldots, d^*\} \cup [1, \varepsilon_0^{-1}) \times \mathbb{N} \subseteq \Omega \subseteq [1, \infty) \times \mathbb{N} \tag{4.56}$$

for some non-negative integer d^* and some $\varepsilon_0 \in [0, 1]$ such that

$$d^* + (1 - \varepsilon_0) > 0.$$

The essence of (4.56) is that for all such Ω, we know that at least one of the parameters (ε^{-1}, d) may go to infinity but not necessarily both of them. Indeed, if $d^* > 0$ then ε may go to 0, so that we are considering multivariate problems with $d \le d^*$, whereas if $\varepsilon_0 < 1$ then d may go to infinity, so that we are considering multivariate problems with $\varepsilon \ge \varepsilon_0$.

Hence, for generalized tractability we assume that $(\varepsilon^{-1}, d) \in \Omega$, and we may choose Ω satisfying (4.56) for some d^* and ε_0. Obviously, the choice of Ω should reflect what is needed in computational practice.

The second way in which generalized tractability may differ from polynomial tractability is how we measure the lack of exponential dependence. We define a *tractability* function

$$T : [1, \infty) \times [1, \infty) \to [1, \infty),$$

which is non-decreasing in both variables and which grows to infinity slower than an exponential function a^x when x tends to infinity for any $a > 1$. More precisely, for a given Ω satisfying (4.56), we assume that for any $a > 1$, $T(x, y)/a^{x+y}$ tends to 0 for $(x, y) \in \Omega$ as $x + y$ approaches infinity. This is equivalent to assuming that

$$\lim_{(x,y) \in \Omega, x+y \to \infty} \frac{\ln T(x, y)}{x + y} = 0. \tag{4.57}$$

With Ω satisfying (4.56) and T satisfying (4.57), we study *generalized tractability*. We say that $S = \{S_d\}$ is (T, Ω)-tractable if there are non-negative numbers C and t such that

$$n(\varepsilon, S_d, \Lambda_d) \leq C\,T(\varepsilon^{-1}, d)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega.$$

We also have generalized strong tractability if we replace $T(\varepsilon^{-1}, d)$ by $T(\varepsilon, 1)$. More precisely, we say that $S = \{S_d\}$ is strongly (T, Ω)-tractable if there is a non-negative number t such that

$$n(\varepsilon, S_d, \Lambda_d) \leq C\,T(\varepsilon^{-1}, 1)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega.$$

In both cases, we are interested in the smallest exponents t; these are called the exponents of (generalized) tractability and strong tractability. Note that generalized tractability coincides with polynomial tractability if we take $\Omega = [1, \infty) \times \mathbb{N}$ and $T(x, y) = xy$. Chapter 8 of this book is devoted to the study of generalized tractability and contains more motivations and examples of interesting T and Ω.

4.5 Notes and Remarks

NR 4:1. In this chapter we survey results from information-based complexity, as they are needed for tractability studies. Many of the results can be found in much greater detail and with proofs in the IBC book [242]. We also survey, however, some more recent results.

NR 4.1:1. The study of the worst case setting for problems with partial information has a relatively long history. The first paper in this area we could trace down[12] is the paper of von Mises [149] from 1933 who proposed to study optimal linear algorithms for univariate integration and the class of functions whose mth derivatives exist and are uniformly bounded. Around 1950, Bückner [22], Nikolskij [159] and Sard [207] studied optimal linear algorithms for integration and approximation. Kiefer [109] studied the nonlinear problem of searching for the maximum of a unimodal function and proved that Fibonacci search is optimal. His results are based on his MIT Master's degree thesis in 1948 and were published in 1953. It seems to us that Kiefer was the first one to find the complexity of a problem with no restriction on the class of algorithms and sample points of function values.

Today there are literally thousands of papers analyzing various problems in the worst case setting.

NR 4.1.3:1. This section is based on [290].

NR 4.2.1:1. This section is based on Creutzig, Wojtaszczyk [34] and on [165], [242].

---

[12]We are grateful to Knut Petras who informed us about the paper of von Mises.

NR 4.2.2:1. Our survey is again based on Creutzig, Wojtaszczyk [34] and on [242].

NR 4.2.3:1. Theorem 4.11 is often stated and proved in the literature for the case when $S \colon H \to G$ is compact. The more general result can be found in Pinkus [193].

NR 4.2.4:1. There are hundreds of papers on n-widths, s-numbers, and sampling numbers for Sobolev embeddings, starting with the paper of Kolmogorov [112] from 1936. These papers are surveyed and some of the results are extended in the recent papers of Vybíral [253], [254]. We give more references in the text.

NR 4.3.1:1. Much more details on the average case setting and its history can be found in [242]. In particular, we want to mention that the first papers dealing with the average case setting for problems with partial information are due to Suldin [229], [230] from 1959 and 1960, and Larkin [130] from 1972. The results reported in this section are based on the results originally obtained in Lee and Wasilkowski [131], Papageorgiou and Wasilkowski [188], Wasilkowski [261], and [267] as well as in the works already mentioned in this section.

NR 4.3.2:1. Much more details on the probabilistic setting can be found in [242]. The results reported in this section are based on the results originally obtained in Lee and Wasilkowski [131], Wasilkowski [260], [261], and [289].

NR 4.3.3:1. The complexity of randomized algorithms were studied and surveyed in Heinrich [80], [81], Wasilkowski [262], and in [160], [165], [242]. The reader can find many references there. Here we only mention a few important works.

Nemirovsky, Yudin [155] studied the problem of global optimization. Integral equations were studied by Emelyanov and Ilin [56], Heinrich [82], [84], Heinrich and Mathé [89], Pfeiffer [190], and in [168]. The randomized complexity of elliptic PDEs was studied by Heinrich [85].

The computation of high dimensional integrals is often performed with randomized algorithms, see Heinrich [80] as well as [160], [165], [242], [262] for "classical" results. Tractability of high dimensional integration in the randomized setting was studied in [219] and by Wasilkowski [264]. Numerical integration with respect to an unknown density was studied in [142] and in Rudolf [203].

NR 4.3.3:2. Concerning Lemma 4.37: Bakhvalov [7] considered the case $n(f, \omega) = n \in \mathbb{N}$ of fixed cardinality. Randomized algorithms with varying cardinality were studied by Wasilkowski [262], in [160], [242], and in many more recent papers. A less explicit form of the lemma is contained in [160, p. 64]. Similar results are contained in Heinrich [81], [82]. With almost the same proof one obtains the inequality

$$e^{\mathrm{ran}}(n) \geq \sqrt{\frac{\alpha - 1}{\alpha}} e^{\mathrm{avg}}(\alpha n, \varrho), \tag{4.58}$$

which holds for $\alpha > 1$, see Pfeiffer [190].

NR 4.3.3:3. Concerning Theorem 4.42: For simplicity we have only used (4.40) and (4.44) for $m = 2n$. Of course one may use (4.58) and (4.44) for a different m to obtain a more general result.

NR 4.4:1. The concept of polynomial tractability was introduced in [287], the concept of generalized tractability in [69].

Chapter 5
Worst Case Setting

This is the first chapter in which we study polynomial and weak tractability for general linear multivariate problems. In this chapter we define multivariate problems over Hilbert spaces, and consider the absolute and normalized error criterion in the worst case setting for the class Λ^{all} of all linear continuous functionals. The class Λ^{std} of function values will be studied in Volumes II and III.

In Section 5.1, we study a sequence of linear multivariate problems $\{S_d\}$ for which S_{d+1} is not necessarily related to S_d. Polynomial and weak tractability conditions are expressed in terms of their singular values. The essence of these conditions is that the singular values must decay sufficiently quickly and have polynomial or non-exponential bounds in terms of d. It is also shown that polynomial and weak tractability for the absolute and normalized error criteria are not related. That is, it may happen that polynomial or weak tractability holds for one of the error criteria and does not for the other.

In Section 5.2, we assume that a sequence of linear multivariate problems is defined by a univariate linear problem and that d-variate problems are obtained by tensor products of the univariate problem. Hence, we now study problems for which all variables and groups of variables play the same role. To omit a trivial case we assume that the univariate problem has at least two non-zero singular values. Polynomial and weak tractability conditions are now expressed only in terms of the singular values of the univariate problem. In this case, we have a number of negative results establishing the curse of dimensionality or polynomial intractability. For example, any linear tensor product problem suffers from the curse of dimensionality for the absolute error criterion if the largest singular value is larger than one, and in the normalized error criterion if the two largest singular values are the same. Any such problem is polynomially intractable for the absolute error criterion if the largest singular value is at least 1 and for the normalized error criteria independently on the largest singular value. There are also a few positive results. In particular, polynomial tractability holds for the absolute error criterion if the largest singular value is less than 1 and if the singular values enjoy a polynomial decay. In fact, polynomial tractability is equivalent to strong polynomial tractability in this case. For the normalized error we can have only weak tractability, which holds if the two largest singular values are different and if the singular values enjoy a logarithmic decay.

In Section 5.3, we study linear weighted tensor product problems. We define such problems and introduce the formal definitions of product, order-dependent, finite-order and finite-diameter weights. The essence of all these definitions is to model multivariate problems for which successive variables or groups of variables play different roles. In particular, we may have d-variate functions that are the sum of functions of a few, say ω^*, variables (finite-order weights) or d-variate functions with decaying depen-

dence on successive variables (product weights). To keep the number of pages in this book relatively small, we do not study tractability of weighted problems for the absolute error criterion. We leave it as an open problem and we believe that the existing proof technique will allow our readers to solve this problem in the near future.

We restrict ourselves to the normalized error criterion and present polynomial and weak tractability conditions in terms of the weights and the singular values of the univariate problem. We now have many positive tractability results, as long as the weights decay sufficiently fast. For example, we have polynomial tractability for product weights iff the sum of some powers of the weights in the d-dimensional case grows at most proportionally to $\ln d$, and strong polynomial tractability iff the latter sum is uniformly bounded in d. For finite-order weights, we always have polynomial tractability, and the degree of d in the estimate of the information complexity is at most ω^*. For finite-diameter weights, where a d-variate function is a sum of functions depending on at most q^* successive variables, this degree is just one.

We also study weak tractability. As we already mentioned, if the two largest singular values are different and the singular values decay logarithmically then even the unweighted problem is weakly tractable. So, the natural question is to verify how much weights can help if the largest singular value is of multiplicity $p \geq 2$. Then indeed we have weak tractability under a suitable assumption on the weights. In particular, for $p = 2$, this assumption says that the number of weights larger than ε^2 cannot be exponential in $\varepsilon^{-1} + d$.

The results of this chapter are illustrated by three linear multivariate problems. We consider the linear Schrödinger equation, and multivariate approximation for weighted Korobov and Sobolev spaces.

The example of multivariate approximation for the weighted Sobolev space of Section 5.4.1 is based on [280], and shows that there are various ways of obtaining weighted problems leading to different tractability results. This example also illustrates an intriguing relationship between smoothness and tractability that we have already discussed in the initial chapters. It turns out that smoothness *hurts* tractability for this example, and tractability is only possible for the smallest value of the smoothness parameter. Even when the problem is tractable, then the conditions on the weights are a little different than before. For example, we have polynomial tractability for finite-order weights. However, we must assume that they are bounded, as opposed to the previous case.

5.1 Linear Problems Defined over Hilbert Spaces

In this section we consider linear multivariate problems defined over Hilbert spaces. We study their polynomial and weak tractability for the absolute and normalized error criteria in the worst case setting and for the class Λ^{all}.

As in Section 4.4 of Chapter 2, we consider the problem $S = \{S_d\}$, where $S_d \colon H_d \to G_d$ is a compact linear operator and H_d and G_d are Hilbert spaces.

We know from Section 4.2 of Chapter 2 that the nth minimal error of the compact operator S_d in the worst case setting for the class Λ^{all} is the $(n+1)$st singular value of S_d which, in turn, is the square root of the $(n+1)$st largest eigenvalue $\lambda_{d,n+1}$ of the compact self-adjoint and positive semi-definite operator $W_d = S_d^* S_d \colon H_d \to H_d$.

We first consider tractability of S for the absolute error criterion. Then the information complexity $n(\varepsilon, d) := n^{\text{wor}}(\varepsilon, S_d, \Lambda^{\text{all}})$ of the multivariate problem S_d for the class Λ^{all} is

$$n(\varepsilon, d) = \min\{n \mid \lambda_{d,n+1} \leq \varepsilon^2\}.$$

The eigenvalues $\{\lambda_{d,j}\}_{j=1}^{\infty}$ can be any non-increasing sequence converging to 0 and hence any function $n \colon \mathbb{R} \times \mathbb{N} \to \mathbb{N}_0$ that is non-increasing in the first variable is possible as $n(\varepsilon, d)$.

It is clear that tractability of S depends on the behavior of the sequence of eigenvalues $\{\lambda_{d,j}\}_{j,d=1}^{\infty}$. To omit the trivial case, we assume that S is not 0, i.e., no S_d is the zero operator, which implies that $\lambda_{d,1} > 0$ for all $d \in \mathbb{N}$. We first study polynomial tractability.

Theorem 5.1. *Consider the non-zero problem $S = \{S_d\}$ for compact linear S_d defined over Hilbert spaces. We study the problem S for the absolute error criterion in the worst case setting and for the class Λ^{all}.*

- *S is polynomially tractable iff there exist $C_1 > 0$, $q_1 \geq 0$, $q_2 \geq 0$ and $\tau > 0$ such that*

$$C_2 := \sup_d \left(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} d^{-q_2} < \infty. \tag{5.1}$$

- *If (5.1) holds then*

$$n(\varepsilon, d) \leq \left(C_1 + C_2^{\tau} \right) d^{\max(q_1, q_2\tau)} \varepsilon^{-2\tau} \quad \text{for all } \varepsilon \in (0, 1] \text{ and } d = 1, 2, \ldots.$$

- *If S is polynomially tractable, so that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some positive C and p with $q \geq 0$, then (5.1) holds with*

$$C_1 = C + 2, \quad q_1 = q, \quad q_2 = 2qp^{-1},$$

and for any τ such that $\tau > p/2$. Then

$$C_2 \leq C^{2/p} \zeta \left(2\tau/p \right)^{1/\tau},$$

where ζ is the Riemann zeta function.

- *S is strongly polynomially tractable iff (5.1) holds with $q_1 = q_2 = 0$. The exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = \inf\{2\tau \mid \tau \text{ satisfies (5.1) with } q_1 = q_2 = 0\}.$$

Proof. Assume first that we have polynomial tractability with $n(\varepsilon, d) \leq Cd^q \varepsilon^{-p}$ for some positive C and p, and $q \geq 0$. This means that $\lambda_{d,n(\varepsilon,d)+1} \leq \varepsilon^2$. Since the eigenvalues $\lambda_{d,j}$ are non-increasing, we have

$$\lambda_{d,\lfloor Cd^q \varepsilon^{-p} \rfloor +1} \leq \varepsilon^2.$$

Suppose that $j = \lfloor Cd^q \varepsilon^{-p} \rfloor + 1$. If we vary $\varepsilon \in (0,1]$ then j takes the values $j = \lfloor Cd^q \rfloor + 1, \lfloor Cd^q \rfloor + 2, \ldots$. We also have $j \leq Cd^q \varepsilon^{-p} + 1$ which is equivalent to $\varepsilon^2 \leq (Cd^q/(j-1))^{2/p}$. Hence

$$\lambda_{d,j} \leq \left(\frac{Cd^q}{j-1} \right)^{2/p} \qquad \text{for all } j \geq \lfloor Cd^q \rfloor + 1.$$

We will be using this inequality for

$$j = \lceil (C+2)d^q \rceil, \lceil (C+2)d^q \rceil + 1, \ldots.$$

This is valid since $\lceil (C+2)d^q \rceil \geq (C+2)d^q \geq \lfloor Cd^q \rfloor + 2 \geq \lfloor Cd^q \rfloor + 1$. We also have $j \geq 2$.

For $\tau > p/2$, we get

$$
\begin{aligned}
\left(\sum_{j=\lceil (C+2)d^q \rceil}^{\infty} \lambda_{d,j}^\tau \right)^{1/\tau} &\leq (Cd^q)^{2/p} \left(\sum_{j=\lceil (C+2)d^q \rceil}^{\infty} (j-1)^{-2\tau/p} \right)^{1/\tau} \\
&\leq (Cd^q)^{2/p} \left(\sum_{j=1}^{\infty} j^{-2\tau/p} \right)^{1/\tau} \\
&= (Cd^q)^{2/p} \zeta(2\tau/p)^{1/\tau}.
\end{aligned}
$$

Thus

$$C_2 := \sup_d \left(\sum_{j=\lceil (C+2)d^q \rceil}^{\infty} \lambda_{d,j}^\tau \right)^{1/\tau} d^{-2q/p} \leq C^{2/p} \zeta(2\tau/p)^{1/\tau} < \infty.$$

Hence, (5.1) holds with $C_1 = C+2, q_1 = q, q_2 = 2q/p$, and any τ such that $\tau > p/2$ and $C_2 \leq C^{2/p} \zeta(2\tau/p)^{1/\tau}$. This also proves the third point of the theorem.

Assume now that (5.1) holds. Since $\lambda_{d,j}$ are ordered, $\lambda_{d,j+1} \leq \lambda_{d,j}$, we have

$$\left(n - \lceil C_1 d^{q_1} \rceil + 1 \right)^{1/\tau} \lambda_{d,n} \leq \left(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^\tau \right)^{1/\tau} \leq C_2 d^{q_2} \qquad (5.2)$$

for $n = \lceil C_1 d^{q_1} \rceil, \lceil C_1 d^{q_1} \rceil + 1, \ldots$.

Choose the smallest n such that

$$\left(n - \lceil C_1 d^{q_1} \rceil + 2 \right)^{-1/\tau} C_2 d^{q_2} \leq \varepsilon^2.$$

Then $\lambda_{d,n+1} \leq \varepsilon^2$ and

$$n = \left\lceil \left(\frac{C_2 d^{q_2}}{\varepsilon^2}\right)^\tau \right\rceil + \lceil C_1 d^{q_1} \rceil - 2 \leq \frac{C_2^\tau d^{q_2 \tau}}{\varepsilon^{2\tau}} + C_1 d^{q_1}$$

$$\leq \frac{C_1 \varepsilon^{2\tau} + C_2^\tau}{\varepsilon^{2\tau}} d^{\max(q_1, q_2 \tau)} \leq \left(C_1 + C_2^\tau\right) d^{\max(q_1, q_2 \tau)} \varepsilon^{-2\tau}.$$

Hence, S is polynomially tractable and $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ with $C = C_1 + C_2^\tau$, $q = \max(q_1, q_2\tau)$ and $p = 2\tau$. This also proves the second point of the theorem.

Strong polynomial tractability of S is proven similarly by taking $q = 0$ in the first part of the proof, and $q_1 = q_2 = 0$ in the second part. The formula for the exponent of strong polynomial tractability follows from the second and third points of the theorem. This completes the proof.   □

The essence of Theorem 5.1 is that polynomial tractability of S fully depends on the behavior of the eigenvalues $\lambda_{d,j}$ modulo the first polynomially-many largest eigenvalues. That is, as long as $j < \lceil C_1 d^{q_1} \rceil$ for some C_1 and q_1, then there is no condition on the eigenvalues $\lambda_{d,j}$. This is quite natural since with polynomially-many information operations we can eliminate the effect of the first polynomially-many largest eigenvalues. However, for $j \geq \lceil C_1 d^{q_1} \rceil$ the behavior of the remaining eigenvalues is essential. A necessary and sufficient condition for polynomial tractability of S is that the sum of their powers must be at most polynomial in d. To obtain strong polynomial tractability of S, we may ignore only finitely-many largest eigenvalues and the sum of the rest of eigenvalues raised to some power must be uniformly bounded in d.

It is easy to check that (5.1) holds iff there exist positive C_1, C_2 and non-negative q_1, q_2 and positive r such that

$$\lambda_{d,n} \leq C_2 d^{q_2} \left(n - \lceil C_1 d^{q_1} \rceil + 1\right)^{-r} \quad \text{for all } n \geq \lceil C_1 d^{q_1} \rceil. \tag{5.3}$$

Indeed, if (5.1) holds then (5.2) implies (5.3) with the same C_1, C_2, q_1, q_2 and $r = 1/\tau$. On the other hand, if (5.3) holds then for any $\tau > 1/r$ we have

$$\left(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^\tau\right)^{1/\tau} \leq C_2 d^{q_2} \zeta(\tau r)^{1/\tau},$$

and (5.1) holds.

The essence of (5.3) is that polynomial tractability of S holds iff, modulo polynomially-many largest eigenvalues, the remaining eigenvalues $\lambda_{d,n}$ are bounded polynomially in d and $(n - \lceil C_1 d^{q_1} \rceil + 1)^{-1}$.

As an example observe that for $\lambda_{d,j} = e^{\alpha \sqrt{d}} j^{-\beta}$ we do *not* have polynomial tractability for $\alpha > 0$, whereas strong polynomial tractability holds for $\alpha \leq 0$ and $\beta > 0$. In the latter case, the exponent of strong polynomial tractability is $2/\beta$.

We now turn to the normalized error criterion for the same class Λ^{all} and prove an analogous theorem characterizing polynomial tractability in terms of the behavior of the eigenvalues $\{\lambda_{d,j}\}$. We now have

$$n(\varepsilon, d) = \min\{n \mid \lambda_{d,n+1} \leq \varepsilon^2 \lambda_{d,1}\}.$$

Theorem 5.2. *Consider the non-zero problem $S = \{S_d\}$ for compact linear S_d defined over Hilbert spaces. We study the problem S for the normalized error criterion in the worst case setting and for the class Λ^{all}.*

- *S is polynomially tractable iff there exist $q_2 \geq 0$ and $\tau > 0$ such that*

$$C_2 := \sup_d \left(\sum_{j=1}^{\infty} \left(\frac{\lambda_{d,j}}{\lambda_{d,1}} \right)^{\tau} \right)^{1/\tau} d^{-q_2} < \infty. \tag{5.4}$$

- *If (5.4) holds then*

$$n(\varepsilon, d) \leq C_2^{\tau} d^{q_2 \tau} \varepsilon^{-2\tau} \quad \text{for all } \varepsilon \in (0, 1] \text{ and } d = 1, 2, \ldots.$$

- *If S is polynomially tractable, so that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some positive C and p with $q \geq 0$, then (5.4) holds with $q_2 = 2q/p$ and any τ such that $\tau > p/2$. Then*

$$C_2 \leq 2^{1/\tau} (C + 2)^{2/p} \zeta(2\tau/p)^{1/\tau}.$$

- *S is strongly polynomially tractable iff (5.4) holds with $q_2 = 0$. The exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = \inf\{2\tau \mid \tau \text{ satisfies (5.4) with } q_2 = 0\}.$$

Proof. Since the proof is similar to the previous one, we only sketch the differences between them. Assuming that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ we now know that

$$\lambda_{d,n(\varepsilon,d)+1} \leq \varepsilon^2 \lambda_{d,1}.$$

This yields as before that

$$\frac{\lambda_{d,j}}{\lambda_{d,1}} \leq \left(\frac{C d^q}{j-1} \right)^{2/p} \quad \text{for } j = \lfloor C d^q \rfloor + 1, \lfloor C d^q \rfloor + 1, \ldots.$$

Obviously, $\lambda_{d,j}/\lambda_{d,1} \leq 1$ for all j.

For $\tau > p/2$, we now obtain

$$\left(\sum_{j=1}^{\infty} \left(\frac{\lambda_{d,j}}{\lambda_{d,1}} \right)^{\tau} \right)^{1/\tau} = \left(\sum_{j=1}^{\lceil (C+2)d^q \rceil - 1} \left(\frac{\lambda_{d,j}}{\lambda_{d,1}} \right)^{\tau} + \sum_{j=\lceil (C+2)d^q \rceil}^{\infty} \left(\frac{\lambda_{d,j}}{\lambda_{d,1}} \right)^{\tau} \right)^{1/\tau}$$

$$\leq \left[(C + 2)d^q + (C d^q)^{2\tau/p} \zeta(2\tau/p) \right]^{1/\tau}.$$

Since $2\tau/p > 1$ and $\zeta(2\tau/p) > 1$, the second term with C replaced by $C + 2$ in the last inequality is larger than the first one. Therefore

$$\left(\sum_{j=1}^{\infty} \left(\frac{\lambda_{d,j}}{\lambda_{d,1}} \right)^{\tau} \right)^{1/\tau} \leq 2^{1/\tau}(C+2)^{2/p}\zeta(2\tau/p)^{1/\tau}d^{2q/p}.$$

This proves (5.4) as well as the third point of the theorem.

Assuming (5.4) we conclude that

$$n^{1/\tau}\lambda_{d,n} \leq C_2 d^{q_2}\lambda_{d,1} \quad \text{for all } n = 1, 2, \ldots.$$

Then $\lambda_{d,n+1} \leq \varepsilon^2 \lambda_{d,1}$ holds for

$$n = \left\lceil \left(\frac{C_2 d^{q_2}}{\varepsilon^2} \right)^{\tau} \right\rceil - 1 \leq C_2^{\tau} d^{q_2 \tau}\varepsilon^{-2\tau}.$$

This proves polynomial tractability of S and the second point of the theorem. Strong polynomial tractability follows as before. □

The main difference between Theorems 5.1 and 5.2 is that for the absolute error criterion the polynomially many largest eigenvalues of $\{\lambda_{d,j}\}$ do not count, whereas for the normalized error criterion the whole sequence of normalized eigenvalues $\{\lambda_{d,j}/\lambda_{d,1}\}$ counts. The reason is that although polynomially-many initial eigenvalues can be arbitrarily large, for the normalized error criterion we consider the ratios $\lambda_{d,j}/\lambda_{d,1}$ which are always at most one. Hence, for the normalized error criterion there is no need to drop the initial polynomial part of the sequence, which was necessary for the absolute error criterion.

It is natural to ask whether polynomial tractabilities for the absolute and normalized error criteria are related. It is easy to see that they are *not*. That is, it may happen that we have, say, polynomial tractability for the absolute error criterion but not for the normalized error criterion or vice versa. Indeed, assume that we have the eigenvalues $\{\lambda_{d,j}\}$ such that

$$\{\lambda_{d,j}\} = \{(j_1 + \alpha)^{-\beta}(j_2 + \alpha)^{-\beta} \cdots (j_d + \alpha)^{-\beta}\}_{j_1, j_2, \ldots, j_d=1}^{\infty}$$

for some $\alpha \geq 0$ and $\beta > 0$. As we shall see in the next section, such eigenvalues may occur for linear tensor product problems.

Observe that $\lambda_{d,j} \leq 1$ and

$$\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} = \left(\sum_{j=1}^{\infty} (j + \alpha)^{-\beta\tau} \right)^d$$

which is finite iff $\beta\tau > 1$. Assuming that $\tau > 1/\beta$ we have

$$
\Big(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \Big)^{1/\tau} = \Big(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} - \sum_{j=1}^{\lceil C_1 d^{q_1} \rceil - 1} \lambda_{d,j}^{\tau} \Big)^{1/\tau}
$$

$$
\geq \Big(\Big(\sum_{j=1}^{\infty} (j+\alpha)^{-\beta\tau} \Big)^{d} - C_1 d^{q_1} \Big)_{+}^{1/\tau}.
$$

We now show that we have polynomial tractability for the absolute error criterion iff

$$
A := \sum_{j=1}^{\infty} (j+\alpha)^{-\beta\tau} \leq 1.
$$

Indeed, if $A > 1$ then

$$
\Big(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \Big)^{1/\tau} \geq \big(A^d - C_1 d^{q_1} \big)_{+}^{\tau}
$$

goes exponentially fast to infinity with d for any C_1, q_1 and τ. Then (5.1) is *not* satisfied for any q_2, and Theorem 5.1 implies polynomial intractability of S. On the other hand if $A \leq 1$ then

$$
\Big(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \Big)^{1/\tau} \leq A^{d/\tau} \leq 1.
$$

By Theorem 5.1, we see that S is strongly polynomially tractable, since (5.1) holds with $q_1 = q_2 = 0$.

Observe that $A \leq 1$ iff $\alpha > 0$ and τ is sufficiently large. Indeed, if $\alpha = 0$ then the first term in A is 1 and the rest of terms are positive. Therefore $A > 1$ for any τ. For $\alpha > 0$ and $\tau > 1/\beta$ we have

$$
A = (1+\alpha)^{-\beta\tau} + (2+\alpha)^{-\beta\tau} + \sum_{j=3}^{\infty} (j+\alpha)^{-\beta\tau}
$$

$$
\leq (1+\alpha)^{-\beta\tau} + (2+\alpha)^{-\beta\tau} + \int_{2}^{\infty} (x+\alpha)^{-\beta\tau} \, dx
$$

$$
= (1+\alpha)^{-\beta\tau} + (2+\alpha)^{-\beta\tau} + \frac{1}{\beta\tau - 1} \frac{1}{(2+\alpha)^{\beta\tau - 1}}.
$$

If τ goes to infinity then the upper bound of A goes to 0. Hence, for sufficiently large τ we have $A \leq 1$, so that S is strongly polynomially tractable. From the last point of Theorem 5.1 we conclude that the exponent of strong polynomial tractability is $p^{\text{str-wor}} = 2\tau^*/\beta$, where τ^* is defined by the condition

$$
\sum_{j=1}^{\infty} (j+\alpha)^{-\tau^*} = 1.
$$

We now turn to the normalized error criterion for the same sequence of eigenvalues. We have

$$\left(\sum_{j=1}^{\infty}\left(\frac{\lambda_{d,j}}{\lambda_{d,1}}\right)^{\tau}\right)^{1/\tau} = \left(\sum_{j=1}^{\infty}\left(\frac{1+\alpha}{j+\alpha}\right)^{\beta\tau}\right)^{d/\tau}.$$

Since the first term of the last sum is 1 independently of α and τ and the remaining terms are positive, the left hand side goes exponentially fast to infinity with d no matter whether $\alpha = 0$ or $\alpha > 0$ and what the value of τ. Due to Theorem 5.1, this implies that the problem S is *polynomially intractable* for arbitrary $\alpha \geq 0$ and $\beta > 0$.

In summary, for $\alpha > 0$ we have polynomial tractability for the absolute error criterion and polynomial intractability for the normalized error criterion.

The opposite case, polynomial intractability for the absolute error criterion and polynomial tractability for the normalized error criterion, can be obtained for the sequence of eigenvalues that we considered before, namely for $\lambda_{d,j} = e^{\alpha\sqrt{d}} j^{-\beta}$ for positive α and β. The lack of polynomial tractability for the absolute error criterion was already discussed, whereas for the normalized error criterion $\lambda_{d,j}/\lambda_{d,1} = j^{-\beta}$, and we even obtain strong polynomial tractability with the exponent $2/\beta$.

We now analyze weak tractability of S in terms of the behavior of the eigenvalues $\{\lambda_{d,j}\}$. The conditions on weak tractability can be presented simultaneously for both the absolute and normalized error criteria by defining

$$\text{CRI}_d = 1 \qquad \text{for the absolute error criterion,}$$
$$\text{CRI}_d = \lambda_{d,1} \quad \text{for the normalized error criterion.}$$

We are ready to prove the following theorem.

Theorem 5.3. *Consider the non-zero problem $S = \{S_d\}$ for compact linear S_d defined over Hilbert spaces. We study the problem S for the absolute or normalized error criterion in the worst case setting and for the class Λ^{all}.*

S is weakly tractable iff

- *we have*

$$\lim_{j\to\infty} \frac{\lambda_{d,j}}{\text{CRI}_d} \ln^2 j = 0 \quad \text{for all } d, \text{ and}$$

- *there exists a function $f : (0, \frac{1}{2}] \to \mathbb{N}$ such that*

$$M := \sup_{\beta\in(0,\frac{1}{2}]} \frac{1}{\beta^2} \sup_{d\geq f(\beta)} \sup_{j\geq\lceil\exp(d\sqrt{\beta})\rceil+1} \frac{\lambda_{d,j}}{\text{CRI}_d} \ln^2 j < \infty.$$

Proof. Assume first that S is weakly tractable. Then for any $\beta \in (0, \frac{1}{2}]$ there exists a positive integer M_β such that for all pairs (ε^{-1}, d) with $(\varepsilon^{-1} + d) \geq M_\beta$, we have $n(\varepsilon, d) \leq \exp(\beta(\varepsilon^{-1} + d))$. Hence, $\lambda_{d,n(\varepsilon,d)+1} \leq \varepsilon^2 \text{CRI}_d$ which implies that

$$\lambda_{d,j} \leq \varepsilon^2 \text{CRI}_d \quad \text{for } j = \lfloor \exp\left(\beta(\varepsilon^{-1} + d)\right)\rfloor + 1.$$

For this index j we have $\ln(j-1) \le \beta(\varepsilon^{-1}+d)$ and $\varepsilon^2 \le \beta^2/(\ln(j-1)-\beta d)_+^2$. Hence, $\lambda_{d,j}/\text{CRI}_d \le \beta^2/(\ln(j-1)-\beta d)_+^2$ for $j = \lfloor \exp\big(\beta(\varepsilon^{-1}+d)\big) \rfloor + 1$ and for all pairs (ε^{-1},d) for which $(\varepsilon^{-1}+d) \ge M_\beta$. By varying ε in the interval $(0,1]$ we obtain

$$\frac{\lambda_{d,j}}{\text{CRI}_d} \le \frac{\beta^2}{(\ln(j-1)-\beta d)_+^2} \quad \text{for all } j = \lfloor e^{\beta M_\beta} \rfloor + 1, \lfloor e^{\beta M_\beta} \rfloor + 2, \ldots.$$

For fixed d and for sufficiently large j, we obtain that $\lambda_{d,j}/\text{CRI}_d$ is of order $\beta^2/\ln^2 j$. Since β can be arbitrarily small, this proves that

$$\lim_{j\to\infty} \frac{\lambda_{d,j}}{\text{CRI}_d} \ln^2 j = 0 \quad \text{for all } d$$

which is the first condition presented in the theorem.

To prove the second condition, we define $f(\beta) = M_\beta$. Then for $d \ge f(\beta)$ and $j \ge \lceil \exp(d\sqrt{\beta}) \rceil + 1$ we have $j \ge \lfloor \exp(\beta M_\beta) \rfloor + 1$, and $\ln(j-1) \ge d\sqrt{\beta}$, which yields $\ln(j-1) - \beta d \ge \ln(j-1)(1-\sqrt{\beta}) > 0$. Since $j \ge 3$, we finally conclude that

$$\frac{1}{\beta^2} \frac{\lambda_{d,j}}{\text{CRI}_d} \ln^2 j \le \frac{\ln^2 j}{\ln^2(j-1)} \frac{1}{1-\sqrt{\beta}} \le \left(\frac{\ln 3}{\ln 2}\right)^2 (2+\sqrt{2}),$$

which proves that $M < \infty$, and completes the first part of the proof.

Assume now that the two conditions presented in the theorem hold. Take an arbitrary $\beta \in (0, \frac{1}{2}]$. Then the first condition implies that there exists an integer $C_\beta > 2$ such that for all $j \ge C_\beta$ and all $d = 1, 2, \ldots, f(\beta)-1$ we have $\lambda_{d,j}/\text{CRI}_d \ln^2 j \le \beta$. Hence,

$$\frac{\lambda_{d,j}}{\text{CRI}_d} \le \frac{\beta}{\ln^2 j} \le \varepsilon^2 \quad \text{for } j = \left\lceil e^{\sqrt{\beta}/\varepsilon} \right\rceil \le 2e^{\sqrt{\beta}/\varepsilon},$$

since $\lceil x \rceil \le 2x$ for $x \ge 1$.

This means that for all $d = 1, 2, \ldots, f(\beta)-1$, and all $\varepsilon \in (0,1]$ we have

$$\ln n(\varepsilon, d) \le \max\left(\ln C_\beta, \frac{\sqrt{\beta}}{\varepsilon} + \ln 2 \right). \tag{5.5}$$

Consider now the case $d \ge f(\beta)$. Then for $j \ge \lceil \exp(d\sqrt{(\beta)}) \rceil + 1$, the second condition yields

$$\frac{\lambda_{d,j}}{\text{CRI}_d} \le \frac{M\beta^2}{\ln^2 j} \le \varepsilon^2 \quad \text{for } j = \max\left(\left\lceil e^{\sqrt{M}\beta/\varepsilon} \right\rceil, \left\lceil e^{(d\sqrt{\beta})} \right\rceil + 1 \right).$$

Since $\ln(x+1) \le \ln x + \ln 2$ for $x \ge 1$, we now obtain

$$\ln n(\varepsilon, d) \le \max\left(\frac{\sqrt{M}\beta}{\varepsilon} + \ln 2, d\sqrt{\beta} + 2\ln 2 \right). \tag{5.6}$$

From (5.5) and (5.6) we conclude that for all $\varepsilon \in (0, 1]$ and all $d \in \mathbb{N}_+$ we have

$$\frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} \leq \max\left(\frac{\ln C_\beta}{\varepsilon^{-1} + d}, \sqrt{\beta} + \frac{\ln 2}{\varepsilon^{-1} + d}, \sqrt{M}\beta + \frac{\ln 2}{\varepsilon^{-1} + d}, \sqrt{\beta} + \frac{2\ln 2}{\varepsilon^{-1} + d}\right).$$

For sufficiently large $\varepsilon^{-1} + d$, the right hand side is less than $2\max(\sqrt{\beta}, \sqrt{M}\beta)$. Since β can be arbitrarily small, this proves that

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0,$$

which means that S is weakly tractable. This completes the proof.   □

Theorem 5.3 states that weak tractability of S is equivalent to the two conditions on the eigenvalues $\lambda_{d,j}$. The first condition simply states that $\lambda_{d,j}$ must go to 0 faster than $\ln^{-2} j$ as j goes to infinity, and this must hold for all d. This condition is quite natural since otherwise $n(\varepsilon, d)$ would be exponential in ε^{-1}. Clearly, the first condition is not enough for weak tractability, since it does not address the dependence on d. The second condition addresses this dependence and is quite complicated. It says that for all $\beta \in (0, \frac{1}{2}]$, large $d \geq f(\beta)$, and large $j \geq \lceil\exp(d\sqrt{\beta})\rceil + 1$, we must have $\lambda_{d,j} \ln^2 j/(\beta^2 \mathrm{CRI}_d)$ uniformly bounded in β, d and j.

We now check these conditions for $\lambda_{d,j} = \exp(\alpha_1 d^{\alpha_2})j^{-\alpha_3}$, for positive α_1, α_2, and α_3. Obviously for such a simple sequence of eigenvalues, we may directly check that for the absolute error criterion we have

$$n(\varepsilon, d) = \left\lceil \frac{\exp\left(\alpha_1 d^{\alpha_2}/\alpha_3\right)}{\varepsilon^{2/\alpha_3}} \right\rceil,$$

so that weak tractability holds iff $\alpha_2 < 1$, whereas for the normalized error criterion we have

$$n(\varepsilon, d) = \left\lceil \frac{1}{\varepsilon^{2/\alpha_3}} \right\rceil,$$

and we even have strong polynomial tractability independently of α_1 and α_2, with the exponent $2/\alpha_3$.

We now check the two conditions presented in Theorem 5.3 just for illustration. We begin with the absolute error criterion, $\mathrm{CRI}_d = 1$. The first condition is obvious since $j^{-\alpha_3} \ln^2 j$ indeed goes to 0. To check the second condition we note that

$$\lambda_{d,j} \ln^2 j = e^{\alpha_1 d^{\alpha_2}} j^{-\alpha_3} \ln^2 j.$$

For $j \geq \lceil\exp(d\sqrt{\beta})\rceil + 1$ we have $j = x\exp(d\sqrt{\beta})$ for some $x \geq 1$, and then

$$\sup_{j\geq\lceil\exp(d\sqrt{\beta})\rceil+1} j^{-\alpha_3} \ln^2 j \leq e^{-d\alpha_3\sqrt{\beta}} \sup_{x\geq 1} x^{-\alpha_3}\left(d\sqrt{\beta} + \ln x\right)^2$$

$$\leq e^{-d\alpha_3\sqrt{\beta}}(2d^2\beta + C)$$

for some absolute constant C, where the last estimate is sharp with respect to d. Hence,

$$\sup_{j \geq \lceil \exp(d\sqrt{\beta})\rceil + 1} \lambda_{d,j} \ln^2 j \leq e^{\alpha_1 d^{\alpha_2} - d\alpha_3 \sqrt{\beta}} (2d^2 \beta + C),$$

and again this bound is sharp in d. Clearly, if $\alpha_2 \geq 1$, then for small β we obtain an exponential function in d; by taking the supremum over large d we obtain infinity independently of the definition of $f(\beta)$. Hence, $\alpha_2 \geq 1$ implies the lack of weak tractability, or equivalently, intractability. Assume then that $\alpha_2 < 1$. Choose $\alpha_4 \in (\alpha_2, 1)$. Then for any positive β we can define a positive integer $f(\beta)$ such that for all $d \geq f(\beta)$ we have

$$\alpha_1 d^{\alpha_2} + \frac{1}{\beta} d^{\alpha_4} \leq d\alpha_3 \sqrt{\beta},$$

$$2\ln \frac{1}{\beta} + \ln(2d^2 \beta + C) \leq \frac{1}{\beta} d^{\alpha_4}.$$

Then

$$\frac{1}{\beta^2} \sup_{d \geq f(\beta)} \sup_{j \geq \lceil \exp(d\sqrt{\beta})\rceil + 1} \lambda_{d,j} \ln^2 j \leq \frac{1}{\beta^2} e^{-d^{\alpha_4}/\beta} (2d^2 \beta + C)$$

$$= \exp\left(2\ln\frac{1}{\beta} + \ln(2d^2\beta + C) - \frac{1}{\beta} d^{\alpha_4}\right)$$

$$\leq 1.$$

This implies that $M < \infty$ and the second condition holds.

For the normalized error criterion, $\mathrm{CRI}_d = \exp(\alpha_1 d^{\alpha_2})$, and

$$\frac{\lambda_{d,j}}{\mathrm{CRI}_d} \ln^2 j = j^{-\alpha_3} \ln^2 j$$

is independent of d. The first condition holds as before, whereas the second condition holds if we define a positive C_β such that

$$j^{-\alpha_3} \ln^2 j \leq \beta^2 \quad \text{for all } j \geq C_\beta,$$

and take

$$f(\beta) = \left\lceil \frac{\ln C_\beta}{\sqrt{\beta}} \right\rceil.$$

We now show that weak tractabilities for the absolute and normalized error criteria are not related. For the previous example,

$$\lambda_{d,j} = \exp(\alpha_1 d^{\alpha_2}) j^{-\alpha_3}$$

with positive α_i, and $\alpha_2 \geq 1$, we do not have weak tractability for the absolute error criterion and we do have weak tractability for the normalized error criterion. Now consider

$$\lambda_{d,j} = \frac{1}{e^d \ln^{2+1/d}(j+1)} \quad \text{for all } d, j \in \mathbb{N}.$$

Then for the absolute error we have

$$\ln n(\varepsilon, d) \leq e^{-d/(2+1/d)} \varepsilon^{-2/(2+1/d)}.$$

Let $\varepsilon^{-1} + d$ go to infinity. Then ε^{-1} or d go to infinity. Assume that only ε^{-1} goes to infinity. Then $(\ln n(\varepsilon, d))/(\varepsilon^{-1} + d)$ goes to 0 since the exponent $2/(2 + 1/d) < 1$. Assume then that d goes to infinity. Now $(\ln n(\varepsilon, d))/(\varepsilon^{-1} + d) \leq e^{-d/2}$ clearly goes to 0. Hence, weak tractability holds for the absolute error criterion.

For the normalized error criterion we have

$$\ln (n(\varepsilon, d) + 1) \geq \varepsilon^{-2/(2+1/d)} \ln 2.$$

Taking $\varepsilon^{-1} = d$ we conclude that

$$\frac{\ln n(1/d, d)}{2d} \geq \frac{\ln 2}{2} d^{-1/(2d+1)} \to \frac{\ln 2}{2} > 0.$$

Hence weak tractability does not hold for the normalized error criterion.

For some eigenvalues it may be difficult to check the second condition of Theorem 5.3 to establish weak tractability. We now provide an easier condition which is, however, only sufficient for weak tractability.

Lemma 5.4. *Consider the non-zero problem $S = \{S_d\}$ for compact linear S_d defined over Hilbert spaces. We study the problem S for the absolute or normalized error criterion in the worst case setting and for the class Λ^{all}.*

If there exists a positive τ such that

$$\lim_{d \to \infty} \frac{\ln \left(\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} / \mathrm{CRI}_d \right)}{d} = 0 \tag{5.7}$$

then S is weakly tractable.

Proof. Since

$$n^{1/\tau} \lambda_{d,n} \leq \left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau}$$

then $\lambda_{d,n} \leq \varepsilon^2 \mathrm{CRI}_d$ holds for

$$n = \left\lceil \left(\frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau}}{\mathrm{CRI}_d} \right)^{\tau} \varepsilon^{-2\tau} \right\rceil.$$

Since $\lceil x \rceil \leq x + 1 \leq 2 \max(x, 1)$ and $\ln \max(x, 1) \leq \max(0, \ln x) = (\ln x)_+$ for all $x \in [0, \infty)$ we have

$$\frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} \leq \frac{\ln 2 + \left(2\tau \ln \varepsilon^{-1} + \tau \ln \left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} / \mathrm{CRI}_d \right)_+}{\varepsilon^{-1} + d}$$

which goes to 0 as $\varepsilon^{-1} + d$ tends to infinity. Hence S is weakly tractable. $\square$

Clearly, (5.7) is only a sufficient condition for weak tractability. Indeed, (5.7) implies that $\lambda_{d,j}$ as a function of j goes polynomially to 0, whereas weak tractability only requires that $\lambda_{d,j} = o(\ln^{-2} j)$. For example, take

$$\lambda_{d,j} = \ln^{-3}(j+1) \quad \text{for all } j, d \in \mathbb{N}.$$

Then $\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau} = \infty$ for any positive τ and Lemma 5.4 does not apply, however, the problem is weakly tractable.

Example: Trade-offs of the Exponents

Theorems 5.1 and 5.2 state necessary and sufficient conditions on strong polynomial and polynomial tractability. If strong tractability holds then we have the formula for its exponent. However, if only polynomial tractability holds then the exponents of d and ε^{-1} are *not* uniquely defined since τ satisfying (5.1) or (5.4) may vary.

We now show that, in general, the exponents of d and ε^{-1} may indeed vastly change. Furthermore, the minimization of one of them may result in the maximization of the other. This means that there is a trade-off between them.

We illustrate this point for the following sequence of the eigenvalues

$$\lambda_{d,j} = \prod_{k=1}^{\min(d, \lceil \ln(d+1) \rceil^s)} j_k^{-\alpha} \quad \text{for all } j \in \mathbb{N}^d$$

for some positive s and α.

Note that the largest eigenvalue is 1, and there is no difference between the absolute and normalized error criteria. For $\tau > 1/\alpha$ we have

$$\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau} = \left(\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^{\tau}\right)^{1/\tau} = (\zeta(\alpha\tau))^{\frac{\min(d, \lceil \ln(d+1) \rceil^s)}{\tau}}$$

$$= (d+1)^{\frac{\ln \zeta(\alpha\tau)}{\tau} \frac{\min(d, \lceil \ln(d+1) \rceil^s)}{\ln(d+1)}}.$$

Since $\zeta(\alpha\tau) > 1$, from Theorem 5.2 we see that strong tractability does *not* hold for any positive s and α.

We now turn to polynomial tractability. Clearly, for $s > 1$ the exponent of $d+1$ goes to infinity and therefore also polynomial tractability does *not* hold. For $s \leq 1$, the exponent of $d+1$ is uniformly bounded in d and we have polynomial tractability. Furthermore we can take q_2 in Theorem 5.2 such that

$$q_2 > \frac{\ln \zeta(\alpha\tau)}{\tau} \limsup_{d \to \infty} \frac{\lceil \ln(d+1) \rceil^s}{\ln(d+1)}.$$

For $s < 1$, the last limit superior is 0 and we can take an arbitrarily small q_2.

The most interesting case is when $s = 1$. Then we can take

$$q_2 = \frac{\ln \zeta(\alpha\tau)}{\tau}$$

and Theorem 5.2 states that

$$n(\varepsilon, d) = \mathcal{O}\big(d^{\ln \zeta(\alpha\tau)} \varepsilon^{-2\tau}\big), \tag{5.8}$$

where the factor in the big $\mathcal{O}$ notation is independent of d and ε^{-1}, and depends only on $\tau > 1/\alpha$.

We now show that the exponents in (5.8) are sharp in the following sense. First of all, note that the exponent 2τ of ε^{-1} must be larger than $2/\alpha$. Indeed for $d = 1$ we have $n(\varepsilon, 1) = \Theta(\varepsilon^{-2/\alpha})$, and for $d = 2$ we have $n(\varepsilon, 2) = \Theta(\varepsilon^{-2/\alpha} \ln \varepsilon^{-1})$. Hence, $2\tau > 2/\alpha$, or $\tau > 1/\alpha$. We turn to the exponent of d and show that for the fixed exponent 2τ of ε^{-1}, the minimal exponent of d is $\ln \zeta(\alpha\tau)$. That is, if

$$n(\varepsilon, d) = \mathcal{O}\big(d^q \varepsilon^{-2\tau}\big) \quad \text{then} \quad q \geq \ln \zeta(\alpha\tau).$$

Indeed, $n(\varepsilon, d) = \mathcal{O}(d^q \varepsilon^{-2\tau})$ implies that

$$\lambda_{d,j} = \mathcal{O}\big(d^{q/\tau} j^{-1/\tau}\big).$$

For $\eta > \tau$, we have

$$\Big(\sum_{j=1}^{\infty} \lambda_{d,j}^{\eta}\Big)^{1/\eta} = \mathcal{O}\big(d^{q/\tau} \zeta(\eta/\tau)^{1/\eta}\big).$$

We also know that the left hand side is equal to $(d + 1)^{\eta^{-1} \ln \zeta(\alpha\eta)}$. This implies that $q/\tau \geq (\ln(\zeta(\alpha\eta)))/\eta$ for all $\eta > \tau$. Letting η tend to τ, we conclude that $q \geq \ln \zeta(\alpha\tau)$, as claimed.

For τ tending to infinity in (5.8), the exponent of d goes to 0 since $\zeta(\alpha\tau)$ goes to 1, whereas the exponent of ε^{-1} goes to infinity. On the other hand, if τ goes to $1/\alpha$ then the exponent of d goes to infinity since $\zeta(\alpha\tau)$ approaches infinity, whereas the exponent of ε^{-1} takes its minimal value $2/\alpha$.

This means we have a trade-off between the exponents of d and ε^{-1}. So how should we choose τ? It depends on how d and ε^{-1} are related for a specific application. For instance, assume that $d = d_\varepsilon = \varepsilon^{-\beta}$ for some positive β. Then

$$n(\varepsilon, d_\varepsilon) = \mathcal{O}\big(\varepsilon^{-(2\tau + \beta \ln \zeta(\alpha\tau))}\big).$$

In this case it seems reasonable to minimize the exponent of ε^{-1}. This means that τ should be the solution of

$$2 + \alpha\beta \frac{\zeta'(\alpha\tau)}{\zeta(\alpha\tau)} = 0,$$

or equivalently

$$\frac{2}{\alpha\beta} \sum_{j=1}^{\infty} \frac{1}{j^{\alpha\tau}} = \sum_{j=1}^{\infty} \frac{\ln j}{j^{\alpha\tau}}.$$

For instance, if we have $\alpha = \beta = 1$ then $\tau = 1.39843\ldots$, and if we have $\alpha = 1$ and $\beta = 2$ then $\tau = 1.68042\ldots$.

We end this example by noting that weak tractability holds for any positive s and α. Indeed, in this case we may apply Lemma 5.4 since for $\tau > 1/\alpha$ we have

$$\frac{\ln\left(\sum_{j\in\mathbb{N}^d} \lambda_{d,j}^{\tau}\right)^{1/\tau}}{d} = \Theta\left(\frac{\lceil\ln(d+1)\rceil^s}{d}\right)$$

and it goes to 0 as d approaches infinity.

Much more can be said about trade-offs between the exponents of d and ε^{-1} for general multivariate problems along the lines indicated in this example. We leave, however, this subject to the reader.

Example: Schrödinger Equation

We illustrate the approach of this section for the Schrödinger equation[1]

$$i\frac{\partial u}{\partial t}(x,t) = -\Delta u(x,t) + q(x)u(x,t) \quad \text{for } x \in I^d := (0,1)^d, \text{ and } t > 0,$$

subject to boundary conditions

$$u(x,t) = 0 \quad \text{for all } x \in \partial I^d, \text{ and } t > 0,$$

and the initial condition

$$u(x,0) = f(x) \quad \text{for all } x \in I^d.$$

Here q is a non-negative continuous function. For this example we assume that q is fixed and consider real functions f from a class $H_d \subseteq L_2(I^d)$. The operator Δ is the Laplacian,

$$\Delta u = \sum_{j=1}^{d} \frac{\partial^2 u}{\partial x_j^2}.$$

It is known that the solution u can be represented as a linear combination of the eigenfunctions of the operator $-\Delta + q$. To be more precise, since the operator

---

[1]For simplicity, we assume that all masses and the normalized Planck constant are one, and boundary conditions are for particles in a box.

$(-\Delta + q)^{-1} \colon L_2(I^d) \to L_2(I^d)$ is self-adjoint and compact, its eigenfunctions $\eta_{d,j}$, which satisfy

$$(-\Delta + q)^{-1}\eta_{d,j}(x) = \beta_{d,j}\eta_{d,j}(x) \quad \text{for all } x \in I^d,$$
$$\eta_{d,j}(x) = 0 \qquad\qquad \text{for all } x \in \partial I^d,$$

form an orthonormal basis of $L_2([0,1]^d)$. Here, $\beta_{d,j}$ are ordered, $\beta_{d,1} \geq \beta_{d,2} \geq \cdots > 0$. Clearly, $\beta_{d,j}$ goes to 0 as j approaches infinity.

The function $f \in L_2(I^d)$ can be then represented as

$$f(x) = \sum_{j=1}^{\infty} f_j \eta_{d,j}(x) \quad \text{with } f_j = \langle f, \eta_{d,j}\rangle_{L_2}.$$

The solution u of the Schrödinger equation is given by

$$u(x,t) = \sum_{j=1}^{\infty} e^{-it\beta_{d,j}^{-1}} f_j \eta_{d,j}(x)$$

for $x \in I^d$ and $t > 0$, if we assume that

$$\sum_{j=1}^{\infty} \beta_{d,j}^{-2}|f_j|^2 < \infty.$$

Clearly, $\|u(\cdot,t)\|_{L_2} = \|f\|_{L_2}$ for all $t > 0$.

For a general function q, it is difficult to find the eigenpairs $(\beta_{d,j}, \eta_{d,j})$. However, if $q \equiv q_0$ is a constant function then it is easy to check that the eigenpairs are given for $j = [j_1, j_2, \ldots, j_d]$ with positive integers j_k by

$$\beta_{d,j} = \left(q_0 + \sum_{k=1}^{d} \pi^2 j_k^2\right)^{-1} \quad \text{and} \quad \eta_{d,j}(x) = 2^{d/2}\prod_{k=1}^{d} \sin(\pi j_k x_k). \qquad (5.9)$$

We now describe the Schrödinger equation as a linear multivariate problem defined over Hilbert spaces and apply the tractability results of the current section. We first define the Schrödinger problem for the largest possible Hilbert space,

$$H_d = \{f \in L_2(I^d) \mid \sum_{j\in\mathbb{N}^d} \beta_{d,j}^{-2}|f_j|^2 < \infty\}, \qquad (5.10)$$

with the inner product $\langle f, g\rangle_{H_d} = \sum_{j\in\mathbb{N}^d} \beta_{d,j}^{-2} f_j g_j$ for $f, g \in H_d$.

The operator $S_d \colon H_d \to L_2([0,1]^d)$ is defined by

$$S_d f = u(\cdot, T) = \sum_{j\in\mathbb{N}^d} e^{-iT\beta_{d,j}^{-1}} f_j \eta_{d,j}$$

i.e., as the solution of the Schrödinger equation for time T. Clearly, S_d is a linear isometry since

$$\|S_d f\|_{L_2([0,1]^d)} = \|f\|_{L_2([0,1]^d)}.$$

We call $S_{\text{linear}} = \{S_d\}$ *the linear Schrödinger problem for a fixed $q \equiv q_0 \geq 0$ and T*, or for brevity, *the linear Schrödinger problem*. We can consider this problem for the class Λ^{all} or Λ^{std} and for the absolute or normalized error criterion. We restrict ourselves in this chapter to the class Λ^{all}.

The operator S_d is compact since the $\beta_{d,j}$'s approach 0 as $|j|$ goes to infinity. More precisely, if we order the eigenvalues $\{\beta_{d,k}\}_{k \in \mathbb{N}} = \{\beta_{d,j}\}_{j \in \mathbb{N}^d}$ then it is possible to check, see (8.16) of Chapter 8, that $\beta_{d,k} = \Theta(k^{-2/d})$ for $k \in \mathbb{N}$ with the factor in the Theta notation dependent on d. It is easy to check that

$$S_d^* f = \sum_{j \in \mathbb{N}^d} e^{\mathrm{i}T\beta_{d,j}^{-1}} \beta_{d,j}^2 f_j \eta_{d,j},$$

$$W_d f = \sum_{j \in \mathbb{N}^d} \beta_{d,j}^2 f_j \eta_{d,j}.$$

Hence, W_d does not depend on T, and its eigenpairs are $\{\lambda_{d,j}, \eta_{d,j}\}_{j \in \mathbb{N}^d}$ with

$$\lambda_{d,j} = \beta_{d,j}^2 \quad \text{and} \quad \lambda_{d,j} = \left(q_0 + \sum_{k=1}^d \pi^2 j_k^2\right)^{-2}.$$

Since $\sqrt{\lambda_{d,j}} = \Theta(|j|^{-2/d})$ we conclude that we conclude that for both error criteria we have

$$n(\varepsilon, d) = \mathcal{O}(\varepsilon^{-d/2})$$

with the factor in the big $\mathcal{O}$ notation depending on d. This shows that the linear Schrödinger equation is *not* polynomially tractable but the last bound is too weak to verify whether the linear Schrödinger equation is weakly tractable.

We now check weak tractability. The largest eigenvalue is obtained for $j_k = 1$, which implies that the initial error is

$$\|S_d\| = \|W_d\|^{1/2} = \frac{1}{q_0 + \pi^2 d}.$$

Observe that if we take $j \in \{1, 2\}^d$ then we obtain 2^d eigenvalues $\lambda_{d,j}$ that are greater or equal to $(q_0 + 4\pi^2 d)^{-2}$. This implies *intractability* for both the absolute and normalized error criteria. Indeed, for the absolute error criteria we take $\varepsilon_d = \frac{1}{2}(q_0 + 4\pi^2 d)^{-1}$ and conclude that

$$n(\varepsilon_d, d) \geq 2^d.$$

This yields

$$\liminf_{\varepsilon_d^{-1} + d \to \infty} \frac{\ln n(\varepsilon_d, d)}{\varepsilon_d^{-1} + d} \geq \frac{\ln 2}{1 + 8\pi^2} > 0.$$

For the normalized error criterion, we take $\varepsilon_d = \frac{1}{2}(q_0 + \pi^2 d)/(q_0 + 4\pi^2 d)$, and then $n(\varepsilon, d)$ is again at least 2^d, which implies that

$$\liminf_{\varepsilon_d^{-1}+d\to\infty} \frac{\ln n(\varepsilon_d, d)}{\varepsilon_d^{-1} + d} \geq \ln 2 > 0.$$

So we have intractability for both error criteria, as claimed.

The intractability of the linear Schrödinger problem means that the space H_d is too large. This space guarantees that the solution of the Schrödinger equation exists, and the information complexity of the problem is $\mathcal{O}(\varepsilon^{-d/2})$, but it is not *not* enough for tractability. To obtain weak or polynomial tractability we must put more restrictions on functions f, and switch to some subspace $\tilde{H}_d$ of H_d.

In fact, it is easy to see that the linear Schrödinger problem is weakly or polynomially tractable iff the approximation problem $\mathrm{APP}_d : \tilde{H}_d \to L_2([0,1]^d)$, with $\mathrm{APP}_d f = f$, is weakly or polynomially tractable. Indeed, this follows from the fact that if we have an algorithm $A_{n,d} f$ for approximating f then we can take $S_d A_{n,d} f$ as an algorithm for approximating $S_d f$, so that isometry of S_d yields that

$$\|S_d f - S_d A_{n,d} f\|_{L_2([0,1]^d)} = \|f - A_{n,d} f\|_{L_2([0,1]^d)}.$$

On the other hand, if we have an algorithm $B_{n,d} f$ for approximating $S_d f$ then, without loss of generality we can assume that $B_{n,d}$ is interpolatory, $B_{n,d} f = S_d \tilde{f}$ with $\tilde{f}$ from the unit ball of $\tilde{H}_d$ sharing the same information as f, and treat $\tilde{f}$ as an approximation of f. Then

$$\begin{aligned}\|f - \tilde{f}\|_{L_2([0,1]^d)} &= \|S_d f - S_d \tilde{f}\|_{L_2([0,1]^d)} \\ &= \|S_d f - B_{n,d} f\|_{L_2([0,1]^d)}.\end{aligned}$$

As we shall see later in the course of the book, we identify many spaces $\tilde{H}_d$ for which approximation is weakly or polynomially tractable. If $\tilde{H}_d \subseteq H_d$ then we can also claim weak or polynomial tractability of the linear Schrödinger problem.

5.2 Linear Tensor Product Problems

In the previous section, we studied linear compact operators S_d without assuming any relations between them. In this section we assume that S_d is a d-fold tensor product of a given linear compact operator S_1. We study polynomial and weak tractability of such problems for the absolute and normalized error criteria in the worst case setting and for the class Λ^{all}.

We now explain the tensor product construction. For $d = 1$, let D_1 be a Borel measurable subset of $\mathbb{R}$, and let H_1 be a separable Hilbert space of real valued functions defined on D_1 with the inner product denoted by $\langle \cdot, \cdot \rangle_{H_1}$. Let G_1 be an arbitrary Hilbert

space, and let $S_1 \colon H_1 \to G_1$ be a compact linear operator[2]. Then the positive semi-definite self-adjoint operator

$$W_1 := S_1^* S_1 \colon H_1 \to H_1$$

is also compact. Let $\{\lambda_i\}$ denote the sequence of non-increasing eigenvalues of W_1, or equivalently, the sequence of squares of the singular values of S_1. If $k = \dim(H_1)$ is finite, then W_1 has finitely many eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then we formally put $\lambda_i = 0$ for $i > k$. In any case, the eigenvalues λ_i converge to 0. As in Section 4.2.3, let $\{(\lambda_i, e_i)\}$ be the eigenpairs of the operator W_1, so that

$$W_1 e_i = \lambda_i e_i \quad \text{with } \lambda_i \geq \lambda_{i+1} \text{ and } \langle e_i, e_j \rangle_{H_1} = \delta_{i,j}$$

for all $i, j = 1, 2, \ldots, \dim(H_1)$.

Then for $n \leq \dim(H_1)$, the algorithm

$$A_n(f) = \sum_{i=1}^{n} \langle f, e_i \rangle_{H_1} S_1 e_i$$

is the nth optimal worst case algorithm and

$$e^{\mathrm{wor}}(n) = e^{\mathrm{wor}}(A_n) = \sqrt{\lambda_{n+1}}.$$

Without loss of generality, we assume that S_1 is not the zero operator, i.e., $\|S_1\| = \sqrt{\lambda_1} > 0$. Recall that the initial error is also $\sqrt{\lambda_1}$.

For $d \geq 2$, let

$$H_d = H_1 \otimes \cdots \otimes H_1$$

be the d-fold tensor product Hilbert space of H_1. This is a space of real valued functions defined on $D_d = D_1 \times \cdots \times D_1 \subseteq \mathbb{R}^d$. Similarly, let $G_d = G_1 \otimes \cdots \otimes G_1$, d times.

We want to cover simultaneously the cases where H_1 has finite and infinite dimension. We write $j \in [1, \dim(H_1) + 1)$ which for the finite dimensional space H_1 means that $j \in [1, \dim(H_1)]$, whereas for the infinite dimensional space H_1, it means that $j \in [1, \infty)$. With this notation in mind, suppose that $\{\eta_j\}_{j \in [1, \dim(H_1)+1)}$ is an orthonormal basis of H_1. For $j = [j_1, j_2, \ldots, j_d]$ with $j_k \in [1, \dim(H_1) + 1)$ and $x = [x_1, x_2, \ldots, x_d]$ with $x_k \in D_1$, define

$$\eta_j(x) = \prod_{k=1}^{d} \eta_{j_k}(x_k).$$

For brevity, we sometimes write $\eta_j = \eta_{j_1} \eta_{j_2} \cdots \eta_{j_d}$.

---

[2]It would be possible to start with a compact operator $S_1 \colon H_1 \to G_1$ between arbitrary separable Hilbert spaces. For example, H_1 could be a space of m-variate functions, and the tensor product construction will give us spaces H_d of $d\,m$-variate functions. This approach has been taken in a number of tractability papers, see e.g., [93], [273]. We prefer to take $m = 1$ for simplicity and to always keep d as the number of variables. More on this point can be also found in NR 5.2:1.

For $i, j \in [1, \dim(H_1) + 1)^d$, we have

$$\langle \eta_i, \eta_j \rangle_{H_d} = \prod_{k=1}^{d} \langle \eta_{i_k}, \eta_{j_k} \rangle_{H_1} = \prod_{k=1}^{d} \delta_{i_k, j_k} = \delta_{i,j}.$$

Hence, $\{\eta_j\}_{j \in [1, \dim(H_1)+1)^d}$ is an orthonormal basis of H_d.

The linear operator S_d is defined as the tensor product operator

$$S_d = S_1 \otimes \cdots \otimes S_1 \colon H_d \to G_d.$$

We have

$$S_d \eta_j = S_1 \eta_{j_1} \otimes S_1 \eta_{j_2} \otimes \cdots \otimes S_1 \eta_{j_d} \in G_d,$$

and for $f = \sum_{j \in [1, \dim(H_1)+1)^d} \langle f, \eta_j \rangle_{H_d} \eta_j \in H_d$, we have

$$S_d f = \sum_{j \in [1, \dim(H_1)+1)^d} \langle f, \eta_j \rangle_{H_d} S_d \eta_j.$$

Note that S_d is compact and that $\|S_d\| = \|S_1\|^d = \lambda_1^{d/2}$. So the initial error is $\lambda_1^{d/2}$, which is exponentially small in d if $\lambda_1 < 1$, is 1 if $\lambda_1 = 1$, and is exponentially large in d if $\lambda_1 > 1$.

We call the multivariate problem $S = \{S_d\}$ a *linear tensor product problem in the worst case setting*. Let

$$W_d := S_d^* S_d : H_d \to H_d.$$

Then W_d is positive semi-definite, self-adjoint and compact. It is easy to check that its eigenpairs $\{(\lambda_{d,j}, e_{d,j})\}$ are given by products of the univariate eigenpairs, i.e.,

$$\{\lambda_{d,j}\}_{j \in \mathbb{N}^d} = \{\lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d}\}_{j_1, j_2, \ldots, j_d \in [1, \infty)},$$
$$\{e_{d,j}\}_{j \in [1, \dim(H_1)+1)^d} = \{e_{j_1} e_{j_2} \cdots e_{j_d}\}_{j_1, j_2, \ldots, j_d \in [1, \dim(H_1)+1)^d},$$

where we use the short notation $e_{d,j} = e_{j_1} e_{j_2} \cdots e_{j_d} := e_{j_1} \otimes e_{j_2} \otimes \cdots \otimes e_{j_d}$, and obviously $e_{d,j}(x) = \prod_{k=1}^{d} e_{j_k}(x_k)$.

We now consider the information complexity for the absolute error criterion in which $\mathrm{CRI}_d = 1$, and for the normalized error criterion in which $\mathrm{CRI}_d = \lambda_1^d$. We have

$$n(\varepsilon, d) := n^{\mathrm{wor}}(\varepsilon, S_d, \Lambda^{\mathrm{all}}) = \left| \left\{ [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d \mid \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2 \mathrm{CRI}_d \right\} \right|,$$

with the convention that the cardinality of the empty set is 0. Hence, $n(\varepsilon, d) = 0$ for $\varepsilon^2 \geq \lambda_1^d / \mathrm{CRI}_d$.

Let $n = n(\varepsilon, d)$. The nth optimal worst case error algorithm for the class Λ^{all} is given by

$$A_n(f) = \sum_{[j_1, j_2, \ldots, j_d]: \, \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2 \mathrm{CRI}_d} \langle f, e_{j_1} e_{j_2} \cdots e_{j_d} \rangle_{H_d} S_d \left(e_{j_1} e_{j_2} \cdots e_{j_d} \right).$$

Note that $n(\varepsilon, d)$ is finite for all $\varepsilon \in (0, 1]$ and all d since $\lim_{j \to \infty} \lambda_j = 0$. For $d \geq 1$, we have

$$
n(\varepsilon, d + 1) = \sum_{j=1}^{\infty} n\left(\frac{\varepsilon \sqrt{\mathrm{CRI}_{d+1}}}{\sqrt{\lambda_j} \sqrt{\mathrm{CRI}_d}}, d \right)
$$

$$
= \sum_{j=1}^{\max\{i \mid \lambda_i > \left[\varepsilon \sqrt{\mathrm{CRI}_{d+1}}/(\lambda_1^{d/2}) \right]^2 \}} n\left(\frac{\varepsilon \sqrt{\mathrm{CRI}_{d+1}}}{\sqrt{\lambda_j} \sqrt{\mathrm{CRI}_d}}, d \right)
$$

$$
= \sum_{j=1}^{n\left(\varepsilon \sqrt{\mathrm{CRI}_{d+1}/\mathrm{CRI}_1}/\lambda_1^{d/2}, 1 \right)} n\left(\frac{\varepsilon \sqrt{\mathrm{CRI}_{d+1}}}{\sqrt{\lambda_j} \sqrt{\mathrm{CRI}_d}}, d \right).
$$

Note that for $\lambda_2 = 0$ the operator S_d is equivalent to a continuous linear functional and can be solved exactly with one information operation. Therefore $n(\varepsilon, d) \leq 1$ for all ε, and the problem S is trivial. Hence we will always assume that $\lambda_2 > 0$, so that the operator W_d has at least 2^d positive eigenvalues. This indicates that the study of tractability is meaningful even if the rest of the eigenvalues $\lambda_j = 0$.

We are ready to study polynomial and weak tractability of linear tensor product problems. We begin with the absolute error criterion.

Theorem 5.5. *Consider the linear tensor product problem in the worst case setting* $S = \{S_d\}$ *with* $\lambda_2 > 0$. *We study this problem for the absolute error criterion and for the class* Λ^{all}.

- *Let* $\lambda_1 > 1$. *Then S is intractable.*

  More precisely, if $\lambda_2 \geq 1$ *then for all* $\varepsilon \in (0, 1)$ *we have*

  $$
  n(\varepsilon, d) \geq 2^d.
  $$

  If $\lambda_2 < 1$ *then define*

  $$
  \alpha = \frac{\ln \lambda_1}{\ln \lambda_1/\lambda_2} \in (0, 1).
  $$

  If $\alpha \geq \frac{1}{2}$, *or equivalently* $\lambda_1 \lambda_2 \geq 1$, *then for all* $\varepsilon \in (0, 1)$ *we have*

  $$
  n(\varepsilon, d) \geq 2^{d-1}.
  $$

  If $\alpha < \frac{1}{2}$, *or equivalently* $\lambda_1 \lambda_2 < 1$, *then for all* $\varepsilon \in (0, 1)$ *we have*

  $$
  n(\varepsilon, d) \geq \exp\left(\left[\alpha \ln \alpha^{-1} + (1 - \alpha) \ln(1 - \alpha)^{-1} \right] d - \tfrac{1}{2} \ln d + \Omega(1) \right)
  $$

  as $d \to \infty$.

- *Let* $\lambda_1 = \lambda_2 = 1$. *Then S is intractable and for all* $\varepsilon \in (0, 1)$ *we have*

  $$
  n(\varepsilon, d) \geq 2^d.
  $$

- *Let $\lambda_1 = 1$ and $\lambda_2 < 1$. Then S is polynomially intractable.*

- *Let $\lambda_1 = 1$.*

  If S is weakly tractable then

  $$\lambda_2 < 1 \quad and \quad \lambda_n = o\left((\ln n)^{-2}\right) \ as\ n \to \infty.$$

  If
  $$\lambda_2 < 1 \quad and \quad \lambda_n = o\left((\ln n)^{-2}(\ln \ln n)^{-2}\right) \ as\ n \to \infty$$
  then S is weakly tractable.

- *Let $\lambda_1 < 1$.*

  If S is weakly tractable then $\lambda_n = o\left((\ln n)^{-2}\right)$ as $n \to \infty$.

  If $\lambda_n = o\left((\ln n)^{-2}(\ln \ln n)^{-2}\right)$ as $n \to \infty$ then S is weakly tractable.

- *Let $\lambda_1 < 1$. Then S is polynomially tractable iff S is strongly polynomially tractable iff there exists a positive r such that*

  $$\lambda_n = \mathcal{O}\left(n^{-r}\right) \quad as\ n \to \infty.$$

  If this holds then the exponent of strong polynomial tractability is

  $$p^{\text{str-wor}} = \inf\left\{2\tau \mid \sum_{j=1}^{\infty} \lambda_j^\tau \le 1\right\}.$$

Proof. Assume first that $\lambda_1 > 1$. For the d-dimensional case, we want to estimate the number of eigenvalues $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}$ that are at least equal to one. For $k \le d$, we take k indices of j_i's equal to two, and the remaining $(d-k)$ indices equal to one. We have $\binom{d}{k}$ eigenvalues equal to $\lambda_1^{d-k}\lambda_2^k = \lambda_1^d(\lambda_2/\lambda_1)^k$, which is a non-increasing function of k. Let k_d be the largest integer k from $[0, d]$ for which $\lambda_1^{d-k}\lambda_2^k \ge 1$. Then the number of eigenvalues greater or equal to 1 is at least equal to the sum of binomial coefficients for $k = 0, 1, \ldots, k_d$. That is, for all $\varepsilon \in (0, 1)$ we have

$$n(\varepsilon, d) \ge \sum_{k=0}^{k_d} \binom{d}{k}.$$

If $\lambda_2 \ge 1$ then $k_d = d$ and the sum of the binomial coefficients is obviously 2^d. Therefore $n(\varepsilon, d) \ge 2^d$, as claimed in the first sub-point of the theorem.

For $\lambda_2 < 1$, we have
$$k_d = \lfloor \alpha d \rfloor.$$

If $\alpha \ge \frac{1}{2}$ then $k_d \ge \lfloor d/2 \rfloor$, and the sum of the binomial coefficients is at least 2^{d-1}. Therefore $n(\varepsilon, d) \ge 2^{d-1}$, as claimed in the second sub-point of the theorem.

If $\alpha < \frac{1}{2}$, we estimate the sum of the binomial coefficients simply by the last term and use the estimate

$$n(\varepsilon, d) \ge \binom{d}{\lfloor \alpha d \rfloor}.$$

Using Stirling's formula, $m! = m^{m+\frac{1}{2}}e^{-m}\sqrt{2\pi}(1 + o(1))$, for factorials and using the fact that $k_d = \lfloor \alpha d \rfloor = \alpha d - \alpha_d$ with $\alpha_d \in [0, 1)$, we conclude for large d that

$$
\begin{aligned}
\ln n(\varepsilon, d) \geq{}& \ln d! - \ln k_d! - \ln(d - k_d)! \\
={}& (d + \tfrac{1}{2}) \ln d - (k_d + \tfrac{1}{2}) \ln k_d - (d - k_d + \tfrac{1}{2}) \ln(d - k_d) \\
& - d + k_d + d - k_d - \ln \sqrt{2\pi} + o(1) \\
={}& (d + \tfrac{1}{2}) \ln d \\
& - (\alpha d + \tfrac{1}{2} - \alpha_d) \left[\ln d + \ln \alpha + \ln \left(1 - \frac{\alpha_d}{\alpha d} \right) \right] + \mathcal{O}(1) \\
& - ((1 - \alpha)d + \tfrac{1}{2} + \alpha_d) \left[\ln d + \ln(1 - \alpha) + \ln \left(1 + \frac{\alpha_d}{(1 - \alpha)d} \right) \right] \\
={}& \left[d + \tfrac{1}{2} - \alpha d - \tfrac{1}{2} + \alpha_d - (1 - \alpha)d - \tfrac{1}{2} - \alpha_d \right] \ln d + \mathcal{O}(1) \\
& - (\alpha d + \tfrac{1}{2} - \alpha_d) \left(\ln \alpha + \mathcal{O}(d^{-1}) \right) \\
& - ((1 - \alpha)d + \tfrac{1}{2} + \alpha_d) \left(\ln(1 - \alpha) + \mathcal{O}(d^{-1}) \right) \\
={}& -\tfrac{1}{2} \ln d + \left[\alpha \ln \alpha^{-1} + (1 - \alpha) \ln(1 - \alpha)^{-1} \right] d + \mathcal{O}(1),
\end{aligned}
$$

which proves the estimate of the last sub-point of the first part of the theorem. Note that all these estimates state that $n(\varepsilon, d)$ is exponential in d and therefore

$$
\limsup_{\varepsilon^{-1}+d\to\infty} \frac{n(\varepsilon, d)}{\varepsilon^{-1} + d} > 0.
$$

Therefore the problem S is intractable, as claimed.

Assume now that $\lambda_1 = \lambda_2 = 1$. By the previous argument we have 2^d eigenvalues equal to one, and $n(\varepsilon, d) \geq 2^d$ for all $\varepsilon \in (0, 1)$, which yields intractability.

Assume then that $\lambda_1 = 1$ and $\lambda_2 < 1$. For any integer k and any $d \geq k$, by the previous argument we have $\binom{d}{k}$ eigenvalues equal to λ_2^k. Choosing $\varepsilon_k = (\lambda_2^k/2)^{1/2}$ we conclude that

$$
n(\varepsilon_k, d) \geq \binom{d}{k} = \Theta(d^k) \quad \text{as } d \to \infty.
$$

Since k can be arbitrarily large, this contradicts polynomial tractability of S.

We now study weak tractability for $\lambda_1 = 1$. Assume that S is weakly tractable. For $d = 1$ we have

$$
\lim_{\varepsilon^{-1}+1\to\infty} \frac{n(\varepsilon, 1)}{\varepsilon^{-1} + 1} = 0 \quad \text{iff} \quad n(\varepsilon, 1) = \exp\left(o(\varepsilon^{-1}) \right) \quad \text{as } \varepsilon \to 0.
$$

Since $n(\varepsilon, 1) = \min\{n \mid \lambda_{n+1} \leq \varepsilon^2\}$ then

$$
n(\varepsilon, 1) = \exp\left(o(\varepsilon^{-1}) \right) \quad \text{iff} \quad \varepsilon^2 = o\left(\ln^{-2} n(\varepsilon, 1) \right) \quad \text{iff} \quad \lambda_n = o\left(\ln^{-2} n \right)
$$

as $n \to \infty$. Clearly, λ_2 must be smaller than 1 since $\lambda_2 = 1$ contradicts weak tractability of S, due to the previous point.

It remains to show that $\lambda_2 < 1$ and $\lambda_n = o\left((\ln n)^{-2}(\ln \ln n)^{-2}\right)$ imply weak tractability of S. It is easy to see that

$$\ln n(\varepsilon, 1) = o\left(\varepsilon^{-1}/\ln \varepsilon^{-1}\right).$$

Indeed,

$$n(\varepsilon, 1) = \min\{n \mid \lambda_{n+1} < \varepsilon^2\} \le \min\{n \mid (\ln n)(\ln \ln n) = o(\varepsilon^{-1})\}.$$

Let $x = \ln n(\varepsilon, 1)$. Then $x \ln x = o(\varepsilon^{-1})$ implies that $x = o(\varepsilon^{-1}/\ln \varepsilon^{-1})$, as claimed.

Note that for $\varepsilon^2 \ge \lambda_2$ we have $n(\varepsilon, d) = 1$. Therefore we may assume that $\varepsilon^2 < \lambda_2$.

We need an estimate from [288] which was already used in Section 3.1.4. For completeness we derive this estimate here. Let

$$\lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2.$$

Let $k = k(j) \le d$ be the number of indices j_i that are at least two. Then $d - k$ indices j_i are equal to one, and $\lambda_2^k > \varepsilon^2$. The last inequality implies that $k \le a_d(\varepsilon)$ with

$$a_d(\varepsilon) = \min\left(d, \left\lceil 2\frac{\ln \varepsilon^{-1}}{\ln \lambda_2^{-1}} \right\rceil - 1\right).$$

Since $\varepsilon^2 < \lambda_2$ we have $a_d(\varepsilon) \ge 1$. Observe also that $\lambda_1 = 1$ implies that

$$\lambda_{j_i} \ge \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2,$$

and therefore $j_i \le n(\varepsilon, 1)$. Hence,

$$n(\varepsilon, d) \le \binom{d}{a_d(\varepsilon)} n(\varepsilon, 1)^{a_d(\varepsilon)} \le \frac{d^{a_d(\varepsilon)}}{a_d(\varepsilon)!} n(\varepsilon, 1)^{a_d(\varepsilon)}. \tag{5.11}$$

By taking logarithms we obtain

$$\ln n(\varepsilon, d) \le a_d(\varepsilon) \ln(d) - \ln a_d(\varepsilon)! + a_d(\varepsilon) \ln n(\varepsilon, 1).$$

Since $a_d(\varepsilon) = \Theta(\min(d, \ln \varepsilon^{-1}))$ we have

$$\frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = \mathcal{O}\left(\frac{\min(d, \ln \varepsilon^{-1}) \ln d}{\varepsilon^{-1} + d} + \frac{\min(d, \ln \varepsilon^{-1}) o(\varepsilon^{-1}/\ln \varepsilon^{-1})}{\varepsilon^{-1} + d}\right). \tag{5.12}$$

Let $x = \max(d, \varepsilon^{-1})$. We claim that $\min(d, \ln \varepsilon^{-1}) \le \ln x$. Indeed, if

$$\min(d, \ln \varepsilon^{-1}) = d$$

then $x = \varepsilon^{-1}$ and $d \le \ln \varepsilon^{-1} \le \ln x$. On the other hand, if

$$\min(d, \ln \varepsilon^{-1}) = \ln \varepsilon^{-1}$$

then since $\varepsilon^{-1} \le x$ we have $\min(d, \ln \varepsilon^{-1}) \le \ln x$, as claimed.

From this and the fact that the function $\varepsilon^{-1}/\ln\varepsilon^{-1}$ is increasing for small ε, we conclude

$$\frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = \mathcal{O}\left(\frac{\ln^2 x}{x} + o(1)\right) = o(1) \tag{5.13}$$

which goes to 0 as x goes to infinity. This means that S is weakly tractable.

We now consider the case $\lambda_1 < 1$. Then clearly $\lambda_2 \le \lambda_1 < 1$. If S is weakly tractable, then $\lambda_n = o\left(\ln^{-2} n\right)$, as before. On the other hand, if

$$\lambda_n = o\left((\ln n)^{-2}(\ln\ln n)^{-2}\right),$$

and $\lambda_2 < 1$, we get weak tractability from the previous point, since the problem with $\lambda_1 = 1$ is harder than the problem with $\lambda_1 < 1$.

We now consider polynomial tractability for $\lambda_1 < 1$. Assume first that S is polynomially tractable. Due to Theorem 5.1, (5.3) holds, which for $d = 1$ says that

$$\lambda_n = \mathcal{O}\left(n^{-r}\right)$$

for some positive r.

We also need to show that polynomial tractability of S implies strong polynomial tractability of S. We will use Theorem 5.1, which says that polynomial tractability holds iff $\sum_{j=\lceil C_1 d^{q_1}\rceil}^{\infty} \lambda_{d,j}^{\tau}$ is bounded by a multiple of $d^{q_2\tau}$ for some positive C_1, τ and non-negative q_1, q_2. Observe that

$$\sum_{j=\lceil C_1 d^{q_1}\rceil}^{\infty} \lambda_{d,j}^{\tau} = \sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} - \sum_{j=1}^{\lceil C_1 d^{q_1}\rceil - 1} \lambda_{d,j}^{\tau}$$

$$= \left(\sum_{j=1}^{\infty} \lambda_j^{\tau}\right)^d - \sum_{j=1}^{\lceil C_1 d^{q_1}\rceil - 1} \lambda_{d,j}^{\tau}$$

$$\ge \left(\sum_{j=1}^{\infty} \lambda_j^{\tau}\right)^d - \lambda_1^d C_1 d^{q_1}.$$

Since $\sum_{j=\lceil C_1 d^{q_1}\rceil}^{\infty} \lambda_{d,j}^{\tau}$ can be bounded by a multiple of $d^{q_2\tau}$, then $\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1$. Note that C_1, q_1 and q_2 are irrelevant for this property. That is, if this bound holds for some appropriate C_1, q_1 and q_2 then it also holds for $C_1 = 1$ and $q_1 = q_2 = 0$. By Theorem 5.1, this implies that S is strongly polynomially tractable.

Finally assume that $\lambda_n = \mathcal{O}(n^{-r})$ for some positive r. We show that there exists a positive τ such that

$$f(\tau) := \sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1.$$

Indeed, there is a positive C such that $\lambda_n \le Cn^{-r}$ for all n. For any $p \ge 2$ and $\tau r > 1$,

we have

$$f(\tau) \leq \lambda_1^\tau + \lambda_2^\tau + \cdots + \lambda_p^\tau + C \sum_{j=p+1}^{\infty} n^{-\tau r}$$

$$\leq p\lambda_1^\tau + C \int_p^\infty x^{-\tau r} \, dx$$

$$= p\lambda_1^\tau + \frac{C}{\tau r - 1} \frac{1}{p^{\tau r - 1}}.$$

Since $\lambda_1 < 1$, this last upper bound tends to 0 as τ goes to infinity. Hence, for large τ, we have $f(\tau) \leq 1$, as claimed. Then (5.1) holds with $C_1 = 1$ and $q_1 = q_2 = 0$, and Theorem 5.1 yields strong polynomial tractability. The formula for the exponent of strong tractability also follows from Theorem 5.1. This completes the proof.   □

We now comment on Theorem 5.5. For $\lambda_1 > 1$, we get intractability of S since $n(\varepsilon, d)$ depends exponentially on d. If $\lambda_1 \lambda_2 \geq 1$, then $n(\varepsilon, d) \geq 2^{d-1}$, whereas for $\lambda_1 \lambda_2 < 1$, we have

$$n(\varepsilon, d) \geq \exp\left(f(\alpha)d - \tfrac{1}{2}\ln d + \Omega(1)\right),$$

where $\alpha = \ln(\lambda_1)/\ln(\lambda_1/\lambda_2) < \frac{1}{2}$ and

$$f(\alpha) = \alpha \ln \alpha^{-1} + (1 - \alpha)\ln(1 - \alpha)^{-1}.$$

Clearly, if λ_2 is small or tends to 0 then S_d tends to a continuous linear functional which (as we know) is a trivial problem for the class Λ^{all}. This corresponds to small values of $f(\alpha)$ since for small λ_2, the value of α is also small and $f(0) = 0$. It is helpful to see the graph of the function f in Figure 5.1, and observe how α affects the factor of the exponential dependence on d.

For $\lambda_1 = 1$, the problem S is polynomially intractable, but can be weakly tractable if $\lambda_2 < 1$ and $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$. We also know that weak tractability requires that $\lambda_n = o(\ln^{-2} n)$. Note that we showed these conditions directly without checking the assumptions of Theorems 5.1 and 5.3.

Note that there is not much difference between the necessary and sufficient conditions presented for weak tractability. Nevertheless, it is not clear whether $\lambda_n = o(\ln^{-2} n)$ is also sufficient for weak tractability. For such eigenvalues we have $\ln n(\varepsilon, 1) = o(\varepsilon^{-1})$. If we now return to (5.12) and take $d = \varepsilon^{-1}$ then the last term is $o(1) \ln \varepsilon^{-1}$, which is why we cannot claim that the corresponding limit is 0. We leave this as the next open problem.

Open Problem 26.

- Consider the linear tensor product problem in the worst case setting $S = \{S_d\}$ with $\lambda_1 = 1$ and $\lambda_2 < 1$. Study this problem for the absolute error criterion and for the class Λ^{all}. Verify whether

$$\lambda_n = o(\ln^{-2} n)$$
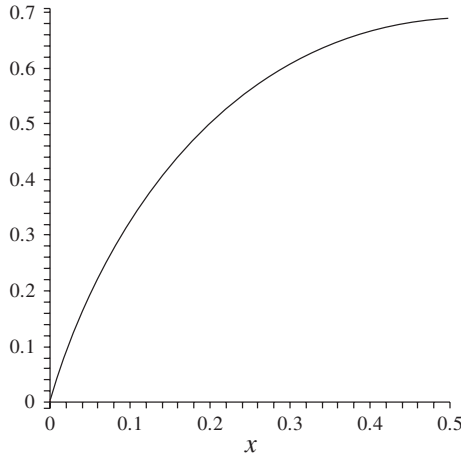
Figure 5.1. Graph of $f(x) = x \ln x^{-1} + (1-x) \ln(1-x)^{-1}$.

is a sufficient condition for weak tractability of S. If not, find a necessary and sufficient condition on $\{\lambda_n\}$ for weak tractability.

For $\lambda_1 < 1$, we show that there is no difference between strong polynomial and polynomial tractability of S and that they both hold iff the univariate eigenvalues are polynomial in n^{-1}, i.e., $\lambda_n = \mathcal{O}(n^{-r})$ for some positive r. We want to stress that the exponent of strong polynomial tractability given by the formula

$$p^{\text{str-wor}} = \inf \left\{ 2\tau \mid \sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1 \right\}$$

is, in general, not related to the speed of convergence of $\{\lambda_j\}$. Indeed, assume that $\lambda_j = Cn^{-r}$ for $r > 0$. Then $\lambda_1 < 1$ holds iff $C < 1$, and it is easy to see that

$$p^{\text{str-wor}} = \frac{2x_C}{r} \quad \text{where } \zeta(x_C) = C^{-x_C/r},$$

where, as always, ζ denotes the Riemann zeta function. Clearly, if C goes to 1 then x_C and $p^{\text{str-wor}}$ both go to infinity independently of r. On the other hand, if C goes to 0 then x_C goes to 1 and $p^{\text{str-wor}}$ goes to $2/r$.

It would be tempting to simplify the formula for the exponent of strong polynomial tractability by $p^{\text{str-wor}} = 2\tau$, where τ is given by the condition

$$\sum_{j=1}^{\infty} \lambda_j^{\tau} = 1.$$

Indeed, this simplified formula is true if such τ exists. However, in general, the existence of such τ cannot be guaranteed. Indeed, take $\lambda_n = Cn^{-1} \ln^{-2}(n+1)$ with a

positive C such that

$$C \sum_{n=1}^{\infty} \frac{1}{n \ln^2(n+1)} < 1.$$

For small C, the last inequality is satisfied since the series is convergent.

Then $\lambda_1 < 1$ and the original formula for $p^{\text{str-wor}}$ gives $p^{\text{str-wor}} \leq 2$ since we can take $\tau = 1$. However, if we take $\tau < 1$ then

$$f(\tau) := \sum_{n=1}^{\infty} \lambda_j^{\tau} = C^{\tau} \sum_{n=1}^{\infty} \frac{1}{j^{\tau} \ln^{2\tau}(n+1)} = \infty,$$

independently of C. In this case, the function f is well defined only for $[1, \infty)$ and there is no τ for which $f(\tau) = 1$. In this case, $p^{\text{str-wor}} = 2$.

We can summarize Theorem 5.5 by saying that for $\lambda_1 > 1$ we have negative tractability results, whereas for $\lambda_1 \leq 1$ we may have positive tractability results. The negative tractability results are easy to explain intuitively. Since we want to approximate the operator S_d whose norm $\|S_d\| = \lambda_1^{d/2}$ is now exponentially *large* in d, there is no surprise that the problem is exponentially hard. On the other hand, if we take $\lambda_1 < 1$ then we approximate the operator S_d whose norm is exponentially *small* in d, and the positive results are perhaps not so surprising. After all, for $\varepsilon \geq \lambda_1^{d/2}$, we can solve the problem even by the zero algorithm. As we will see later, for some classical problems related to L_2 discrepancy, we have $\lambda_1 = 3^{-1/2}$ and then for, say, $d = 360$, the zero algorithm has the worst case error $3^{-180} \approx 10^{-85.9}$. It is then really hard to imagine that someone could be interested in solving such a problem with $\varepsilon < 3^{-180}$.

This discussion suggests that linear tensor product problems are properly normalized only when $\lambda_1 = 1$. If we have at least a double largest eigenvalue, that is $\lambda_1 = \lambda_2 = 1$, the problem S is intractable. However, if the largest eigenvalue is simple, $\lambda_2 < 1$, then we still do not have polynomial tractability, but we may have weak tractability, which holds iff λ_n goes to 0 faster than $(\ln n)^{-2}(\ln \ln n)^{-2}$. We will return to this example in Chapter 8 when generalized tractability is studied and present more estimates of $n(\varepsilon, d)$ depending on the speed of convergence of $\{\lambda_n\}$.

We now turn to the normalized error criterion. The information complexity in this case is

$$n(\varepsilon, d) = \left| \{[j_1, j_2, \ldots, j_d] \in \mathbb{N}^d \mid \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_d} > \varepsilon^2 \lambda_1^d \} \right|.$$

If we define $\lambda_j' = \lambda_j / \lambda_1$ then

$$n(\varepsilon, d) = \left| \{[j_1, j_2, \ldots, j_d] \in \mathbb{N}^d \mid \lambda_{j_1}' \lambda_{j_2}' \cdots \lambda_{j_d}' > \varepsilon^2 \} \right|.$$

This corresponds to the absolute error criterion for the univariate eigenvalues λ_j' with $\lambda_1' = 1$, Hence, we obtain tractability conditions analogous to those for the absolute error criterion for the case $\lambda_1 = 1$. For completeness we summarize these conditions in the next theorem.

Theorem 5.6. *Consider the linear tensor product problem $S = \{S_d\}$ for the normalized error criterion in the worst case setting and for the class Λ^{all} with $\lambda_2 > 0$.*

- *Let $\lambda_1 = \lambda_2$. Then S is intractable and for all $\varepsilon \in (0,1)$ we have*

$$n(\varepsilon, d) \geq 2^d.$$

- *Let $\lambda_2 < \lambda_1$. Then S is polynomially intractable.*
- *If S is weakly tractable then*

$$\lambda_2 < \lambda_1 \quad and \quad \lambda_n = o\left((\ln n)^{-2}\right) \ as \ n \to \infty.$$

If

$$\lambda_2 < \lambda_1 \quad and \quad \lambda_n = o\left((\ln n)^{-2}(\ln \ln n)^{-2}\right) \ as \ n \to \infty$$

then S is weakly tractable.

The main difference between the absolute and normalized error criteria is that we may have strong polynomial tractability for the absolute error criterion but *not* for the normalized error criterion. For the latter criterion we may have only weak tractability. To regain polynomial tractability for the normalized error criterion we must introduce *weights* of the underlying spaces; this is the subject of the next section.

Example: Approximation for Korobov Spaces

We illustrate the results of this section for the multivariate approximation problem $S_d = \mathrm{APP}_d \colon H_d \to G_d$, where $\mathrm{APP}_d f = f$. We take $H_d = H_d^{\mathrm{kor}} = H_{d,\alpha}$ as the Korobov space introduced in Appendix A, and $G_d = L_2([0,1]^d)$. Since $H_{d,\alpha}$ and $L_2([0,1]^d)$ are both tensor product Hilbert spaces and App_d is a linear tensor product operator, this approximation problem is an example of a linear tensor product problem in the worst case setting.

It is easy to find the operator $W_1 = \mathrm{APP}_1^* \mathrm{APP}_1 \colon H_{1,\alpha} \to H_{1,\alpha}$. We have

$$(W_1 f)(x) = \sum_{h \in \mathbb{Z}} \varrho_{1,\alpha}^{-1}(h) \hat{f}(h) e_h(x),$$

where, as in Appendix A, the coefficients $\hat{f}(h)$'s denote the Fourier coefficients of $f \in H_{1,\alpha}$, and $\{e_h\}$, where $e_h(x) = \varrho_{1,\alpha}(h)^{-1/2} \exp(2\pi i h x)$, is an orthonormal basis of $H_{1,\alpha}$, and

$$\varrho_{1,\alpha}(h) = \beta_1^{-1} \delta_{0,h} + \beta_2^{-1}(1 - \delta_{0,h})|h|^{2\alpha}$$

with positive β_1 and β_2.

The eigenpairs of W_1 are $\{\lambda_h, e_h\}_{h \in \mathbb{Z}}$ with $\lambda_h = \varrho_{1,\alpha}^{-1}(h)$. We have

$$\{\lambda_h\}_{h \in \mathbb{Z}} = \left\{ \beta_1, \beta_2, \beta_2, \frac{\beta_2}{2^{2\alpha}}, \frac{\beta_2}{2^{2\alpha}}, \ldots, \frac{\beta_2}{j^{2\alpha}}, \frac{\beta_2}{j^{2\alpha}}, \ldots \right\}.$$

It is easy to see that the largest eigenvalue is $\lambda_1 = \max(\beta_1, \beta_2)$, and the second largest eigenvalue is $\lambda_2 = \beta_2$. Furthermore, $\lambda_1 = \lambda_2$ iff $\beta_1 \leq \beta_2$.

For $d > 1$, the eigenpairs of $W_d = \text{APP}_d^* \text{APP}_d : H_{d,\alpha} \to H_{d,\alpha}$ are $\{\lambda_h, e_h\}_{h \in \mathbb{Z}^d}$. As always for linear tensor product problems, the eigenvalues are of product form and

$$\{\lambda_h\}_{h \in \mathbb{Z}^d} = \{\lambda_{h_1} \lambda_{h_2} \cdots \lambda_{h_d}\}_{h_j \in \mathbb{Z}}.$$

Assume for a moment that $\alpha = 0$. For $d = 1$, we then have at most two distinct eigenvalues β_1 and β_2, and the eigenvalue β_2 has infinite multiplicity. This means that W_1 is *not* compact. For $d \geq 1$, we have eigenvalues $\beta_1^k \beta_2^{d-k}$ for $k = 0, 1, \ldots, d$ and each eigenvalue $\beta_1^k \beta_2^{d-k}$ for $k = 0, 1, \ldots, d-1$ has infinite multiplicity. Hence W_d is *not* compact and $\|W_d\| = \max(\beta_1, \beta_2)^d$. For $\varepsilon^2 < \max(\beta_1, \beta_2)^{d-1} \beta_2$, the information complexity $n(\varepsilon, d) = \infty$, and therefore the approximation problem is intractable for both the absolute and normalized error criteria.

For $\alpha > 0$, the eigenvalues converge to 0 and the operators W_d are compact. We now apply Theorem 5.5 for the approximation problem $\text{APP} = \{\text{APP}_d\}$ and the absolute error criterion, and conclude:

- APP is intractable iff $\max(\beta_1, \beta_2) > 1$ or $\beta_1 \leq \beta_2 = 1$.

- APP is weakly tractable iff $\beta_2 < \beta_1 = 1$ or $\max(\beta_1, \beta_2) < 1$.

- APP is strongly polynomially tractable iff APP is polynomially tractable iff $\max(\beta_1, \beta_2) < 1$. If this holds then the exponent of strong polynomial tractability is $p^{\text{str-wor}} = 2\tau$ with $\tau > (2\alpha)^{-1}$ being the unique solution of

$$\beta_1^\tau + 2\beta_2^\tau \zeta(2\tau\alpha) = 1. \tag{5.14}$$

We now switch to the normalized error criterion. From Theorem 5.6 we conclude:

- APP is intractable iff $\beta_1 \leq \beta_2$.

- APP is weakly tractable iff $\beta_2 < \beta_1$.

- APP is polynomially intractable independently of β_1 and β_2.

We stress that weak and polynomial tractability for both the absolute and normalized criteria depends only on β_1 and β_2 as long as $\alpha > 0$. We can achieve polynomial tractability only for the absolute error criterion iff both β_1 and β_2 are smaller than one. In this case, we even obtain strong polynomial tractability, and the exponent $p^{\text{str-wor}}$ also depends on α. Clearly, if α goes to infinity then $p^{\text{str-wor}}$ tends to $2\tau^*$ with τ^* satisfying the equation

$$\beta_1^{\tau^*} + 2\beta_2^{\tau^*} = 1.$$

For $\beta_1 = \beta_2 = \beta < 1$ we have $\tau^* = \ln(3)/\ln(\beta^{-1})$.

If α goes to 0 then $p^{\text{str-wor}}$ tends to infinity. On the other hand, if α is fixed and $\max(\beta_1, \beta_2)$ tends to 0 then the exponent $p^{\text{str-wor}}$ tends to $(2\alpha)^{-1}$.

We indicated in Appendix A that for $\alpha = r \geq 1$ being an integer, the choice $\beta_1 = 1$ and $\beta_2 = (2\pi)^{-r}$ leads to the norm involving only derivatives of functions. Note that the absolute and normalized errors are now the same, and we have weak tractability and polynomial intractability for such β_i.

Example: Schrödinger Equation (Continued)

We know that the linear Schrödinger problem is weakly or polynomially tractable if the initial condition functions f belong to a space $\tilde{H}_d$ that is a subset of the space H_d defined by (5.10) and the approximation problem for $\tilde{H}_d$ is weakly or polynomially tractable. Could we choose $\tilde{H}_d$ as the Korobov space $H_d^{\mathrm{kor}} = H_{d,\alpha}$? The answer is *no* since the constant function $f(x) \equiv 1$ belongs to the Korobov space but not to the space H_d.

For simplicity we restrict the Korobov space to functions which are orthogonal to all $\eta_{d,j}$ with odd components. More precisely, for $j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d$, let $2j = [2j_1, 2j_2, \ldots, 2j_d]$ and consider the function $\eta_{d,2j}$ given by (5.9). We first compute the Fourier coefficients of this function. Let $h = [h_1, h_2, \ldots, h_d] \in \mathbb{Z}^d$ and let $|h| = [|h_1|, |h_2|, \ldots, |h_d|]$. Then for $|h| \neq 2j$ we have

$$\widehat{\eta_{d,2j}}(h) = 0,$$

whereas for $|h| = 2j$, i.e., $h_k = \pm 2j_k$, we have

$$\widehat{\eta_{d,2j}}(h) = i^d,$$

and 2^d Fourier coefficients of $\eta_{d,2j}$ are non-zero. Hence $\eta_{d,2j} \in H_{d,\alpha}$. Since $\varrho_{d,\alpha}(h) = \varrho_{d,\alpha}(|h|)$ we obtain $\|\eta_{d,2j}\|_{H_{d,\alpha}} = 2^{d/2} \varrho_{d,\alpha}^{1/2}(2j)$. Therefore

$$\tilde{\eta}_{d,2j} = 2^{-d/2} \varrho_{d,\alpha}^{-1/2}(2j) \eta_{d,2j} \quad \text{for all } j \in \mathbb{N}^d$$

are orthonormal in $H_{d,\alpha}$ and orthogonal in $L_2 = L_2([0,1]^d)$. We define

$$\tilde{H}_d = \big\{ f \in L_2 \mid f = \textstyle\sum_{j \in \mathbb{N}^d} \big\langle f, \tilde{\eta}_{d,2j} \big\rangle_{L_2} \tilde{\eta}_{d,2j} \text{ with } \sum_{j \in \mathbb{N}^d} \big| \big\langle f, \tilde{\eta}_{d,2j} \big\rangle_{L_2} \big|^2 < \infty \big\}.$$

Note that for $f \in \tilde{H}_d$ we have

$$
\begin{aligned}
\|f\|_{H_{d,\alpha}}^2 &= \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}(h) |\hat{f}(h)|^2 \\
&= \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}(h) \Big| \sum_{j \in \mathbb{N}^d} \big\langle f, \tilde{\eta}_{d,2j} \big\rangle_{L_2} 2^{-d/2} \varrho_{d,\alpha}^{-1/2}(2j) \widehat{\eta_{d,2j}}(h) \Big|^2 \\
&= \sum_{h \in \mathbb{Z}^d, j \in \mathbb{N}^d} \big| \big\langle f, \tilde{\eta}_{d,2j} \big\rangle_{L_2} \big|^2 \delta_{|h|,2j} 2^{-d} \\
&= \sum_{j \in \mathbb{N}^d} \big| \big\langle f, \tilde{\eta}_{d,2j} \big\rangle_{L_2} \big|^2 < \infty.
\end{aligned}
$$

This means that $\tilde{H}_d \subseteq H_{d,\alpha}$.

We now show that for $\alpha \geq 2$, we have $\tilde{H}_d \subseteq H_d$. For $j \in \mathbb{N}^d$ and $f \in L_2$, we obtain

$$
\begin{aligned}
f_{2j} &= \left\langle f, \eta_{d,2j} \right\rangle_{L_2} \\
&= 2^{d/2} \sum_{h \in \mathbb{Z}^d} \hat{f}(h) \prod_{k=1}^{d} \int_0^1 \exp\left(2\pi \mathrm{i} h_k x\right) \sin\left(2\pi j_k x\right) \mathrm{d}x \\
&= \frac{\mathrm{i}^d}{2^{d/2}} \sum_{\varepsilon_j \in \{-1,1\}} (-1)^{\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_d} \, \hat{f}\left(2\varepsilon_1 j_1, 2\varepsilon_2 j_2, \ldots, 2\varepsilon_d j_d\right).
\end{aligned}
$$

Therefore, for $f \in \tilde{H}_d$ we have

$$
\begin{aligned}
\sum_{j \in \mathbb{N}^d} \beta_{d,j}^{-2} |f_j|^2 &= \sum_{j \in \mathbb{N}^d} \beta_{d,2j}^{-2} |f_{2j}|^2 \\
&\leq \sum_{j \in \mathbb{N}^d} \sum_{\varepsilon_j \in \{-1,1\}} \left(q_0 + 4\pi^2 \sum_{k=1}^{d} j_k^2\right)^2 |\hat{f}(2\varepsilon_1 j_1, 2\varepsilon_2 j_2, \ldots, 2\varepsilon_d j_d)|^2 \\
&= \sum_{j \in \mathbb{N}^d} \sum_{\varepsilon_j \in \{-1,1\}} \frac{\left(q_0 + 4\pi^2 \sum_{k=1}^{d} j_k^2\right)^2}{\varrho_\alpha(2j)} \varrho_\alpha(2j) |\hat{f}(2\varepsilon_1 j_1, 2\varepsilon_2 j_2, \ldots, 2\varepsilon_d j_d)|^2 \\
&\leq \|f\|_{H_{d,\alpha}}^2 \sup_{j \in \mathbb{N}^d} \frac{\left(q_0 + 4\pi^2 \sum_{k=1}^{d} j_k^2\right)^2}{\varrho_\alpha(2j)}.
\end{aligned}
$$

Hence, we conclude that $\tilde{H}_d \subseteq H_d$ iff

$$
A := \sup_{j \in \mathbb{N}^d} \frac{\left(q_0 + 4\pi^2 \sum_{k=1}^{d} j_k^2\right)^2}{\varrho_\alpha(2j)} = \left(\frac{\beta_2}{2^{2\alpha}}\right)^d \sup_{j \in \mathbb{N}^d} \frac{\left(q_0 + 4\pi^2 \sum_{k=1}^{d} j_k^2\right)^2}{\prod_{k=1}^{d} j_k^{2\alpha}} < \infty.
$$

It is easy to check that $A < \infty$ iff $\alpha \geq 2$. Indeed, if $A < \infty$ then we can take $j_k = 1$ for all k except, say, p and go with j_p to infinity. Then 2α must be at least 4, i.e., $\alpha \geq 2$. On the other hand, if $\alpha \geq 2$ then

$$
A \leq \left(\frac{\beta_2}{2^{2\alpha}}\right)^d \sup_{j \in \mathbb{N}^d} \frac{\left(q_0 + 4d \max_{k \in [d]} j_k^2\right)^2}{(\max_{k \in [d]} j_k)^{2\alpha}} < \infty,
$$

as claimed.

Hence, for $\alpha \geq 2$ we can apply the previous result that tractability of the linear Schrödinger problem for $\tilde{H}_d$ is equivalent to tractability of the approximation problem for $\tilde{H}_d$. Consider then the approximation problem $\mathrm{APP}_d f = f$ with $\mathrm{APP}_d : \tilde{H}_d \to L_2$. It is easy to check that $W_d = \mathrm{APP}_d^* \mathrm{APP}_d : \tilde{H}_d \to \tilde{H}_d$ has the eigenpairs

$$
W_d \tilde{\eta}_{d,2j} = \varrho_{d,\alpha}^{-1} \tilde{\eta}_{d,2j} \quad \text{for all } j \in \mathbb{N}^d.
$$

Hence, the eigenvalues are

$$\lambda_{d,j} = \varrho_{d,\alpha}^{-1} = \prod_{k=1}^{d} \frac{\beta_2}{2^{2\alpha} j_k^{2\alpha}} \quad \text{for all } j \in \mathbb{N}^d.$$

They are of product form and for $d = 1$ the largest eigenvalue

$$\lambda_1 = \frac{\beta_2}{2^{2\alpha}}$$

has multiplicity 1 since the second largest eigenvalue is

$$\lambda_2 = \frac{\beta_2}{2^{4\alpha}}.$$

We now apply Theorem 5.5 for the approximation problem $\text{APP} = \{\text{APP}_d\}$ and the absolute error criterion, and conclude:

- the linear Schrödinger problem for $\widetilde{H}_d$ is intractable iff $\beta_2 > 2^{2\alpha}$.

- the linear Schrödinger problem for $\widetilde{H}_d$ is weakly tractable iff $\beta_2 \leq 2^{2\alpha}$.

- the linear Schrödinger problem for $\widetilde{H}_d$ is strongly polynomially tractable iff the linear Schrödinger problem for $\widetilde{H}_d$ is polynomially tractable iff $\beta_2 < 2^{2\alpha}$. If this holds then the exponent of strong polynomial tractability is $p^{\text{str-wor}} = 2\tau$ with $\tau > (2\alpha)^{-1}$ being the unique solution of

$$\left(\frac{\beta_2}{2^{2\alpha}}\right)^{\tau} \zeta(2\tau\alpha) = 1.$$

We now switch to the normalized error criterion. Clearly, the initial error is $\lambda_{d,\vec{1}}^{1/2}$ where $\vec{1} = [1, 1, \ldots, 1]$, and $\lambda_{d,j}/\lambda_{d,\vec{1}} = \prod_{k=1}^{d} j_k^{-2\alpha}$ is independent of β_2 for all $j \in \mathbb{N}^d$. From Theorem 5.6 we conclude:

- the linear Schrödinger problem for $\widetilde{H}_d$ is weakly tractable.

- the linear Schrödinger problem for $\widetilde{H}_d$ is polynomially intractable.

5.3 Linear Weighted Tensor Product Problems

In this section we introduce *weights* which play a major role in tractability studies. We first introduce weighted Hilbert spaces of d-variate functions, with each group of variables controlled by its corresponding weight. Since we have 2^d groups of variables indexed by subsets $\mathfrak{u}$ of the set $\{1, 2, \ldots, d\}$, we have 2^d weights each

denoted by $\gamma_{d,\mathfrak{u}}$. The definition of weighted Hilbert space is similar to the ANOVA decomposition presented in Chapter 3.

Then we discuss different types of weights such as product weights, order-dependent weights, finite-order weights, and finite-diameter weights. Our special emphasis is on product weights, where the jth variable is controlled by the weight $\gamma_{d,j}$ and the groups of variables in $\mathfrak{u}$ are controlled by $\prod_{j\in\mathfrak{u}} \gamma_{d,j}$, and on finite-order weights where the weights $\gamma_{d,\mathfrak{u}}$ are 0 if the cardinality of the subset $\mathfrak{u}$ is larger than, say, $|\mathfrak{u}| > \omega^*$ for some ω^* independent of d.

Having weighted Hilbert spaces, it is easy to obtain linear weighted tensor product operators and to study their polynomial and weak tractability. The main point is to find necessary and sufficient conditions on the weights such that polynomial or weak tractability holds. By now, it should be intuitively clear that for sufficiently quickly decaying weights we should obtain positive tractability results. In this way, we may change many negative results on the curse of dimensionality or polynomial intractability for the unweighted linear problems by shrinking the classes of functions to corresponding weighted classes.

5.3.1 Weighted Hilbert Spaces

We now show how the ANOVA decomposition presented in Section 3.1.6 of Chapter 3 can be generalized to an arbitrary tensor product Hilbert space. This will allow us to introduce weights measuring the influence of each groups of variables and obtain weighted Hilbert spaces.

As in the previous section, let $H_d = H_1 \otimes H_1 \otimes \cdots \otimes H_1$ be the d-fold tensor product space of real valued functions defined on D_1^d, where H_1 is a separable Hilbert space of univariate real functions defined on $D_1 \subseteq \mathbb{R}$.

To simplify notation, we assume that $\dim(H_1) = \infty$. Let $\{\eta_j\}_{j\in\mathbb{N}}$ be an orthonormal basis of H_1, i.e., $\langle \eta_i, \eta_j \rangle_{H_1} = \delta_{i,j}$. In general, η_1 can be chosen as any normalized element of H_1. However, it will be simpler if we take $\eta_1 \equiv 1$, which obviously can be done as long as $1 \in H_1$ and $\|1\|_{H_1} = 1$.

For $d \geq 2$, let $j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d$ and $x = [x_1, x_2, \ldots, x_d] \in D_1^d$. Define

$$\eta_j(x) = \prod_{k=1}^d \eta_{j_k}(x_k).$$

Then $\{\eta_j\}_{j\in\mathbb{N}^d}$ is an orthonormal basis of H_d and for any function f from H_d we have

$$f(x) = \sum_{j\in\mathbb{N}^d} c_j \eta_j(x),$$

where $c_j = \langle f, \eta_j \rangle_{H_d}$. Note that $\|f\|_{H_d}^2 = \sum_{j\in\mathbb{N}^d} c_j^2 < \infty$.

It will be convenient to use the notation

$$[d] := \{1, 2, \ldots, d\} \quad \text{for all } d \in \mathbb{N}.$$

For each $j \in \mathbb{N}^d$ define

$$\mathfrak{u}(j) = \{k \mid j_k \geq 2\}.$$

Clearly, $\mathfrak{u}(j) \subseteq [d]$ and by varying j we can obtain any of the 2^d subsets of $[d]$. For example, for $j = [1, 1, \ldots, 1]$ we have $\mathfrak{u}(j) = \emptyset$, and for any $j = [j_1, j_2, \ldots, j_d]$ with all $j_k \geq 2$ we have $\mathfrak{u}(j) = [d]$. For $\mathfrak{u} \subseteq [d]$, by $\bar{\mathfrak{u}}$ we denote the complement of $\mathfrak{u}$, i.e, $\bar{\mathfrak{u}} = [d] \setminus \mathfrak{u}$.

We can rewrite the function η_j as

$$\eta_j(x) = \prod_{k \in \mathfrak{u}(j)} \eta_{j_k}(x_k) \prod_{k \in \overline{\mathfrak{u}(j)}} \eta_1(x_k).$$

Hence, if $\eta_1 \equiv 1$ the second product is one, and the function η_j depends *only* on variables from $\mathfrak{u}(j)$. For general η_1, the function η_j depends on variables not in $\mathfrak{u}(j)$ in a preassigned way through the function η_1.

The series for the function f can be also rewritten as

$$\begin{aligned}
f(x) &= \sum_{j \in \mathbb{N}^d} c_j \prod_{k \in \mathfrak{u}(j)} \eta_{j_k}(x_k) \prod_{k \in \overline{\mathfrak{u}(j)}} \eta_1(x_k) \\
&= \sum_{\mathfrak{u} \subseteq [d]} \left(\sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} c_j \prod_{k \in \mathfrak{u}} \eta_{j_k}(x_k) \right) \prod_{k \in \bar{\mathfrak{u}}} \eta_1(x_k).
\end{aligned}$$

For any subset $\mathfrak{u}$ of $[d]$, let $f_{\mathfrak{u}} : D_1^d \to \mathbb{R}$ be defined by

$$f_{\mathfrak{u}}(x) = \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} c_j \eta_j(x) = \left(\sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} c_j \prod_{k \in \mathfrak{u}} \eta_{j_k}(x_k) \right) \prod_{k \in \bar{\mathfrak{u}}} \eta_1(x_k).$$

Hence, $f_{\mathfrak{u}}$ is represented as the product of two factors

$$f_{\mathfrak{u}}(x) = f_{\mathfrak{u},1}(x_{\mathfrak{u}}) h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}),$$

where

$$f_{\mathfrak{u},1}(x_{\mathfrak{u}}) := \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} c_j \prod_{k \in \mathfrak{u}} \eta_{j_k}(x_k) \quad \text{and} \quad h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) := \prod_{k \in \bar{\mathfrak{u}}} \eta_1(x_k).$$

For example, take $\mathfrak{u} = \emptyset$. Then $f_{\emptyset,1}(x_\emptyset) = \langle f, \eta_{1,1,\ldots,1} \rangle_{H_d}$ and $h_{\emptyset,2}(x_\emptyset) = h_{[d],2}(x) = \prod_{k=1}^d \eta_1(x_k)$. For $\mathfrak{u} = [d]$ we have $f_{[d],1}(x) = f_{[d]}(x)$ and $h_{\emptyset,2}(x_\emptyset) = 1$.

For an arbitrary $\mathfrak{u}$, the first factor $f_{\mathfrak{u},1}$ is a function that depends only on variables in $\mathfrak{u}$ and belongs to the tensor product Hilbert space $H_{\mathfrak{u}}$ of $|\mathfrak{u}|$ copies of $\tilde{H}_1$. Here

$$\tilde{H}_1 = \{f \in H_1 \mid \langle f, \eta_1 \rangle_{H_1} = 0\}$$

is a closed subspace of H_1, and $\{\eta_j\}_{j \in \mathbb{N} \setminus \{1\}}$ is its orthonormal basis.

The second factor $h_{\bar{\mathfrak{u}},2}$ is a function that depends only on variables not in $\mathfrak{u}$ through the function η_1. If $\eta_1 \equiv 1$, then the second factor $h_{\bar{\mathfrak{u}},2}$ is always one. For general η_1,

the function $f_\mathfrak{u}$ has a preassigned dependence on variables not in $\mathfrak{u}$ and may have arbitrary dependence on variables in $\mathfrak{u}$ as long as the first factor $f_{\mathfrak{u},1}$ belongs to $H_\mathfrak{u}$. In this sense we say that $f_\mathfrak{u}$ depends only on variables in $\mathfrak{u}$.

For $\mathfrak{u}, \mathfrak{v} \subseteq [d]$ and $\mathfrak{u} \neq \mathfrak{v}$ we have

$$\langle f_\mathfrak{u}, f_\mathfrak{v} \rangle_{H_d} = \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} \sum_{k \in \mathbb{N}^d : \mathfrak{u}(k) = \mathfrak{v}} c_j c_k \langle \eta_j, \eta_k \rangle_{H_d} = 0$$

since $j \neq k$. By the same argument, we have for $f, g \in H_d$ and $\mathfrak{u} \neq \mathfrak{v}$,

$$\langle f_\mathfrak{u}, g_\mathfrak{v} \rangle_{H_d} = 0.$$

Hence, the $\{f_\mathfrak{u}\}$ are orthogonal. For $\mathfrak{u} = \mathfrak{v}$, we have

$$\|f_\mathfrak{u}\|_{H_d}^2 = \langle f_\mathfrak{u}, f_\mathfrak{u} \rangle_{H_d} = \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} c_j^2 = \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} \langle f_\mathfrak{u}, \eta_j \rangle_{H_d}^2.$$

Going back to the series of f, we conclude that

$$f = \sum_{\mathfrak{u} \subseteq [d]} f_u.$$

Hence, we have decomposed an arbitrary function f from H_d as the sum of mutually orthogonal functions $f_\mathfrak{u}$, each of which depends only on variables from $\mathfrak{u}$. This resembles the ANOVA decomposition (3.9) presented in Section 3.1.6 of Chapter 3, especially if we can take $\eta_1 \equiv 1$.

For $f, g \in H_d$, note that

$$\begin{aligned}
\langle f_\mathfrak{u}, g_\mathfrak{u} \rangle_{H_d} &= \langle f_{\mathfrak{u},1} h_{\bar{\mathfrak{u}},2}, g_{\mathfrak{u},1} h_{\bar{\mathfrak{u}},2} \rangle_{H_d} \\
&= \langle f_{\mathfrak{u},1}, g_{\mathfrak{u},1} \rangle_{H_\mathfrak{u}} \langle h_{\bar{\mathfrak{u}},2}, h_{\bar{\mathfrak{u}},2} \rangle_{H_{\bar{\mathfrak{u}}}} \\
&= \langle f_{\mathfrak{u},1}, g_{\mathfrak{u},1} \rangle_{H_\mathfrak{u}}.
\end{aligned}$$

Hence we have

$$\langle f, g \rangle_{H_d} = \sum_{\mathfrak{u} \subseteq [d]} \langle f_\mathfrak{u}, g_\mathfrak{u} \rangle_{H_d} = \sum_{\mathfrak{u} \subseteq [d]} \langle f_{\mathfrak{u},1}, g_{\mathfrak{u},1} \rangle_{H_\mathfrak{u}},$$

and

$$\|f\|_{H_d}^2 = \sum_{\mathfrak{u} \subseteq [d]} \|f_\mathfrak{u}\|_{H_d}^2 = \sum_{\mathfrak{u} \subseteq [d]} \|f_{\mathfrak{u},1}\|_{H_\mathfrak{u}}^2.$$

The formula for the norm of f tells us that the contribution of each $f_\mathfrak{u}$ is the same. This means that any group of variables is equally important in their contribution to the norm of f.

Suppose that we know a priori that some groups of variables are more important than the others or that the function f does not depend on some groups of variables. Such situations can be modeled by introducing a sequence of *weights* $\gamma = \{\gamma_{d,\mathfrak{u}}\}$,

where $d \in \mathbb{N}$ and $\mathfrak{u}$ is an arbitrary subset of $[d]$. We assume first that all $\gamma_{d,\mathfrak{u}}$ are positive. The zero weight $\gamma_{d,\mathfrak{u}}$ can be obtained by the limiting process and by adopting the convention that $0/0 = 0$.

For positive weights, we define the *weighted* Hilbert space $H_{d,\gamma}$ as a separable Hilbert space that is algebraically the same as the space H_d, and whose inner product for $f, g \in H_d$ is given by

$$\langle f, g \rangle_{H_{d,\gamma}} := \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \langle f_\mathfrak{u}, g_\mathfrak{u} \rangle_{H_d} = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \langle f_{\mathfrak{u},1}, g_{\mathfrak{u},1} \rangle_{H_\mathfrak{u}}.$$

Note that the norms of H_d and $H_{d,\gamma}$ are equivalent, with

$$\frac{1}{\max_\mathfrak{u} \gamma_{d,\mathfrak{u}}^{1/2}} \, \|f\|_{H_d} \; \le \; \|f\|_{H_{d,\gamma}} \; \le \; \frac{1}{\min_\mathfrak{u} \gamma_{d,\mathfrak{u}}^{1/2}} \, \|f\|_{H_d},$$

although the equivalence factors may arbitrarily depend on d. We also have

$$\|f\|_{H_{d,\gamma}}^2 = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \|f_\mathfrak{u}\|_{H_d}^2 = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \sum_{j \in \mathbb{N}^d : \mathfrak{u}(j) = \mathfrak{u}} \langle f_\mathfrak{u}, \eta_j \rangle_{H_d}^2.$$

For all $j \in \mathbb{N}^d$ and $\mathfrak{u} \subseteq [d]$, define

$$\eta_{j,\gamma} = \gamma_{d,\mathfrak{u}(j)}^{1/2} \, \eta_j.$$

It is easy to check that $\{\eta_{j,\gamma}\}_{j \in \mathbb{N}^d}$ is an orthonormal basis of $H_{d,\gamma}$. It is also easy to verify that for all $f \in H_d$ and $j \in \mathbb{N}^d$, we have

$$\langle f, \eta_j \rangle_{H_d} = \gamma_{d,\mathfrak{u}(j)}^{1/2} \langle f, \eta_{j,\gamma} \rangle_{H_{d,\gamma}} = \gamma_{d,\mathfrak{u}(j)} \langle f, \eta_j \rangle_{H_{d,\gamma}}.$$

We now consider general weights for which some $\gamma_{d,\mathfrak{u}}$ may be 0. If $\gamma_{d,\mathfrak{u}} = 0$ then we assume that $f_\mathfrak{u} = 0$, and the term for $\mathfrak{u}$ is absent in the inner product formula. In this case, the space $H_{d,\gamma}$ is a *proper* subspace of H_d and consists of linear combinations of functions $f_\mathfrak{u}$ corresponding to non-zero weights $\gamma_{d,\mathfrak{u}}$. Define

$$\mathbb{N}_\gamma^d = \{ j \in \mathbb{N}^d \mid \gamma_{d,\mathfrak{u}(j)} > 0 \}$$

as the set of indices j for which the corresponding weight $\gamma_{d,\mathfrak{u}(j)}$ is positive. Then $\{\eta_{j,\gamma}\}_{j \in \mathbb{N}_\gamma^d}$ is an orthonormal basis of $H_{d,\gamma}$. We always assume that for each d at least one weight $\gamma_{d,\mathfrak{u}(j)}$ is positive, and therefore the set $\mathbb{N}_\gamma^d$ is never empty. In fact, if the only non-zero weight is $\gamma_{d,\emptyset}$ then $\mathbb{N}_\gamma^d = \{[1, 1, \ldots, 1]\}$; otherwise $\mathbb{N}_\gamma^d$ has infinitely many elements.

Consider the unit ball of the space $H_{d,\gamma}$. From the first formula for the norm of f, it is clear that we can change the influence of each $f_\mathfrak{u}$ by a proper choice of the weight $\gamma_{d,\mathfrak{u}}$. For small $\gamma_{d,\mathfrak{u}}$, the functions in the unit ball of $H_{d,\gamma}$ must have small $f_\mathfrak{u}$, and if $\gamma_{d,\mathfrak{u}} = 0$ then we know a priori that all $f_\mathfrak{u} = 0$, which means that we consider

functions that do not depend on the group $\mathfrak{u}$ of variables. For example, if $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u}$ for which $|\mathfrak{u}| \geq 3$, then we know a priori that functions can depend only on groups of two variables. In the sequel, we will specify different conditions on weights that control the behavior of functions in a more qualitative way.

We stress that for general weights the space $H_{d,\gamma}$ is not a tensor product space, although its construction is based on the tensor product spaces $H_{\mathfrak{u}}$. For example, consider the weights $\gamma_{d,\mathfrak{u}} = 1$ for all $\mathfrak{u}$ for which $|\mathfrak{u}| \leq \omega$ for some integer ω independent of d, and $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u}$ for which $|\mathfrak{u}| > \omega$. Then $H_{d,\gamma}$ is the direct sum of the tensor product spaces $H_{\mathfrak{u}}$ for all $\mathfrak{u}$ with $|\mathfrak{u}| \leq \omega$. On the other hand, if $\gamma_{d,\mathfrak{u}} = \prod_{k \in \mathfrak{u}} \gamma_{d,k}$ for some positive $\gamma_{d,k}$, then the space $H_{d,\gamma}$ is the tensor product space $H_{1,\gamma_{d,1}} \otimes H_{1,\gamma_{d,2}} \otimes \cdots \otimes H_{d,\gamma_{d,d}}$. In particular, if $\gamma_{d,k} = 1$ for all k then we obviously get $H_{d,\gamma} = H_d$.

We finally comment on the case when the Hilbert space H_1 has a reproducing kernel $K_1 \colon D_1 \times D_1 \to \mathbb{R}$. This property is not needed for the class Λ^{all}; however, it is necessary for the class Λ^{std} that will be extensively studied later in Volumes II and III. For an arbitrary point $x \in D_1$, define the linear functional $L_x(f) = f(x)$ for all $f \in H_1$. It is well known, see the book of Aronszajn [2], that a Hilbert space has a reproducing kernel iff the L_x's are continuous. Hence, we must assume that H_1 has a reproducing kernel when we study the class Λ^{std}.

The basic properties of reproducing kernel Hilbert spaces are collected in Appendix A. Here we only mention that $K(x, y) = K(y, x)$ for all $x, y \in D_1$, $K_1(\cdot, x) \in H_1$ for any $x \in D_1$, and

$$f(x) = \langle f, K_1(\cdot, x) \rangle_{H_1} \quad \text{for all } f \in H_1 \text{ and } x \in D_1.$$

This is the main property of a reproducing kernel Hilbert space, which makes it possible to deduce many important properties of the whole Hilbert space from the reproducing kernel alone.

It is easy to see that if H_1 is a reproducing kernel Hilbert space, then the weighted Hilbert space $H_{d,\gamma}$ is also a reproducing kernel Hilbert space for any choice of the weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$. Furthermore, the reproducing kernel $K_{d,\gamma}$ of the space $H_{d,\gamma}$ is fully determined by the kernel K_1, the function η_1 and the weights $\gamma_{d,\mathfrak{u}}$. We have

$$K_{d,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} K_{\mathfrak{u}}(x_{\mathfrak{u}}, y_{\mathfrak{u}}) h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) h_{\bar{\mathfrak{u}},2}(y_{\bar{\mathfrak{u}}}) \quad \text{for all } x, y \in D_1^d,$$

where the reproducing kernel $K_{\mathfrak{u}}$ of the space $H_{\mathfrak{u}}$ has the form

$$K_{\mathfrak{u}}(x_{\mathfrak{u}}, y_{\mathfrak{u}}) = \prod_{k \in \mathfrak{u}} K_1(x_k, y_k).$$

Indeed, observe first that

$$\left(K_{d,\gamma}(\cdot, x) \right)_{\mathfrak{u},1} = \gamma_{d,\mathfrak{u}} h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) K_{\mathfrak{u}}(\cdot, x_{\mathfrak{u}}).$$

Since $f = \sum_{\mathfrak{u}} f_{\mathfrak{u}}$ with $f_{\mathfrak{u},1} \in H_{\mathfrak{u}}$ and $f_{\mathfrak{u},1}(x_{\mathfrak{u}}) = \langle f_{\mathfrak{u},1}, K_{\mathfrak{u}}(\cdot, x_{\mathfrak{u}}) \rangle_{H_{\mathfrak{u}}}$, we have

$$
\begin{aligned}
f(x) &= \sum_{\mathfrak{u} \subseteq [d]} f_{\mathfrak{u},1}(x_{\mathfrak{u}}) h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) \\
&= \sum_{\mathfrak{u} \subseteq [d]} \langle f_{\mathfrak{u},1}, K_{\mathfrak{u}}(\cdot, x_{\mathfrak{u}}) \rangle_{H_{\mathfrak{u}}} h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) \\
&= \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \langle f_{\mathfrak{u},1}, \gamma_{d,\mathfrak{u}} h_{\bar{\mathfrak{u}},2}(x_{\bar{\mathfrak{u}}}) K_{\mathfrak{u}}(\cdot, x_{\mathfrak{u}}) \rangle_{H_{\mathfrak{u}}} \\
&= \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \langle f_{\mathfrak{u},1}, \left(K_{d,\gamma}(\cdot, x) \right)_{\mathfrak{u},1} \rangle_{H_{\mathfrak{u}}} = \langle f, K_{d,\gamma}(\cdot, x) \rangle_{H_d},
\end{aligned}
$$

as claimed.

For general weights, the reproducing kernel $K_{d,\gamma}$ cannot be represented as the product of univariate reproducing kernels. However, if $\gamma_{d,\mathfrak{u}} = c_d \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ then

$$
K_{d,\gamma}(x, y) = c_d \prod_{j=1}^{d} \left(\eta_1(x_j) \eta_1(y_j) + \gamma_{d,j} K_1(x_j, y_j) \right).
$$

5.3.2 Types of Weights

So far we considered an arbitrary sequence of weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ with $\gamma_{d,\mathfrak{u}} \geq 0$. In this section we specify some assumptions on the weights $\gamma_{d,\mathfrak{u}}$. We obtain different types of weights for which we later prove tractability results.

The first type of weights studied for tractability of multivariate problems was product weights, see [216] where such weights were introduced with no dependence on d, and [271] where the dependence of the product weights on d was allowed. The first use of general weights $\gamma_{d,\mathfrak{u}}$ can be found in Example 1 of [94] and was suggested by Art Owen. The reader is referred to [177] for a survey of tractability results for product weights mostly for multivariate integration.

We say that $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ is a sequence of *product weights* iff for all $d \in \mathbb{N}$ and for all $\mathfrak{u} \subseteq [d]$ we have

$$
\gamma_{d,\mathfrak{u}} = c_d \prod_{j \in \mathfrak{u}} \gamma_{d,j}
$$

for some positive c_d and some arbitrary non-negative sequence $\{\gamma_{d,j}\}$ such that

$$
\gamma_{d,d} \leq \gamma_{d,d-1} \leq \cdots \leq \gamma_{d,2} \leq \gamma_{d,1}.
$$

If additionally $\gamma_{d,j} = \gamma_j$ then γ is called a sequence of *product weights independent of dimension*. Here, we use the convention that the product over the empty set is one. That is, $\gamma_{d,\emptyset} = c_d$. In many cases, $c_d = 1$ has been studied but, as we shall see later, it is sometimes convenient to allow an arbitrary positive c_d. For product weights, all 2^d weights $\gamma_{d,\mathfrak{u}}$ are specified by $d + 1$ numbers.

The essence of product weights is that we monitor the influence of the jth variable for the d-dimensional case through the weight $\gamma_{d,j}$. The ordering of $\gamma_{d,j}$ means that the first variable is the most important one, followed by the second variable, and so on. Obviously, if all $\gamma_{d,j}$ are the same, say $\gamma_{d,j} \equiv \beta_d$, then all variables are equally important. For product weights, the group of variables $x_{\mathfrak{u}}$ is monitored by the weight $c_d \prod_{j \in \mathfrak{u}} \gamma_{d,j}$. If all $\gamma_{d,j}$'s are equal to β_d then the group of variables $x_{\mathfrak{u}}$ is monitored by the weight $c_d \beta_d^{|\mathfrak{u}|}$. Observe that all groups of variables are equally important only for $\beta_d = 1$. The unweighted case corresponds to $c_d = \gamma_{d,j} = 1$.

As we shall see in Volume II, strong polynomial tractability of problems such as multivariate integration for the normalized error criterion holds if c_d is arbitrary and $\sum_{j=1}^{d} \gamma_{d,j}$ is uniformly bounded in d. For product weights independent of dimension this holds if $\sum_{j=1}^{\infty} \gamma_j < \infty$. For equal product weights $\gamma_{d,j} = \beta_d$, it is enough to guarantee that $d\beta_d$ is uniformly bounded in d which holds, for example, if $\beta_d = d^{-1}$. Hence, even if all variables are equally important, the problem will be strongly polynomially tractable if the groups of k variables are monitored with the weight $c_d d^{-k}$.

For product weights, the Hilbert space $H_{d,\gamma}$ is the tensor product space of the $H_{1,\gamma_{d,j}}$, and if H_1 has a reproducing kernel K_1 then $H_{d,\gamma}$ has the reproducing kernel $K_{d,\gamma}$ of product form given in the previous subsection.

We now turn to order-dependent weights introduced in [45]. We say that $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ is a sequence of *order-dependent weights* iff for all $d \in \mathbb{N}$ and for all $\mathfrak{u} \subseteq [d]$ we have

$$\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$$

for some arbitrary non-negative sequence $\{\Gamma_{d,j}\}_{j \in [d], d \in \mathbb{N}}$.

If the weights are order-dependent, then for any $k \in [d]$, each group of k variables is equally important and is monitored by the weight $\Gamma_{d,k}$. Equivalently, if we consider the decomposition of the function $f = \sum_{\mathfrak{u}} f_{\mathfrak{u}}$ from the weighted space $H_{d,\gamma}$ then all terms $f_{\mathfrak{u}}$ corresponding to the same cardinality of $\mathfrak{u}$ are equally important. For order-dependent weights, all 2^d weights are specified by $d + 1$ numbers. Note that the order-dependent weights are product weights iff $\Gamma_{d,k} = c_d a_d^k$ for some positive c_d and non-negative a_d. If all $\Gamma_{d,k} = 1$ then, as before, all groups of variables are equally important and we are back to the unweighted case.

As we shall see later, polynomial tractability of some multivariate problems for the class Λ^{all} requires that $\sum_{\mathfrak{u}} \gamma_{d,\mathfrak{u}}^{\tau}$ is polynomially bounded in d for some positive τ. For order-dependent weights, this is equivalent to requiring that

$$\sum_{k=0}^{d} \binom{d}{k} \Gamma_{d,k}^{\tau}$$

is polynomially bounded in d. For instance, this holds for $\Gamma_{d,k} = c_d$ iff $2^d c_d^{\tau}$ is polynomially bounded in d, i.e., when $c_d = \mathcal{O}(2^{-d/\beta} d^q)$ for some positive β and non-negative q.

We now turn to finite-order weights introduced in [45].   We believe that this type of weights is probably the most important one and captures the behavior of many computational multivariate problems for large d.

We say that $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ is a sequence of *finite-order weights* iff there exists an integer ω such that

$$\gamma_{d,\mathfrak{u}} = 0 \quad \text{for all } d \text{ and for all } \mathfrak{u} \text{ with } |\mathfrak{u}| > \omega. \tag{5.15}$$

The finite-order weights are of *order* ω^* if ω^* is the smallest integer with the property (5.15).

The essence of finite-order weights is that an arbitrary function $f = \sum_{\mathfrak{u}} f_{\mathfrak{u}}$ from the weighted space $H_{d,\gamma}$ is the sum of functions $f_{\mathfrak{u}}$, each depending on at most ω^* variables. This is a very powerful property that will enable us to obtain many polynomial tractability results for linear and some non-linear multivariate problems.

For finite-order weights, we have at most $\sum_{k=0}^{\omega^*} \binom{d}{k} = \mathcal{O}(d^{\omega^*})$ non-zero weights. In fact, it is easy to show, see also [273], that

$$\sum_{k=0}^{\omega^*} \binom{d}{k} \leq 2 d^{\omega^*}.$$

The inequality is obvious for $d = 1$ or for $\omega^* = 0$. For $d \geq 2$ and $\omega^* \geq 1$, we use induction on ω^*. For $\omega^* = 1$ it is again obvious, whereas for $\omega^* + 1$ we have

$$\sum_{k=0}^{\omega^*+1} \binom{d}{k} \leq 2 d^{\omega^*} + \binom{d}{\omega^* + 1} \leq 2 d^{\omega^*} + \frac{d^{\omega^*+1}}{(\omega^* + 1)!}$$

$$\leq d^{\omega^*+1} \left(\frac{2}{d} + \frac{1}{(\omega^* + 1)!} \right) \leq 2 d^{\omega^*+1},$$

as claimed. Obviously, for large d we have $\sum_{k=0}^{\omega^*} \binom{d}{k} = d^{\omega^*}(1 + o(1))/\omega^*!$.

We stress that the total number of non-zero finite-order weights ω^* is sometimes significantly less than d^{ω^*}. For example, consider functions often studied in physics which are given as a sum of Coulomb pair potentials,

$$f(x) = \sum_{1 \leq i < j \leq m} \frac{1}{\|\vec{x}_i - \vec{x}_j\|},$$

where $x = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m]$ with $\vec{x}_j \in \mathbb{R}^3$, and the Euclidean norm $\|\cdot\|$ of vectors, see for instance the book of Glimm and Jaffe [64]. In this case $d = 3m$. Since f is not defined for $\vec{x}_i = \vec{x}_j$, usually f is modified by taking a small positive η and by considering

$$f_\eta(x) = \sum_{1 \leq i < j \leq m} \frac{1}{\left(\|\vec{x}_i - \vec{x}_j\|^2 + \eta \right)^{1/2}}.$$

Hence, the terms of f and f_η depend on groups of two 3-dimensional vectors, i.e., each term depends on six variables. In this case we thus have $\omega^* = 6$ although the total number of non-zero terms is only of order d^2 not d^6.

Another subclass of finite-order weights with only a few non-zero weights was recently pointed to us by Creutzig [33]. He proposed to call them *finite-diameter weights*. That is, $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ is a sequence of *finite-diameter weights* if there exists an integer $q \geq 1$ such that

$$\gamma_{d,\mathfrak{u}} = 0 \quad \text{for all } d \text{ and for all } \mathfrak{u} \text{ with diam}(\mathfrak{u}) \geq q. \tag{5.16}$$

Here, the diameter of $\mathfrak{u}$ is $\text{diam}(\mathfrak{u}) := \max_{i,j \in \mathfrak{u}} |i - j|$, and we take $\text{diam}(\emptyset) = 0$ by convention.

The essence of finite-diameter weights is that $f = \sum_{\mathfrak{u}} f_{\mathfrak{u}}$ may have non-zero $f_{\mathfrak{u}}$ only for functions depending on at most q successive components. Examples of financial mathematics seem to correspond, at least approximately, to finite-diameter weights since they correspond to time dependent phenomena that are given as sums of terms depending mostly on what has occurred in the recent past.

The finite-diameter weights are of *order q^**, with $q^* \geq 1$, if q^* is the smallest integer with the property (5.16). It is easy to see that finite-diameter weights of order q^* are also finite-order weights of order q^*, since $|\mathfrak{u}| > q^*$ implies that $\text{diam}(\mathfrak{u}) \geq q^*$.

The important property of finite-diameter weights is that the total number of non-zero weights is only *linear* in d, independently of their order q^*. More precisely, for finite-diameter weights of order q^*, the total number of non-zero weights is at most 2^d for $d \leq q^*$ and

$$2^{q^*-1}d - (q^* - 2)2^{q^*-1} \quad \text{for } q^* < d.$$

The case $q^* < d$ is obviously more interesting.

This formula can be proven as follows. For $d \leq q^*$ we have $\text{diam}(\mathfrak{u}) \leq d-1 < q^*$. Since there is no restriction on non-zero weights $\gamma_{d,\mathfrak{u}}$, we have at most 2^d non-zero weights. For $q^* < d$, we count for $k = 0, 1, \ldots, q^* - 1$ how many weights $\gamma_{d,\mathfrak{u}}$ can be non-zero for $\text{diam}(\mathfrak{u}) = k$. For $k = 0$ we have clearly $d + 1$ such sets, $\{\emptyset\}, \{1\}, \ldots, \{d\}$. For $k > 0$, the smallest and largest elements of $\mathfrak{u}$ with $\text{diam}(\mathfrak{u}) = k$ are i and $i + k$ with $i \in \{1, 2, \ldots, d - k\}$. So we have $d - k$ choices for the smallest element. Note that all $\mathfrak{u} = \{i, i_2, i_3, \ldots, i_{k-1}, i+k\}$ with $\{i_2, i_3, \ldots, i_{k-1}\} \subseteq \{i+1, i+2, \ldots, i+k-1\}$ have diameter k and we have 2^{k-1} such $\mathfrak{u}$. Hence, the total number of $\mathfrak{u}$ with $\text{diam}(\mathfrak{u}) = k$ is $(d - k)2^{k-1}$, so that the total number of non-zero finite-diameter weights of order $q^* < d$ is

$$1 + d + \sum_{k=1}^{q^*-1} (d - k)2^{k-1} = 2^{q^*-1}(d - q^* + 2),$$

as claimed. Note that for the both cases, $d \leq q^*$ and $q^* < d$, the total number of non-zero finite-diameter weights is at most

$$2^{\min(q^*,d)-1}d - (\min(q^*, d) - 2)2^{\min(q^*,d)-1}.$$

From one point of view, positive polynomial tractability results for finite-order weights may seem to be *not* surprising, at least for linear multivariate problems S_d. Indeed, instead of approximating $S_d f$ one can switch, due to linearity of S_d, to approximation of $S_{\mathfrak{u}} f := S_d f_{\mathfrak{u}}$, which is an instance of at most an ω^*-dimensional problem. So instead of solving one d-dimensional linear problem, we have to solve at most $2d^\omega$ linear problems, each of which is at most ω^*-dimensional. So, if even the information complexity of each ω^*-dimensional problem is exponential in ω^* but polynomial in ε^{-1}, we will get polynomial tractability for S_d. There is, however, a problem with this reasoning. First of all, we may get information only about f, and the previous approach requires the information about each $f_{\mathfrak{u}}$. But if we use the class Λ^{all} as in this chapter we still can use information on all $f_{\mathfrak{u}}$ since any continuous linear functional on $f_{\mathfrak{u}}$ can be treated as a linear continuous functional on f. Hence, the previous reasoning is at least so far well founded for the class Λ^{all}. What then happens if we use function values, that is the class Λ^{std}? There is some hope also in this case, as shown in [126], which we will not pursue here.

There is also a potential problem related to scaling. Namely, the problems S_d and $S_{\mathfrak{u}}$ may have very different initial errors. So as long as we use the normalized error criterion we would need to solve all problems $S_{\mathfrak{u}}$ to within $\varepsilon \|S_d\| / (2d^{\omega^*})$. The division by $2d^{\omega^*}$ is indeed needed since the individual errors of approximating $S_{\mathfrak{u}} f_{\mathfrak{u}}$ may add up when we approximate $S_d f$. Hence, still assuming that we have a polynomial dependence of the information complexity for the problem $S_{\mathfrak{u}}$ under the absolute error criterion on ε^{-1}, say, $n(\varepsilon, S_{\mathfrak{u}}) = \mathcal{O}(\varepsilon^{-p_{\mathfrak{u}}})$ with an unspecified dependence on ω^* in the big $\mathcal{O}$ notation, we will obtain a total cost of the form

$$\sum_{\mathfrak{u}:|\mathfrak{u}|\le\omega^*} \mathcal{O}\left(d^\omega \varepsilon^{-1}\|S_d\|\right)^{p_{\mathfrak{u}}}.$$

Is this cost really polynomial in d? It is not clear how $\|S_d\|$ depends on d and that is why we cannot be sure that we have polynomial tractability for the normalized error criterion. Well, let us make one more concession and switch to the absolute error criterion. Then $\|S_d\|$ disappears from the last formula, and the total cost is

$$\mathcal{O}\left(d^{\omega^*(1+\max_{\mathfrak{u}:|\mathfrak{u}|\le\omega^* } p_{\mathfrak{u}})} \varepsilon^{-\max_{\mathfrak{u}:|\mathfrak{u}|\le\omega^* } p_{\mathfrak{u}}}\right).$$

This is polynomial in d and ε^{-1} but the degrees of these polynomials can be quite high, and, in particular, never less than ω^* for d.

There is one more point of criticism of this ad-hoc analysis. Namely, we did not consider the influence of finite-order weights. Obviously if all non-zero weights are large the problem becomes harder. On the other hand, if one of the non-zero weights is small, say $\gamma_{d,\mathfrak{u}}$, then the problem $S_{\mathfrak{u}}$ is less important than problems with larger weights.

In any case, it should be clear by now to the reader that, although there is something valid in this intuitive statement that finite-order weights should imply polynomial tractability, we need a rigorous analysis, especially when S_d is non-linear or when the class Λ^{std} is used or when the normalized error criterion is chosen. We shall see in

the course of this book that polynomial and sometimes strong polynomial tractability indeed hold for finite-order weights even for the class Λ^{std}, for the normalized error criterion and for linear and selected non-linear multivariate problems.

5.3.3 Weighted Operators

In Section 5.2 we considered linear tensor product operators $S_d \colon H_d \to G_d$. We now extend their definition to the weighted Hilbert spaces $H_{d,\gamma}$.

We define $S_{d,\gamma} \colon H_{d,\gamma} \to G_d$ as $S_{d,\gamma} f = S_d f$ which makes sense since $H_{d,\gamma} \subseteq H_d$. The information complexity of $S_{d,\gamma}$ for the class Λ^{all} depends on the eigenvalues of the operator

$$W_{d,\gamma} := S_{d,\gamma}^* S_{d,\gamma} \colon H_{d,\gamma} \to H_{d,\gamma}.$$

Although $S_{d,\gamma} = S_d$, we cannot claim that $W_{d,\gamma} = W_d$ since the inner product of $H_{d,\gamma}$ is, in general, different than the inner product of H_d, and therefore $S_{d,\gamma}^*$ may be different than S_d^*. We now find the eigenpairs of $W_{d,\gamma}$ in terms of the eigenpairs $(\lambda_{d,j}, e_{d,j})$ of W_d, the latter given in Section 5.2.

We assume that the functions η_j from the previous section are chosen as the eigenfunctions of the operator W_d, so that

$$\eta_j = e_j \quad \text{for all } j \in \mathbb{N}^d. \tag{5.17}$$

We comment on this assumption. As we shall see in a moment, this assumption will easily allow us to find all eigenpairs of $W_{d,\gamma}$ which, in turn, will allow us to find necessary and sufficient conditions on tractability. If this assumption is not satisfied then tractability analysis may be harder. We will encounter this problem for S_d being a linear functional for which only the class Λ^{std} makes sense. As we shall see a different proof technique is then needed.

Assuming (5.17) and remembering that $\{\eta_{j,\gamma} = \gamma_{d,\mathfrak{u}(j)}^{1/2} \eta_j\}_{j \in \mathbb{N}_\gamma^d}$ is an orthonormal basis of $H_{d,\gamma}$, we have

$$
\begin{aligned}
\langle \eta_{i,\gamma}, W_{d,\gamma} \eta_{j,\gamma} \rangle_{H_{d,\gamma}} &= \langle S_{d,\gamma} \eta_{i,\gamma}, S_{d,\gamma} \eta_{j,\gamma} \rangle_{G_d} = \langle S_d \eta_{i,\gamma}, S_d \eta_{j,\gamma} \rangle_{G_d} \\
&= \left(\gamma_{d,\mathfrak{u}(i)} \gamma_{d,\mathfrak{u}(j)} \right)^{1/2} \langle S_d \eta_i, S_d \eta_j \rangle_{G_d} \\
&= \left(\gamma_{d,\mathfrak{u}(i)} \gamma_{d,\mathfrak{u}(j)} \right)^{1/2} \langle \eta_i, W_d \eta_j \rangle_{H_d} = \gamma_{d,\mathfrak{u}(j)} \lambda_{d,j} \delta_{i,j}.
\end{aligned}
$$

This proves that $W_{d,\gamma} \eta_{j,\gamma}$ is orthogonal to all $\eta_{i,\gamma}$ with $i \neq j$. Hence $W_{d,\gamma} \eta_{j,\gamma}$ must be parallel to $\eta_{j,\gamma}$ and then

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

where

$$\lambda_{d,\gamma,j} = \gamma_{d,\mathfrak{u}(j)} \lambda_{d,j}.$$

Hence, the eigenfunctions of $W_{d,\gamma}$ are the same as W_d, however, they are differently normalized, whereas the eigenvalues of $W_{d,\gamma}$ are the eigenvalues of W_d multiplied by the weights $\gamma_{d,\mathfrak{u}(j)}$.

It is clear that the operator $W_{d,\gamma}$ is compact for any weights γ, since the $\lambda_{d,j}$ are products of the eigenvalues for the univariate problem and they tend to 0. Observe that

$$\lambda_{d,j} = \Big(\prod_{k \in \mathfrak{u}(j)} \lambda_{j_k} \Big) \lambda_1^{d-|\mathfrak{u}(j)|} \le \lambda_2^{|\mathfrak{u}(j)|} \lambda_1^{d-|\mathfrak{u}(j)|},$$

since the eigenvalues $\{\lambda_j\}$ are ordered. This proves that the initial error is

$$\|S_{d,\gamma}\|_{H_{d,\gamma} \to G_d} = \max_{j \in \mathbb{N}^d} \big(\gamma_{d,\mathfrak{u}(j)} \lambda_{d,j} \big)^{1/2} = \max_{\mathfrak{u} \subseteq [d]} \big(\gamma_{d,\mathfrak{u}} \lambda_2^{|\mathfrak{u}|} \lambda_1^{d-|\mathfrak{u}|} \big)^{1/2}.$$

We call $S_\gamma = \{S_{d,\gamma}\}$ a *linear weighted tensor product problem in the worst case setting*, or simply a *weighted* problem.

5.3.4 Tractability of Weighted Problems

We are ready to study polynomial and weak tractability of the weighted problem $S_\gamma = \{S_{d,\gamma}\}$ for the class Λ^{all}.

To keep the number of pages in this book relatively reasonable, we leave the case of tractability of S_γ for the absolute error criterion as an open problem for the reader.

Open Problem 27.

- Consider the weighted problem $S_\gamma = \{S_{d,\gamma}\}$ as defined in this section with $\lambda_2 > 0$. Find necessary and sufficient conditions for polynomial and weak tractability of S_γ for the absolute error criterion in the worst case setting and for the class Λ^{all}.

We now address tractability of S_γ for the normalized error criterion. Let $\{\lambda_{d,\gamma,j}\}$ be the ordered sequence of the eigenvalues of $W_{d,\gamma}$. We know that

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

$$= \Big\{ \gamma_{d,\mathfrak{u}(j)} \lambda_1^{d-|\mathfrak{u}(j)|} \lambda_2^{|\mathfrak{u}(j)|} \prod_{k \in \mathfrak{u}(j)} \frac{\lambda_{j_k}}{\lambda_2} \Big\}_{j \in \mathbb{N}_\gamma^d}.$$

The largest eigenvalue is given if we take $j = [j_1, j_2, \ldots, j_d]$ with $j_k \in \{1, 2\}$ such that $j_k = 2$ iff $k \in \mathfrak{u}$. Then $\mathfrak{u}(j) = \mathfrak{u}$ and

$$\lambda_{d,\gamma,1} = \max_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \lambda_1^{d-|\mathfrak{u}|} \lambda_2^{|\mathfrak{u}|} = \lambda_1^d \max_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \Big(\frac{\lambda_2}{\lambda_1} \Big)^{|\mathfrak{u}|}.$$

Note that for $\lambda_2 = 0$, we have $\lambda_{d,\gamma,1} = \gamma_{d,\emptyset}\lambda_1^d$ and $\lambda_{d,\gamma,j} = 0$ for all $j \geq 2$. Then $n(\varepsilon, d) \leq 1$ and the problem S_γ is trivial in the class Λ^{all}. Hence we assume $\lambda_2 > 0$ in what follows.

For the normalized error criterion, tractability depends on $\{\lambda_{d,\gamma,j}/\lambda_{d,\gamma,1}\}$. We first introduce the *normalized weights* $\gamma^* = \{\gamma_{d,\mathfrak{u}}^*\}_{\mathfrak{u}\subseteq[d],d\in\mathbb{N}}$ as

$$\gamma_{d,\mathfrak{u}}^* = \frac{\gamma_{d,\mathfrak{u}}(\lambda_2/\lambda_1)^{|\mathfrak{u}|}}{\max_{\mathfrak{v}\subseteq[d]} \gamma_{d,\mathfrak{v}}(\lambda_2/\lambda_1)^{|\mathfrak{v}|}},$$

and the *normalized eigenvalues* $\lambda^* = \{\lambda_j^*\}_{j\in\mathbb{N}}$ as

$$\lambda_j^* = \frac{\lambda_{j+1}}{\lambda_2}.$$

Clearly, $\gamma_{d,\mathfrak{u}}^*, \lambda_j^* \in [0, 1]$. Then

$$\left\{\frac{\lambda_{d,\gamma,j}}{\lambda_{d,\gamma,1}}\right\}_{j\in\mathbb{N}} = \left\{\gamma_{d,\mathfrak{u}(j)}^* \prod_{k\in\mathfrak{u}(j)} \lambda_{j_k-1}^*\right\}_{j\in\mathbb{N}_\gamma^d}. \tag{5.18}$$

Note that if γ is a sequence of order-dependent or finite-order weights, then γ^* is also a sequence of order-dependent or finite-order weights.

For product weights $\gamma_{d,\mathfrak{u}} = c_d \prod_{j\in\mathfrak{u}} \gamma_{d,j}$, we have

$$\gamma_{d,\mathfrak{u}}^* = c_d^* \prod_{j\in\mathfrak{u}} \gamma_{d,j}^* \quad \text{for } c_d^* = \frac{1}{\max_{\mathfrak{v}\subseteq[d]} \prod_{j\in\mathfrak{v}} \gamma_{d,\mathfrak{v}}^*}, \quad \gamma_{d,j}^* = \gamma_{d,j}(\lambda_2/\lambda_1),$$

and so γ^* is also a set of product weights. Note, however, that this point requires an arbitrary c_d. Indeed, assume that we adopt the definition of product weights with $c_d = 1$ always. Then $c_d^* = 1$ iff $\max_{\mathfrak{v}\subseteq[d]} \gamma_{d,\mathfrak{v}}(\lambda_2/\lambda_1)^{|\mathfrak{v}|} = 1$. We have

$$\max_{\mathfrak{v}\subseteq[d]} \gamma_{d,\mathfrak{v}}(\lambda_2/\lambda_1)^{|\mathfrak{v}|} = \max_{\mathfrak{v}} \prod_{j\in\mathfrak{v}} \gamma_{d,j}\lambda_2/\lambda_1 = \max_{k=0,k\in[d]} \prod_{j=1}^k \gamma_{d,j}\lambda_2/\lambda_1.$$

Since $\gamma_{d,\emptyset} = 1$, the last maximum is at least 1, and it is 1 iff $\gamma_{d,1}\lambda_2/\lambda_1 \leq 1$ for all d, which obviously does not have to hold in general. That explains why we decided to allow to have an arbitrary positive c_d in the definition of product weights.

We will be using the concept of the so-called *sum-exponent* p_ψ of a sequence $\psi = \{\psi_{d,k}\}_{k\in[d],d\in\mathbb{N}}$ for $\psi_{d,k} \in [0, 1]$ which was defined in [271] as

$$p_\psi = \inf \left\{\tau \geq 0 \mid \limsup_{d\to\infty} \sum_{k=1}^d \psi_{d,k}^\tau < \infty\right\}, \tag{5.19}$$

with the convention that $\inf \emptyset = \infty$.

Without loss of generality assume that $\{\psi_{d,k}\}$ are ordered, $\psi_{d,d} \leq \psi_{d,d-1} \leq \cdots \leq \psi_{d,1}$. Then $p_\psi < \infty$ iff $\psi_{d,k} = \mathcal{O}(k^{-\beta})$ for some positive β, and with the big $\mathcal{O}$ factor independent of d. Indeed, let $p_\psi < \infty$. Then

$$\limsup_d \sum_{k=1}^d \psi_{d,k}^s < \infty \quad \text{implies} \quad \limsup_d \sum_{k=1}^d \psi_{d,k}^\tau < \infty \quad \text{for any } \tau \geq s.$$

Hence, for $\tau > p_\psi$ there exists a positive C such that for all d and all $k \leq d$, we have

$$k\psi_{d,k}^\tau \leq \sum_{j=1}^d \psi_{d,j}^\tau \leq C.$$

This means that $\psi_{d,k} \leq C^\beta k^{-\beta}$ with $\beta = 1/\tau$, as claimed. On the other hand, if $\psi_{d,k} \leq C k^{-\beta}$ then clearly $p_\psi \leq 1/\beta$.

Hence, $p_\psi < \infty$ means that the numbers $\psi_{d,k}$ decay polynomially in k^{-1} and uniformly in d.

Now set $\psi_{d,k} = \lambda_k^*$, so that there is no dependence on d. Then

$$p_{\lambda^*} = \inf\left\{\tau \geq 0 \mid \sum_{k=1}^\infty \left[\lambda_k^*\right]^\tau < \infty\right\}.$$

We are ready to present a theorem characterizing polynomial tractability for the weighted case.

Theorem 5.7. *Consider the linear weighted tensor product problem in the worst case setting $S_\gamma = \{S_{d,\gamma}\}$ for compact linear $S_{d,\gamma}: H_{d,\gamma} \to G_d$ defined over Hilbert spaces $H_{d,\gamma}$ and G_d with $\lambda_2 > 0$. We assume that for each d, the weight $\gamma_{d,\mathfrak{u}} > 0$ for at least one non-empty $\mathfrak{u}$. We study the problem S_γ for the normalized error criterion and for the class Λ^{all}.*

- *S_γ is polynomially tractable iff $p_{\lambda^*} < \infty$ and there exist $q_2 \geq 0$ and $\tau > p_{\lambda^*}$ such that*

$$C_2 := \sup_d \left(\sum_{\mathfrak{u} \subseteq [d]} \left[\gamma_{d,\mathfrak{u}}^*\right]^\tau \left(\sum_{j=1}^\infty \left[\lambda_j^*\right]^\tau \right)^{|\mathfrak{u}|} \right)^{1/\tau} d^{-q_2} < \infty. \qquad (5.20)$$

- *If (5.20) holds then*

$$n(\varepsilon, d) \leq C_2^\tau d^{q_2\tau} \varepsilon^{-2\tau} \quad \text{for all } \varepsilon \in (0, 1] \text{ and } d = 1, 2, \ldots.$$

- *If S_γ is polynomially tractable, so that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some positive C and p with $q \geq 0$, then (5.20) holds with $q_2 = 2q/p$ and any τ such that $\tau > p/2$. Then*

$$C_2 \leq 2^{1/\tau}(C + 2)^{2/p} \zeta(2\tau/p)^{1/\tau}.$$

- S_γ is strongly polynomially tractable iff (5.20) holds with $q_2 = 0$. The exponent of strong polynomial tractability is

$$p^{\text{str-wor}} = \inf\{2\tau \mid \tau > p_{\lambda*} \text{ and satisfies (5.20) with } q_2 = 0\}.$$

- *For product weights* $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ *with* $\gamma_{d,j+1} \le \gamma_{d,j}$, *let* $\gamma_{d,j}^* = \gamma_{d,j} \lambda_2 / \lambda_1$.

  - *Let* $\lambda_3 > 0$. *Then* S_γ *is polynomially tractable iff* $p_{\lambda*} < \infty$ *and there exists* $\tau > p_{\lambda*}$ *such that*

    $$\limsup_{d \to \infty} \frac{\sum_{j=1}^d \min\left(1, [\gamma_{d,j}^*]^\tau\right)}{\ln d} < \infty. \tag{5.21}$$

    If (5.21) *holds then for any*

    $$q > \frac{1}{\tau} \left(\sum_{j=1}^\infty \left(\frac{\lambda_{j+1}}{\lambda_1}\right)^\tau \right) \limsup_{d \to \infty} \frac{\sum_{j=1}^d \min\left(1, [\gamma_{d,j}^*]^\tau\right)}{\ln d},$$

    there exists a positive C *such that*

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

  - *Let* $\lambda_3 = 0$. *Define* $\beta = \{\beta_{d,j}\}_{j \in [d], d \in \mathbb{N}}$ *as*

    $$\beta_{d,j} = \min\left([\gamma_{d,j}^*]^{-1}, \gamma_{d,j}^*\right) \in [0, 1].$$

    Then S_γ *is polynomially tractable iff there exists* $\tau > 0$ *such that*

    $$\limsup_{d \to \infty} \frac{\sum_{j=1}^d \beta_{d,j}^\tau}{\ln d} < \infty. \tag{5.22}$$

    If (5.22) *holds then for any*

    $$q > \frac{1}{\tau} \limsup_{d \to \infty} \frac{\sum_{j=1}^d \beta_{d,j}^\tau}{\ln d}$$

    there exists a positive C *such that*

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

  - *Let* $\lambda_3 > 0$. *Then* S_γ *is strongly polynomially tractable iff* $p_{\lambda*} < \infty$ *and* $p_{\gamma**} < \infty$, *where* $\gamma^{**} = \{\min(1, \gamma_{d,k}^*)\}$. *If this holds, the exponent of strong polynomial tractability is*

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

– *Let $\lambda_3 = 0$. Then S_γ is strongly polynomially tractable iff $p_\beta < \infty$. If this holds, the exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = 2p_\beta.$$

- *For order-dependent weights $\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$, we have*
  – *S_γ is polynomially tractable iff $p_{\lambda^*} < \infty$ and there exist $q_2 \geq 0$ and $\tau > p_{\lambda^*}$ such that*

$$C_2 := \sup_d \left(\sum_{k=0}^{d} \binom{d}{k} [\Gamma_{d,k}^*]^\tau \left(\sum_{j=1}^{\infty} [\lambda_j^*]^\tau \right)^k \right)^{1/\tau} d^{-q_2} < \infty, \quad (5.23)$$

  where

$$\Gamma_{d,k}^* = \frac{\Gamma_{d,k}(\lambda_2/\lambda_1)^k}{\max_{j=0,1,\ldots,d} \Gamma_{d,j}(\lambda_2/\lambda_1)^j}$$

  are the normalized order-dependent weights.
  – *If (5.23) holds then*

$$n(\varepsilon, d) \leq C_2 d^{q_2\tau} \varepsilon^{-2\tau} \quad \text{for all } \varepsilon \in (0, 1] \text{ and } d = 1, 2, \ldots.$$

  – *S_γ is strongly polynomially tractable iff $p_{\lambda^*} < \infty$ and (5.23) holds with $q_2 = 0$. Then the exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = \inf\{2\tau \mid \tau > p_{\lambda^*} \text{ and satisfies (5.23) with } q_2 = 0\}.$$

- *For finite-order weights $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > \omega^*$ with order ω^*, we have*
  – *S_γ is polynomially tractable iff $p_{\lambda^*} < \infty$. Then for any $\tau > p_{\lambda^*}$,*

$$
\begin{aligned}
n(\varepsilon, d) &\leq \left(\sum_{j=1}^{\infty} [\lambda_j^*]^\tau \right)^{\omega^*} \left| \{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \neq 0\} \right| \varepsilon^{-2\tau} \\
&\leq 2 \left(\sum_{j=1}^{\infty} [\lambda_j^*]^\tau \right)^{\omega^*} d^{\omega^*} \varepsilon^{-2\tau}.
\end{aligned}
\tag{5.24}
$$

  Obviously, for finite-diameter weights of order q^* with $q^* < d$ we may replace $2d^{\omega^*}$ in the last estimate by $2^{q^*-1}(d - q^* + 2)$.
  – *S_γ is strongly polynomially tractable iff $p_{\lambda^*} < \infty$ and there exists $\tau > p_{\lambda^*}$ such that*

$$\sup_d \left(\sum_{\mathfrak{u} \subseteq [d] |\mathfrak{u}| \leq \omega^*} [\gamma_{d,\mathfrak{u}}^*]^\tau \left(\sum_{j=1}^{\infty} \lambda_j^* \right)^{|\mathfrak{u}|} \right)^{1/\tau} < \infty. \tag{5.25}$$

  The exponent of strong tractability is

$$p^{\text{str-wor}} = \inf\{2\tau \mid \tau > p_{\lambda^*} \text{ and satisfies (5.25)}\}.$$

Proof. From Theorem 5.2 we know that polynomial tractability for the normalized error criteria requires that some powers of the eigenvalues are summable. In our case we have

$$M_{d,\tau} := \left(\sum_{j=1}^{\infty} \left(\frac{\lambda_{d,\gamma,j}}{\lambda_{d,\gamma,1}} \right)^{\tau} \right)^{1/\tau} = \left(\sum_{\mathfrak{u} \subseteq [d]} [\gamma^*_{d,\mathfrak{u}}]^{\tau} \left(\sum_{j=1}^{\infty} [\lambda^*_j]^{\tau} \right)^{|\mathfrak{u}|} \right)^{1/\tau}.$$

Since $\gamma_{d,\mathfrak{u}} > 0$ for at least one non-empty $\mathfrak{u}$, we conclude that $M_{d,\tau}$ is finite iff $p_{\lambda^*} < \infty$ and if we take $\tau > p_{\lambda^*}$ or $\tau = p_\lambda$ if the corresponding infimum in the definition of p_λ is attained. Then C_2 in (5.20) is the same as C_2 in (5.4) and the first four points of Theorem 5.7 coincide with the four points of Theorem 5.2 and are presented for completeness.

Assume now that we have product weights. We now have $\gamma^*_{d,j} = \gamma_{d,j} \lambda_2 / \lambda_1$, and $\gamma^* = \{\gamma_{d,\mathfrak{u}}\}$ is also the sequence of product weights $\gamma^*_{d,\mathfrak{u}} = c^*_d \prod_{j \in \mathfrak{u}} \gamma^*_{d,j}$, where

$$c^*_d = \frac{1}{\max_{\mathfrak{v} \subseteq [d]} \prod_{j \in \mathfrak{v}} \gamma^*_{d,j}}.$$

Let

$$r_{d,\gamma^*} = \left| \{ j \in [d] \mid \gamma^*_{d,j} \geq 1 \} \right|$$

be the number of the normalized weights $\gamma^*_{d,j}$ that are at least one. Then

$$c^*_d = \frac{1}{\prod_{j=1}^{r_{d,\gamma^*}} \gamma^*_{d,j}} = \frac{1}{\prod_{j=1}^{d} \max\left(1, \gamma^*_{d,j}\right)}.$$

Therefore

$$M^{\tau}_{d,\tau} = [c^*_d]^{\tau} \sum_{\mathfrak{u} \subseteq [d]} \prod_{k \in \mathfrak{u}} \left([\gamma^*_{d,k}]^{\tau} \left(\sum_{j=1}^{\infty} [\lambda^*_j]^{\tau} \right) \right)$$

$$= [c^*_d]^{\tau} \prod_{k=1}^{d} \left(1 + [\gamma^*_{d,k}]^{\tau} \sum_{j=1}^{\infty} [\lambda^*_j]^{\tau} \right)$$

$$= \prod_{k=1}^{d} \left(\frac{1 + [\gamma^*_{d,k}]^{\tau} \sum_{j=1}^{\infty} (\lambda_{j+1}/\lambda_2)^{\tau}}{(1, [\gamma^*_{d,k}]^{\tau})} \right).$$

Assume first that $\lambda_3 > 0$. Note that $p_\lambda = p_{\lambda^*}$, so that the last series is convergent for $\tau > p_{\lambda^*}$. Let

$$\alpha_\tau = \sum_{j=1}^{\infty} \left(\frac{\lambda_{j+1}}{\lambda_2} \right)^{\tau} \geq 1 + \left(\frac{\lambda_3}{\lambda_2} \right)^{\tau} > 1.$$

Then

$$M_{d,\tau} = \prod_{k=1}^{r_{d,\gamma^*}} \left(\frac{1}{[\gamma^*_{d,k}]^{\tau}} + \alpha_\tau \right)^{1/\tau} \prod_{k=r_{d,\gamma^*}+1}^{d} \left(1 + [\gamma^*_{d,k}]^{\tau} \alpha_\tau \right)^{1/\tau}.$$

Observe that

$$M_{d,\tau} \geq \alpha_\tau^{r_{d,\gamma*}/\tau}.$$

S_γ is polynomially tractable iff $\{M_{d,\tau} d^{-q_2}\}$ is uniformly bounded in d for some q_2. Since $\alpha_\tau > 1$ this implies that if S_γ is polynomially tractable then

$$r_{\gamma*} := \limsup_d \frac{r_{d,\gamma*}}{\ln d} < \infty.$$

If $r_{\gamma*} = \infty$ then $M_{d,\tau}$ goes to infinity faster than any power of d and therefore S_γ is *not* polynomially tractable. Note that $r_{\gamma*} = \infty$ implies that

$$\limsup_d \frac{\sum_{j=1}^d \min\left(1, [\gamma_{d,j}^*]^\tau\right)}{\ln d} = \infty.$$

Hence, polynomial tractability of S_γ implies that $r_{\gamma*} < \infty$, and if (5.21) holds then $r_{\gamma*} < \infty$. This means that we can assume without loss of generality that $r_{\gamma*}$ is finite or equivalently that $r_{d,\gamma*} = \mathcal{O}(\ln d)$.

For $r_{\gamma*}$ finite, the first product for $k \in [1, r_{d,\gamma*}]$ in the definition of $M_{d,\tau}$ is polynomially bounded. So to guarantee polynomial tractability of S_γ we must check that the second product for $k \in [r_{d,\gamma*} + 1, d]$ is also polynomially bounded. Note that

$$d^{-q_2} \prod_{k=r_{d,\gamma*}+1}^d \left(1 + [\gamma_{d,k}^*]^\tau \alpha_\tau\right)^{1/\tau}$$

$$= \exp\left(-\ln(d)\left[q_2 - \frac{1}{\tau \ln d} \sum_{k=r_{d,\gamma*}+1}^d \ln\left(1 + [\gamma_{d,k}^*]^\tau \alpha_\tau\right)\right]\right).$$

So it is uniformly bounded in d iff

$$q_2 > \frac{1}{\tau} \limsup_{d \to \infty} \frac{\sum_{k=r_{d,\gamma*}+1}^d \ln\left(1 + x_{d,k}\right)}{\ln d} \quad \text{for } x_{d,k} = [\gamma_{d,k}^*]^\tau \alpha_\tau.$$

We have $x_{d,k} \leq \alpha_\tau$. For $x \in [0, \alpha_\tau]$ it is easy to check that

$$\frac{1}{1 + \alpha_\tau} x \leq \ln(1 + x) \leq x. \tag{5.26}$$

Therefore

$$\overline{\lim}_d \frac{\sum_{k=r_{d,\gamma*}+1}^d \ln\left(1 + x_{d,k}\right)}{\ln d} < \infty \quad \text{iff} \quad \overline{\lim}_d \frac{\sum_{k=r_{d,\gamma*}+1}^d x_{d,k}}{\ln d} < \infty$$

$$\text{iff} \quad \overline{\lim}_d \frac{\sum_{k=r_{d,\gamma*}+1}^d [\gamma_{d,k}^*]^\tau}{\ln d} < \infty.$$

Finally notice that

$$\frac{\sum_{k=1}^{d} \min\left(1, \left[\gamma_{d,j}^*\right]^\tau\right)}{\ln d} = \frac{r_{d,\gamma^*}}{\ln d} + \frac{\sum_{k=r_{d,\gamma^*}+1}^{d} \left[\gamma_{d,k}^*\right]^\tau}{\ln d}.$$

That is why (5.21) is a necessary and sufficient condition for polynomial tractability of S_γ.

To obtain the estimate on the exponent q, note that $M_{d,\tau}$ can be also written as

$$M_{d,\tau} = \prod_{k=1}^{d} \left(\alpha_\tau \min\left(1, \left[\gamma_{d,k}^*\right]^\tau\right) + \min\left(1, \left[\gamma_{d,k}^*\right]^{-\tau}\right)\right)^{1/\tau}$$

$$\leq \prod_{k=1}^{d} \left(\alpha_\tau \min\left(1, \left[\gamma_{d,k}^*\right]^\tau\right) + 1\right)^{1/\tau}.$$

Then

$$M_{d,\tau} d^{-q} \leq \exp\left(-\ln(d)\left[q - \frac{\alpha_\tau}{\tau} \frac{\sum_{k=1}^{d} \min\left(1, \left[\gamma_{d,k}^*\right]^\tau\right)}{\ln d}\right]\right).$$

So it is uniformly bounded in d for q given in the theorem. The rest of this point is easy.

Assume now that $\lambda_3 = 0$. Then $\alpha_\tau = 1$ and $p_{\lambda^*} = 0$. Note that for any non-negative a we have

$$\frac{1+a}{\max(1, a)} = 1 + \min\left(a^{-1}, a\right).$$

Therefore we can now rewrite $M_{d,\tau}$ as

$$M_{d,\tau} = \prod_{k=1}^{d} \left(1 + \min\left(\left[\gamma_{d,k}^*\right]^{-\tau}, \left[\gamma_{d,k}^*\right]^\tau\right)\right) = \prod_{j=1}^{d} \left(1 + \beta_{d,j}^\tau\right).$$

The rest follows by the same reasoning as in the previous point.

We turn to strong polynomial tractability of S_γ for product weights. Then $\{M_{d,\tau}\}$ must be uniformly bounded in d. For $\lambda_3 > 0$, this holds iff $r_{d,\gamma^*} = \mathcal{O}(1)$ and $\left\{\sum_{k=r_{d,\gamma^*}+1}^{d} [\gamma_{d,k}^*]^\tau\right\}$ is uniformly bounded in d. This, in turn, holds iff $p_{\gamma^{**}}$ is finite. This means that strong polynomial tractability holds iff p_{λ^*} and p_γ are finite, as claimed. This and the fourth point of the theorem give the formula for the exponent of strong polynomial tractability.

For $\lambda_3 = 0$, the sequence $\{M_{d,\tau}\}$ is uniformly bounded in d iff $\sum_{j=1}^{d} \beta_{d,j}^\tau$ is uniformly bounded in d, which holds iff $p_\beta < \infty$. The rest follows easily.

We now switch to order-dependent weights. Note that C_2 in (5.23) is now the same as C_2 in (5.20), and this point follows from the first part of the theorem.

We are ready to address the last point of the theorem for finite-order weights. We already know that $p_{\lambda^*} < \infty$ is necessary for polynomial tractability of S_γ. For any

finite-order weights, consider C_2 given by (5.20) with $\tau > p_\lambda$ and $q_2 = \omega^*/\tau$. We have

$$C_2 = \sup_d \left(\sum_{\mathfrak{u}:\mathfrak{u}\subseteq[d],|\mathfrak{u}|\leq\omega^*} [\gamma_{d,\mathfrak{u}}^*]^\tau \left(\sum_{j=1}^\infty [\lambda_j^*]^\tau \right)^{|\mathfrak{u}|} \right)^{1/\tau} d^{-q_2}.$$

Since $\gamma_{d,\mathfrak{u}}^* \leq 1$, the cardinality of the sum with respect to $\mathfrak{u}$ is at most $2d^{\omega^*}$, and the sum with respect to j consists of the first term $[\lambda_1^*]^\tau = 1$ and therefore the powers of this sum are non-decreasing, we obtain

$$C_2 \leq \left(\sum_{j=1}^\infty [\lambda_j^*]^\tau \right)^{\omega^*/\tau} \left| \{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \neq 0\} \right|^{1/\tau} d^{-\omega^*/\tau} \leq 2^{1/\tau} \left(\sum_{j=1}^\infty [\lambda_j^*]^\tau \right)^{\omega^*/\tau}.$$

Then the estimate (5.24) on $n(\varepsilon, d)$ follows from the second point of the theorem. Strong polynomial tractability is the same as in the fourth point of Theorem 5.2 specified for finite-order weights. This completes the proof. □

We now comment on Theorem 5.7. The essence of this theorem is that we know necessary and sufficient conditions on polynomial tractability of the weighted problem S_γ. These conditions may be simplified for special weights. For product weights, polynomial tractability is equivalent to the normalized (univariate) eigenvalues having a finite sum-exponent and to the sum of a power of the normalized weights growing at most logarithmically in d. Strong polynomial tractability holds iff the sum-exponents of both the normalized eigenvalues and weights are finite, that is, the sums of their powers are uniformly bounded in d. The sum-exponent p_{λ^*} of the normalized eigenvalues measures the smoothness of the univariate problem; for $\lambda_n^* = \Theta(n^{-r})$ it is equal to $1/r$. The sum-exponent p_{γ^*} of the normalized weights measures the decay of the weights and, in general, has nothing in common with p_{λ^*}. We stress that the exponent of strong polynomial tractability $p^{\text{str-wor}} = 2\max(p_{\lambda^*}, p_\gamma)$ can be much larger than p_{λ^*}. For example, take $\gamma_{d,j} = j^{-\beta}$ for small positive β such that $\beta \ll r$, and assume that $\lambda_1 = \lambda_2$. Then $\gamma^* = \gamma$ and $p^{\text{str-wor}} = p_{\gamma^*} = 1/\beta$, which is much larger than $1/r$. Hence, we see again that smoothness has nothing to do with tractability, although it sets a lower bound on the exponent of strong polynomial tractability since $p^{\text{str-wor}} \geq 2p_{\lambda^*}$ for any normalized weights.

Observe that for the unweighted case, $\gamma_{d,\mathfrak{u}} = 1$, we have product weights and the condition (5.21) is *not* satisfied as well as $p_{\gamma^*} = \infty$. This means that polynomial tractability does not hold, and, in this case, Theorem 5.7 coincides with Theorem 5.6.

For general product weights, polynomial tractability does not necessarily imply strong polynomial tractability, see [271]. Indeed, take a problem with $p_{\lambda^*} < \infty$, and consider the weights $\gamma_{d,j} = 1$ for $j = 1, 2, \ldots, \lceil \ln d \rceil$, and $\gamma_{d,j} = 0$ for $j = \lceil \ln d \rceil + 1, \ldots, d$. Then (5.21) or (5.22) holds for every positive τ, so that we have polynomial tractability. On the other hand, $p_{\gamma^{**}} = \infty$ and strong tractability does not hold.

The situation changes if we consider product weights independent of d, i.e., $\gamma_{d,j} = \gamma_j$ and $\gamma_1 \geq \gamma_2 \geq \cdots$. Then strong polynomial tractability is *equivalent*

to polynomial tractability, see Theorem 2 in [271]. Indeed, it is enough to show that polynomial tractability implies strong polynomial tractability. We now have $\gamma^*_{d,j} = \gamma_j \lambda_2 / \lambda_1$, and (5.21) or (5.22) implies that $\lim_j \gamma_j = 0$. Therefore there exists a positive integer j_0 and a positive number C such that

$$\frac{j \gamma_j^\tau}{\ln j} \leq C \quad \text{for all } j \geq j_0,$$

with τ satisfying (5.21) or (5.22). Hence,

$$\gamma_j \leq C^{1/\tau} \left(\frac{\ln j}{j} \right)^{1/\tau} \quad \text{for all } j \geq j_0.$$

This yields that $p_{\gamma^{**}} = p_{\gamma^*} = p_\gamma \leq \tau < \infty$, and strong tractability holds.

We now turn to the case of finite-order weights. We stress that as long as the sum-exponent of the normalized eigenvalues is finite, S_γ is polynomially tractable for *arbitrary* finite-order weights. The only dependence on d is through the cardinality of the set of non-zero finite-order weights. In general, this cardinality is of order d^{ω^*}, which explains why the bound on the information complexity $n(\varepsilon, d)$ depends *exponentially* on the order ω^* of weights. Indeed, ω^* is the degree of d, as well as the degree of the factor depending on the normalized eigenvalues. As long as ω^* is relatively small, this exponential dependence causes no problem. On the other hand, if ω^* is large, the situation is changed; although $n(\varepsilon, d)$ depends polynomially on d we may be not quite satisfied with the bound. It is easy to see that this exponential dependence on ω^* is generally unavoidable. Indeed, consider the simple case in which $\gamma_{d,\mathfrak{u}} = 1$ for all $|\mathfrak{u}| = \omega^*$, and $\gamma_{d,\mathfrak{u}} = 0$ otherwise. For $d = 1$, assume that $\lambda_1 = \lambda_2 = 1$ and $\lambda_j = 0$ for $j \geq 3$. Then we have $\binom{d}{\omega^*}$ multivariate problems with ω^* variables, each having 2^{ω^*} positive eigenvalues all equal to one. This implies that

$$n(\varepsilon, d) = 2^{\omega^*} \binom{d}{\omega^*} = \Theta\big((2d)^{\omega^*}\big) \quad \text{for all } \varepsilon \in [0, 1),$$

as claimed.

Obviously, the situation is better for some finite-order weights. For example, if we take finite-diameter weights of order q^*, then the dependence on d in $n(\varepsilon, d)$ is at most linear. However, note that we may have an exponential dependence on the order q^*, since the number of non-zero finite-diameter weights may indeed depend exponentially on q^*.

We now compare Theorems 5.6 and 5.7. For $\lambda_1 = \lambda_2$ we know that the unweighted problem S is intractable, whereas the weighted problem S_γ can even be strongly polynomial tractable if both the sum-exponents p_{λ^*} and p_{γ^*} are finite. In fact, one can check that for $\gamma_{d,\mathfrak{u}} = 1$ and for $\lambda_1 = \lambda_2$ we have $\lambda_1^* = 1$ and therefore C_2 in (5.20) is at least $\sup_d 2^{d/\tau} d^{-q_2} = \infty$ for all $\tau > 0$, which agrees with Theorem 5.6. For $\lambda_2 < \lambda_1$, we know that the unweighted problem S is polynomially intractable, and again for properly decaying normalized weights its weighted counterpart S_γ can even be strongly polynomial tractable.

We summarize this discussion by stressing the following. The unweighted problem that can be intractable or weakly tractable may become polynomially tractable or strongly polynomial tractable for the weighted case.

We are ready to analyze weak tractability of the weighted problem S_γ. We first remind the reader that the last point of Theorem 5.6 says that the unweighted problem S is weakly tractable if $\lambda_2 < \lambda_1$ and $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$, whereas $\lambda_2 = \lambda_1$ implies intractability of S.

It would be natural to hope that the weights can only help and should preserve weak tractability as long as $\lambda_2 < \lambda_1$ and λ_n goes to 0 at least as fast as indicated; we might even hope to get weak tractability if $\lambda_2 = \lambda_1$ under some conditions on the weights. We now show that this intuition is correct modulo a proper normalization of the weights.

First we present an example of showing that unless the weights are properly normalized, we may even *loose* weak tractability. The point here is that although we start with a single largest eigenvalue λ_1, we select weights such that the largest normalized eigenvalue is at least double and the weighted problem is intractable. Indeed, assume for simplicity that $0 < \lambda_2 < \lambda_1$ and the remaining eigenvalues are 0, i.e., $\lambda_j = 0$ for $j \geq 3$. For the d-dimensional case we have 2^d positive eigenvalues. More precisely for $k = 0, 1, \ldots, d$ we have $\binom{d}{k}$ eigenvalues equal to $\lambda_1^d (\lambda_2/\lambda_1)^k$. Define the weights $\gamma_{d,\mathfrak{u}} = (\lambda_1/\lambda_2)^{|\mathfrak{u}|}$. Then $\gamma_{d,\mathfrak{u}}^* = 1$ and all 2^d eigenvalues for the weighted case are just one. Hence for the weighted problem S_γ, we have $n(\varepsilon, d) = 2^d$ for all $\varepsilon \in [0, 1)$, and indeed we lost weak tractability.

The reason we have lost weak tractability is that we use exponentially large weights. Indeed, for $\mathfrak{u} = [d]$ we defined $\gamma_{d,[d]} = (\lambda_1/\lambda_2)^d$ and we transformed the largest eigenvalue λ_1^d of multiplicity 1 for the unweighted case to the largest normalized eigenvalue 1 of multiplicity 2^d.

To prevent such an unnatural behavior we restrict ourselves to weights

$$\gamma_{d,\emptyset} = 1 \text{ and } \gamma_{d,\mathfrak{u}} \in [0, 1] \quad \text{for all non-empty } \mathfrak{u} \subseteq [d]. \tag{5.27}$$

We are ready to formulate a theorem on weak tractability of S_γ that covers the cases of arbitrary multiplicity of the largest eigenvalue λ_1.

Theorem 5.8. *Consider the linear weighted tensor product problem in the worst case setting $S_\gamma = \{S_{d,\gamma}\}$ for compact linear $S_{d,\gamma} \colon H_{d,\gamma} \to G_d$ defined over Hilbert spaces $H_{d,\gamma}$ and G_d with $\lambda_2 > 0$. We study the problem S_γ for the normalized error criterion and for the class Λ^{all}. The weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ satisfies (5.27). Let $n(\varepsilon, d) = n(\varepsilon, S_{d,\gamma})$ denote the information complexity of $S_{d,\gamma}$.*

- *Let λ_1 be of multiplicity one, i.e., $\lambda_2 < \lambda_1$.*

  If there is a non-zero weight $\gamma_{d,\mathfrak{u}}$ for a non-empty $\mathfrak{u}$, and S_γ is weakly tractable, then $\lambda_n = o((\ln n)^{-2})$ as $n \to \infty$.

  If $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$ then S_γ is weakly tractable for all weight sequences.

- *Let λ_1 be of multiplicity p with $p \geq 2$, i.e., $\lambda_{p+1} < \lambda_p = \lambda_{p-1} = \cdots = \lambda_1$, and $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$. Define*

$$m_p(\varepsilon, d) = \sum_{\mathfrak{u} \subseteq [d]:\ \gamma_{d,\mathfrak{u}} > \varepsilon^2} (p-1)^{|\mathfrak{u}|}.$$

Then

- *we have*

$$1 \leq \frac{n(\varepsilon, d)}{m_p(\varepsilon, d)} = \exp\left(o(\varepsilon^{-1} + d)\right) \quad as\ \varepsilon^{-1} + d \to \infty.$$

Hence, S_γ is weakly tractable iff

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln m_p(\varepsilon, d)}{\varepsilon^{-1} + d} = 0.$$

- *For product weights $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ with*

$$0 \leq \gamma_{d,d} \leq \gamma_{d,d-1} \leq \cdots \leq \gamma_{d,1} \leq 1,$$

define $k(\varepsilon, d, \gamma) = k$ to be the element $k \in [d]$ such that

$$\prod_{j=1}^{k} \gamma_{d,j} > \varepsilon^2 \quad and \quad \prod_{j=1}^{k+1} \gamma_{d,j} \leq \varepsilon^2.$$

If such a k does not exist, set $k(\varepsilon, d, \gamma) = d$.
Then S_γ is weakly tractable iff

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{k(\varepsilon, d, \gamma)}{\varepsilon^{-1} + d} = 0.$$

- *For order-dependent weights $\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$ with*

$$0 \leq \Gamma_{d,d} \leq \Gamma_{d,d-1} \leq \cdots \leq \Gamma_{d,1} \leq 1,$$

define $k(\varepsilon, d, \gamma) = k$ to be the element $k \in [1, d]$ such that

$$\Gamma_{d,k} > \varepsilon^2 \quad and \quad \Gamma_{d,k+1} \leq \varepsilon^2.$$

If such a k does not exist, set $k(\varepsilon, d, \gamma) = d$.
Then S_γ is weakly tractable iff

$$\lim_{\varepsilon^{-1}+d \to \infty} \ln\left(\sum_{j=0}^{k(\varepsilon,d,\gamma)} (p-1)^j \binom{d}{j}\right)(\varepsilon^{-1} + d)^{-1} = 0.$$

– *For finite-order weights $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > \omega^*$ with order ω^*, S_γ is always weakly tractable.*

Proof. For weights satisfying (5.27) we have $\gamma^*_{d,\mathfrak{u}} = \gamma_{d,\mathfrak{u}}(\lambda_2/\lambda_1)^{|\mathfrak{u}|}$. As before, denote $\lambda'_j = \lambda_j/\lambda_1$. Then the eigenvalues for the weighted case are

$$\left\{ \frac{\lambda_{d,\gamma,j}}{\lambda_{d,\gamma,1}} \right\} = \left\{ \gamma_{d,\mathfrak{u}(j)} \prod_{k \in \mathfrak{u}(j)} \lambda'_{j_k} \right\}.$$

Let $n^{\mathrm{unw}}(\varepsilon, k)$ denote the information complexity of the unweighted k-dimensional problem S_k, so that

$$n^{\mathrm{unw}}(\varepsilon, k) = \left| \left\{ j \in (\mathbb{N} \setminus \{1\})^k \mid \lambda'_{j_1} \lambda'_{j_2} \cdots \lambda'_{j_k} > \varepsilon^2 \right\} \right|.$$

Then

$$n(\varepsilon, d) = \sum_{\mathfrak{u} \subseteq [d]} n^{\mathrm{unw}}(\varepsilon \gamma_{d,\mathfrak{u}}^{-1/2}, |\mathfrak{u}|).$$

Observe that for $\gamma_{d,\mathfrak{u}} \leq \varepsilon^2$ we have $\varepsilon \gamma_{d,\mathfrak{u}}^{-1/2} \geq 1$ and $n^{\mathrm{unw}}(\varepsilon \gamma_{d,\mathfrak{u}}^{-1/2}, |\mathfrak{u}|) = 0$. Therefore we can rewrite the last equality as

$$n(\varepsilon, d) = \sum_{\mathfrak{u} \subseteq [d],\ \gamma_{d,\mathfrak{u}} > \varepsilon^2} n^{\mathrm{unw}}(\varepsilon \gamma_{d,\mathfrak{u}}^{-1/2}, |\mathfrak{u}|).$$

Assume first that $\lambda_2 < \lambda_1$. If $\gamma_{d,\mathfrak{u}}$ is non-zero for some $k := |\mathfrak{u}| > 0$ then weak tractability of S_γ implies that S_k is also weakly tractable, which can only happen if $\lambda_n = o((\ln n)^{-2})$, as claimed. On the other hand, we have

$$\gamma_{d,\mathfrak{u}(j)} \prod_{k \in \mathfrak{u}(j)} \lambda'_{j_k} = \gamma_{d,\mathfrak{u}(j)} \prod_{k=1}^d \lambda'_{j_k} \leq \prod_{k=1}^d \lambda'_{j_k}.$$

The eigenvalues $\left\{ \prod_{k=1}^d \lambda'_{j_k} \right\}_{j \in \mathbb{N}^d}$ correspond to the unweighted case for the normalized error criterion studied in Theorem 5.6. This shows that the weighted problem S_γ is not harder than the unweighted problem S for the normalized error criterion and that weak tractability of S implies weak tractability of S_γ as long as $\lambda_2 < \lambda_1$ and $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$. This completes the proof of the first point.

We switch to the more interesting case when the largest eigenvalue λ_1 has multiplicity p with $p \geq 2$. As we know from Theorem 5.6, this implies that the unweighted problem is intractable.

Observe that for $\gamma_{d,\mathfrak{u}} > \varepsilon^2$ we have $n^{\mathrm{unw}}(\varepsilon \gamma^{-1/2}, |\mathfrak{u}|) \geq (p-1)^{|\mathfrak{u}|}$, since there are $(p-1)^{|\mathfrak{u}|}$ eigenvalues equal to one. Therefore

$$n(\varepsilon, d) \geq m_p(\varepsilon, d).$$

We need to estimate $n^{\mathrm{unw}}(\varepsilon, k)$ from above. Note that we now have $j_k \in \mathbb{N} \setminus \{1\}$, and the multiplicity of the largest eigenvalue $\lambda_2' = 1$ is $p-1$, and that $\lambda_{p+1} < 1$. We use an approach similar to that in the proof of (5.11). For $\lambda_{p+1} > 0$, we define

$$a_k(\varepsilon) = \min \left(k, \left\lceil 2\frac{\ln \varepsilon^{-1}}{\ln \left[\lambda_{p+1}'\right]^{-1}} \right\rceil - 1 \right).$$

For $\lambda_{p+1} = 0$ we take $a_k(\varepsilon) = 0$. We claim that

$$n^{\mathrm{unw}}(\varepsilon, k) \leq (p-1)^{k-a_k(\varepsilon)} \binom{d}{a_k(\varepsilon)} n^{\mathrm{unw}}(\varepsilon, 1)^{a_k(\varepsilon)}.$$

Indeed this follows from the same proof as in (5.11), with the only difference being that now at least $k - a_k(\varepsilon)$ indices take integer values from the interval $[2, p]$ and at most $a_k(\varepsilon)$ from $[p+1, n^{\mathrm{unw}}(\varepsilon, 1)]$. This explains the extra factor $(p-1)^{k-a_k(\varepsilon)}$. We know that $n^{\mathrm{unw}}(\varepsilon, 1) = \exp(o(\varepsilon^{-1}/\ln \varepsilon^{-1}))$ since $\lambda_n = o((\ln n)^{-2}(\ln n \ln n)^{-2})$. We proved in (5.13) that

$$\frac{\ln \binom{k}{a_k(\varepsilon)} n^{\mathrm{unw}}(\varepsilon, 1)^{a_k(\varepsilon)}}{\varepsilon^{-1} + k} = o(1) \quad \text{as } \varepsilon^{-1} + k \to \infty.$$

This implies that

$$n(\varepsilon, d) = \exp\left(o(\varepsilon^{-1} + d)\right) \sum_{\mathfrak{u}:\gamma_{d,\mathfrak{u}} > \varepsilon^2} (p-1)^{|\mathfrak{u}|} = \exp\left(o(\varepsilon^{-1} + d)\right) m_p(\varepsilon, d),$$

as claimed. Obviously

$$\lim_{\varepsilon^{-1} + d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = \lim_{\varepsilon^{-1} + d \to \infty} \frac{\ln m_p(\varepsilon, d)}{\varepsilon^{-1} + d},$$

completing the proof of this point of the theorem.

Now consider product weights. Observe that $\gamma_{d,\mathfrak{u}} > \varepsilon^2$ with $j = |\mathfrak{u}|$ implies that $\gamma_{d,[j]} > \varepsilon^2$. Hence, $\gamma_{d,\mathfrak{u}} > \varepsilon^2$ only for $\mathfrak{u}$ being a subset of $[k]$, where $k = k(\varepsilon, d, \gamma)$. For $j = |\mathfrak{u}| \leq k$ we have $\binom{k}{j}$ such subsets, which yields that

$$m_p(\varepsilon, d) = \sum_{j=0}^{k(\varepsilon,d,\gamma)} (p-1)^j \binom{k}{j} = p^{k(\varepsilon,d,\gamma)},$$

and the result follows from the previous point.

Next, consider order-dependent weights. Then $\gamma_{d,\mathfrak{u}} > \varepsilon^2$ iff $|\mathfrak{u}| \leq k(\varepsilon, d, \gamma)$, which holds for all subsets of $[d]$ of cardinality at most $k(\varepsilon, d, \gamma)$. This yields

$$m_p(\varepsilon, d) = \sum_{j=0}^{k(\varepsilon,d,\gamma)} (p-1)^j \binom{d}{j}$$

and the rest is obvious.

Finally, for arbitrary finite-order weights we even have polynomial tractability, see Theorem 5.7, which obviously implies weak tractability. This completes the proof.   □

Clearly, the most important part of Theorem 5.8 is the case of a multiple largest eigenvalue λ_1, i.e., when $p \geq 2$. Assuming that λ_n decays faster than $(\ln n)^{-2}(\ln \ln n)^{-2}$, weak tractability holds iff

$$\ln m_p(\varepsilon, d) = o(\varepsilon^{-1} + d).$$

Observe that for the unweighted case $\gamma_{d,\mathfrak{u}} = 1$, we have $m_p(\varepsilon, d) = 2^d$ and $S_\gamma = S$ is not weakly tractable. In this case, Theorem 5.8 coincides with Theorem 5.6. However, we can have weak tractability of S_γ for equal and sufficiently small weights. Indeed, let $\gamma_{d,\gamma} = c_d$. Then $m_p(\varepsilon, d) = 0$ if $c_d \leq \varepsilon^2$, and $m_p(\varepsilon, d) = p^d$ if $c_d > \varepsilon^2$. S_γ is weakly tractable iff

$$\lim_{d \to \infty} d c_d^{1/2} = 0,$$

which holds iff $c_d = o(d^{-2})$. This holds independently of $p \geq 2$. Hence, if all groups of variables $x_\mathfrak{u}$ are monitored with a weight $o(d^{-2})$, then weak tractability holds.

For product weights, we must have $k(\varepsilon, d, \gamma) = o(\varepsilon^{-1} + d)$ to obtain weak tractability. We can double-check whether this condition holds when we have polynomial tractability. As we know, polynomial tractability for product weights holds if $\sum_{j=1}^{d} \gamma_{d,j}^\tau = \mathcal{O}(\ln d)$ for some positive τ. Then $\gamma_{d,j} = \mathcal{O}(j^{-1/\tau} \ln d)$. It is not difficult to check that

$$k(\varepsilon, d, \gamma) = \mathcal{O}\left(\max\left(\ln \varepsilon^{-1}, (\ln d)^\tau\right)\right),$$

which obviously satisfies $k(\varepsilon, d\gamma) = o(\varepsilon^{-1} + d)$.

We now choose product weights for which weak tractability holds but polynomial tractability does not. Take $\gamma_{d,j} = c \in (0, 1]$. For $c = 1$ all weights $\gamma_{d,\mathfrak{u}} = 1$, and we have intractability, whereas for $c \in (0, 1)$ we do not have polynomial tractability, since $\sum_{j=1}^{d} \gamma_{d,j}^\tau = c^\tau d$ goes faster than $\ln d$ to infinity. However, we have weak tractability since

$$k(\varepsilon, d, \gamma) = \min\left(d, \lceil 2\ln(\varepsilon^{-1})/\ln(c^{-1})\rceil - 1\right).$$

Then for $x = \max(d, \varepsilon^{-1})$ we have $k(\varepsilon, d, \gamma) = \mathcal{O}(\ln x)$, as in the proof of Theorem 5.5, from which weak tractability easily follows.

One can also construct product weights that go to 0 and for which weak tractability holds but polynomial tractability does not. Let

$$\gamma_{d,j} = [\ln(e - 1 + j)]^{-\beta}$$

for a positive β. Note that $\gamma_{d,j} \leq \gamma_{d,1} = 1$. Obviously polynomial tractability for such product weights does not hold, since $\sum_{j=1}^{d} \gamma_{d,j}^\tau$ grows faster than $\ln d$. On the

other hand, let $k^* = k^*(\varepsilon) = \lceil \ln \varepsilon^{-1} \rceil$. Then

$$\prod_{j=1}^{2k^*} \gamma_{d,j} \leq \prod_{j=k^*+1}^{2k^*} \gamma_{d,j} \leq \left(\ln(k^*)\right)^{-\beta k^*} \leq \varepsilon^{\beta \lceil \ln \varepsilon^{-1} \rceil} \leq \varepsilon^2$$

for small ε. This shows that $k(\varepsilon, d, \gamma) \leq 2\lceil \ln \varepsilon^{-1} \rceil$, from which weak tractability holds.

We now discuss order-dependent weights. Let $\Gamma_{d,j} = j^{-\beta}$. It is easy to check that S_γ is weakly tractable iff $\beta > 2$. Indeed, we now have

$$k = k(\varepsilon, d, \gamma) = \Theta(\min(d, \varepsilon^{-2/\beta}))$$

and

$$\sum_{j=1}^{k} (p-1)^j \binom{d}{j} = \Theta\left(\min\left(p^d, [(p-1)d]^{\varepsilon^{-2/\beta}}\right)\right),$$

which implies weak tractability iff $\beta > 2$.

Obviously, the result for finite-order weights is not surprising in view of the fact that S_γ is polynomially tractable for all finite-order weights. As an illustration, one can check that $m_p(\varepsilon, d) = \mathcal{O}(d^{\omega^*}(p-1)^{\omega^*})$ for finite-order weights, which also yields weak tractability.

Example: Approximation for Weighted Korobov Space

We illustrate the results of this section for the multivariate approximation problem for the weighted Korobov space $H_{d,\gamma}^{\mathrm{Kor}} = H_{d,\alpha,\gamma}$ defined in Appendix A. That is, we have $S_d = \mathrm{APP}_d : H_{d,\gamma}^{\mathrm{Kor}} \to L_2([0,1]^d)$ with $\mathrm{APP}_d f = f$.

Proceeding as we did in the example of approximation for Korobov spaces, it is easy to check that the operator $W_d = \mathrm{APP}_d^* \mathrm{APP}_d : H_{d,\alpha,\gamma} \to H_{d,\alpha,\gamma}$ has the eigenpairs $\{\lambda_{h,\gamma}, e_{h,\gamma}\}_{h \in \mathbb{Z}_\gamma^d}$ with $\lambda_{h,\gamma} = \varrho_{d,\alpha,\gamma}(h)^{-1}$, $e_{h,\gamma}$ given in Appendix A, and

$$\mathbb{Z}_\gamma^d = \{h \in \mathbb{Z}^d \mid \gamma_{d,\mathfrak{u}_h} > 0\} \quad \text{for } \mathfrak{u}_h = \{k \mid h_k \neq 0\}.$$

We have

$$\{\lambda_{h,\gamma}\}_{h \in \mathbb{Z}_\gamma^d} = \left\{\gamma_{d,\mathfrak{u}_h} \beta_1^{d-|\mathfrak{u}_h|} \beta_2^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} |h_j|^{-2\alpha}\right\}_{h \in \mathbb{Z}_\gamma^d}.$$

Note that each eigenvalue does not depend on signs of the components of the vector $h \in \mathbb{Z}^d$. If some component of h is 0 then obviously the sign does not matter and therefore the eigenvalue

$$\gamma_{d,\mathfrak{u}_h} \beta_1^{d-|\mathfrak{u}_h|} \beta_2^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} |h_j|^{-2\alpha}$$

has multiplicity at least $2^{|\mathfrak{u}_h|}$.

Let $\{\lambda_{d,\gamma,j}\}_{j\in\mathbb{N}}$ be the ordered sequence of $\{\lambda_{h,\gamma}\}_{h\in\mathbb{Z}_\gamma^d}$. That is, $\{\lambda_{d,\gamma,j}\}_{j\in\mathbb{N}} = \{\lambda_{h,\gamma}\}_{h\in\mathbb{Z}^d}$ and $\lambda_{d,\gamma,j} \geq \lambda_{d,\gamma,j+1}$ for $j \in \mathbb{N}$. Clearly,

$$\lambda_{d,\gamma,1} = \beta_1^d \max_{\mathfrak{u}\subseteq[d]} \gamma_{d,\mathfrak{u}} \left(\frac{\beta_2}{\beta_1}\right)^{|\mathfrak{u}|}.$$

For $\alpha = 0$, all eigenvalues are of the form $\beta_1^d \gamma_{d,\mathfrak{u}}(\beta_2/\beta_1)^{|\mathfrak{u}|}$, and such eigenvalues have infinite multiplicity if $\gamma_{d,\mathfrak{u}} > 0$ for $|\mathfrak{u}| > 0$. If we assume that $\gamma_{d,\mathfrak{u}} > 0$ for some non-empty $\mathfrak{u}$ then the information complexity $n(\varepsilon, d) = \infty$ for $\varepsilon^2 < \beta_1^d \gamma_{d,\mathfrak{u}}(\beta_2/\beta_1)^{|\mathfrak{u}|}$. This implies that the problem is intractable for both the absolute and normalized error criteria.

Assume that $\alpha > 0$. As we have done in this section, consider only the normalized error criterion. Note that we cannot yet apply the results of this section to our problem since the eigenvalues $\{\lambda_{h,\gamma}\}$ are defined over $h \in \mathbb{Z}_\gamma^d$ whereas the theorems of this section require them to be indexed over $j \in \mathbb{N}^d$. Obviously, it is possible to reduce the case $h \in \mathbb{Z}_\gamma^d$ to the case $j \in \mathbb{N}^d$, and we now show how to do this. Define

$$\gamma_{d,\mathfrak{u}}^* = \frac{\gamma_{d,\mathfrak{u}}(\beta_2/\beta_1)^{|\mathfrak{u}|}}{\max_{\mathfrak{v}\subseteq[d]} \gamma_{d,\mathfrak{v}}(\beta_2/\beta_1)^{|\mathfrak{v}|}},$$

and the sequence of eigenvalues

$$\lambda_{2k-1} = \lambda_{2k} = k^{-2\alpha} \quad \text{for } k = 1, 2, \dots.$$

Note that $\lambda_1 = \lambda_2 = 1$ and therefore $\lambda_j^* = \lambda_{j+1}/\lambda_2 = \lambda_{j+1}$. We can also equivalently write $\lambda_j = 1/\lceil j/2 \rceil^{2\alpha}$ for all $j \in \mathbb{N}$.

We now have

$$\left\{\frac{\lambda_{d,\gamma,j}}{\lambda_{d,\gamma,1}}\right\}_{j\in\mathbb{N}} = \left\{\gamma_{d,\mathfrak{u}_h}^* \prod_{j\in\mathfrak{u}_h} |h_j|^{-2\alpha}\right\}_{h\in\mathbb{Z}^d}.$$

We will show that

$$\left\{\gamma_{d,\mathfrak{u}_h}^* \prod_{j\in\mathfrak{u}_h} |h_j|^{-2\alpha}\right\}_{h\in\mathbb{Z}^d} = \left\{\gamma_{d,\mathfrak{u}(j)}^* \prod_{k\in\mathfrak{u}(j)} \lambda_{j_k-1}^*\right\}_{j\in\mathbb{N}^d},$$

so that (5.18) holds. Here, as always, $\mathfrak{u}(j) = \{k \mid j_k \geq 2\}$. This enables us to apply the results of this section.

Observe that each eigenvalue in both sequences parameterized by $h \in \mathbb{Z}^d$ and $j \in \mathbb{N}^d$ has multiplicity $2^{|\mathfrak{u}_h|}$ and $2^{|\mathfrak{u}(j)|}$, respectively. This has already been explained for $h \in \mathbb{Z}^d$, whereas for $j \in \mathbb{N}^d$ it follows from the fact that each eigenvalue λ_j has multiplicity 2. More precisely, let $\delta_k \in \{0, 1\}$ for $k \in \mathfrak{u}(j)$. Then for the vector j' with component $j_k' = j_k - \delta_k$ for $k \in \mathfrak{u}(j)$ with even j_k, and $j_k = j_k + \delta_k$ for $k \in \mathfrak{u}(j)$ with odd j_k, and $j_k = 1$ for $k \notin \mathfrak{u}(j)$, we obtain the same eigenvalue. Hence, we have $2^{|\mathfrak{u}(j)|}$ such eigenvalues, as claimed.

We now prove the equality of the sequences parameterized by $h \in \mathbb{Z}^d$ and by $j \in \mathbb{N}^d$. Take an arbitrary $h \in \mathbb{Z}^d$ and consider the eigenvalue $\gamma_{d,\mathfrak{u}_h}^* \prod_{k \in \mathfrak{u}_h} |h_k|^{-2\alpha}$ of multiplicity $2^{|\mathfrak{u}_h|}$. Define $j_k = 2|h_k|$ for $k \in \mathfrak{u}_h$ and $j_k = 1$ for $k \notin \mathfrak{u}_h$. Then $\mathfrak{u}(j) = \mathfrak{u}_h$ and $\lambda_{j_k} = \lambda_{j_k-1}^* = |h_k|^{-2\alpha}$. Therefore

$$\gamma_{d,\mathfrak{u}_h}^* \prod_{k \in \mathfrak{u}_h} |h_k|^{-2\alpha} = \gamma_{d,\mathfrak{u}(j)}^* \prod_{k \in \mathfrak{u}(j)} \lambda_{j_k-1}^*.$$

Observe that the same equality is obtained for $\lambda_{j_k-1} = \lambda_{j_k}^*$. Hence, for all $k \in \mathfrak{u}_h$, we have two eigenvalues in the sequence parameterized by $j \in \mathbb{Z}^d$. Varying k, we have $2^{|\mathfrak{u}_h|}$ such eigenvalues. Hence, we preserve the multiplicity of the original eigenvalue.

The proof of the reverse inclusion is similar. For $\gamma_{d,\mathfrak{u}(j)}^* \prod_{k \in \mathfrak{u}(j)} \lambda_{j_k-1}^*$ we define $h_k = \pm 2j_k$ if $k \in \mathfrak{u}(j)$ and $h_k = 0$ for $k \notin \mathfrak{u}(j)$. Then $\mathfrak{u}_h = \mathfrak{u}(j)$ and clearly

$$\gamma_{d,\mathfrak{u}(j)}^* \prod_{k \in \mathfrak{u}(j)} \lambda_{j_k-1}^* = \gamma_{d,\mathfrak{u}_h}^* \prod_{k \in \mathfrak{u}_h} \lceil j_k/2 \rceil^{-2\alpha} = \gamma_{d,\mathfrak{u}_h}^* \prod_{k \in \mathfrak{u}_h} |h_k|^{-2\alpha}.$$

Furthermore, the last equalities are achieved for $2^{|\mathfrak{u}(j)|}$ vectors h. This completes the proof.

We are ready to apply Theorems 5.7 and 5.8 for the weighted approximation problem $\mathrm{APP}_\gamma = \{\mathrm{APP}_d\}$. For polynomial tractability we assume, as in Theorem 5.7, that for each d, at least one weight $\gamma_{d,\mathfrak{u}} > 0$ for $|\mathfrak{u}| > 0$. It is easy to check that we have $p_{\lambda^*} = (2\alpha)^{-1}$. Hence APP_γ is polynomially tractable iff there exist $q_2 > 0$ and $\tau > (2\alpha)^{-1}$ such that

$$C_2 = \sup_d \left(\sum_{\mathfrak{u} \subseteq [d]} \left[\gamma_{d,\mathfrak{u}}^* \right]^\tau (2\zeta(2\tau\alpha))^{|\mathfrak{u}|} \right)^{1/\tau} d^{-q_2} < \infty. \tag{5.28}$$

If this holds then

$$n(\varepsilon, d) \le C_2^\tau d^{q_2\tau} \varepsilon^{-2\tau}.$$

Strong polynomial tractability holds iff we can take $q_2 = 0$ in the definition of C_2 above.

Consider now product weights with $\gamma_{d,\emptyset} = 1$, $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ and $\gamma_{d,j} \in [0,1]$. Assume for simplicity that $\beta_2 \le \beta_1$. Then $\max_{\mathfrak{v} \subseteq [d]} \gamma_{d,\mathfrak{v}} (\beta_2/\beta_1)^{|\mathfrak{v}|} = 1$, and

$$\gamma_{d,\mathfrak{u}}^* = \prod_{j \in \mathfrak{u}} \gamma_{d,j} \frac{\beta_2}{\beta_1}.$$

Then APP_γ is polynomially tractable iff there exists $\tau > (2\alpha)^{-1}$ such that

$$A := \left(\frac{\beta_2}{\beta_1} \right)^\tau \limsup_d \frac{\sum_{j=1}^d \gamma_{d,j}^\tau}{\ln d} < \infty.$$

If this holds, then for any

$$q > \frac{(2\zeta(2\tau\alpha) - 1)A}{\tau}$$

there exists a positive C such that

$$n(\varepsilon, d) \le C d^q \varepsilon^{-2\tau}.$$

Furthermore, APP_γ is strongly tractable for product weights iff $p_\gamma < \infty$. If this holds, then the exponent of strong polynomial tractability is $2\max((2\alpha)^{-1}, p_\gamma)$.

For finite-order weights, APP_γ is always polynomially tractable. More precisely, for any $\tau > (2\alpha)^{-1}$ we have

$$n(\varepsilon, d) \le 2\,(2\zeta(2\tau\alpha) - 1)\,d^{\omega^*}\varepsilon^{-2\tau},$$

where ω^* is the order of the finite-order weights.

For finite-order weights, strong polynomial tractability holds iff there exists $\tau > (2\alpha)^{-1}$ such that

$$\sup_d \Big(\sum_{\mathfrak{u}\subseteq[d],\,|\mathfrak{u}|\le\omega^*} [\gamma_{d,\mathfrak{u}}^*]^\tau [2\zeta(2\tau\alpha) - 1]^{|\mathfrak{u}|} \Big)^{1/\tau} < \infty.$$

If this holds, then the exponent of strong polynomial tractability is 2τ with the smallest τ satisfying the last condition. Note that for $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} \in [0, 1]$, we have $\gamma_{d,\mathfrak{u}}^* = \gamma_{d,\mathfrak{u}}(\beta_2/\beta_1)^{|\mathfrak{u}|}$.

We now turn to weak tractability for weights satisfying (5.27). As we know, even the unweighted APP_γ is weakly tractable iff $\beta_2 < \beta_1$. Hence, we now consider the case $\beta_2 = \beta_1$ and apply Theorem 5.8. Since the largest eigenvalue for $d = 1$ has now multiplicity 3, we conclude that APP_γ is weakly tractable iff

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{m_3(\varepsilon, d)}{\varepsilon^{-1} + d} = 0, \tag{5.29}$$

where

$$m_3(\varepsilon, d) = \sum_{\mathfrak{u}\subseteq[d]:\,\gamma_{d,\mathfrak{u}}>\varepsilon^2} 2^{|\mathfrak{u}|}.$$

The last condition can be simplified for specific weights such as product weights as in Theorem 5.8.

Example: Trade-offs of the Exponents (Continued)

We slightly generalize the eigenvalues considered so far for this example. Let $g: \mathbb{N} \to \mathbb{N}$ and $g(d) \le d$ for all $d \in \mathbb{N}$. Consider the following eigenvalues

$$\lambda_{d,j} = \prod_{k=1}^{g(d)} j_k^{-\alpha} \quad \text{for all } j \in \mathbb{N}^d$$

for some positive α. For $g(d) = \min(d, \lceil \ln(d+1) \rceil^s)$ we obtain the eigenvalues previously studied.

Note that this sequence corresponds to product weights for which $\gamma_{d,j} = 1$ for $j = 1, 2, \ldots, g(d)$ and $\gamma_{d,j} = 0$ for $j = g(d) + 1, g(d) + 2, \ldots, d$. The largest eigenvalue is 1, and the absolute and normalized error criteria coincide. Furthermore, the multiplicity of the largest eigenvalue is 1.

From Theorem 5.6 we know that weak tractability holds for any function g. We want to check when strong polynomial and polynomial tractability hold. From Theorem 5.7 it is easy to check that strong polynomial tractability holds iff

$$\limsup_{d \to \infty} g(d) < \infty.$$

If the last condition is satisfied then only finitely many dimensions are considered and therefore the problem is strongly polynomially tractable. It is quite natural to consider the opposite case when

$$\limsup_{d \to \infty} g(d) = \infty,$$

and turn to polynomial tractability. Then (5.21) states that polynomial tractability holds iff

$$A_g := \limsup_{d \to \infty} \frac{g(d)}{\ln d} < \infty.$$

Hence $A_g < \infty$ iff $g(d)$ is at most a multiple of $\ln d$. If $A_g < \infty$ then for all

$$\tau > 1/\alpha \quad \text{and} \quad q > A_g \ln \zeta(\alpha \tau)$$

we have

$$n(\varepsilon, d) = \mathcal{O}\left(d^q \varepsilon^{-2\tau}\right),$$

with the factors in the big $\mathcal{O}$ notation depending only on τ and q. Again, if A_g is positive then we have a trade-off between the exponent of d and ε^{-1}. In particular, the exponent of d can be arbitrarily small if we take sufficiently large τ at the expense of the exponent of ε^{-1} which then becomes arbitrarily large.

Example: Schrödinger Equation (Continued)

We already considered the linear Schrödinger equation for $\tilde{H}_d$ which is a subset of the Korobov space $H_{d,\alpha}$ for $\alpha \geq 2$. We briefly comment on what happens if we replace the unweighted space $\tilde{H}_d$ by a weighted space $\tilde{H}_{d,\gamma}$ repeating the process of replacing $H_{d,\alpha}$ by $H_{d,\alpha,\gamma}$.

It is easy to see that $\tilde{H}_d$ is also a subset of $H_{d,\alpha,\gamma}$ for $\alpha \geq 2$ and $\gamma_{d,[d]} > 0$. Indeed, $\eta_{d,2j} \in H_{d,\alpha,\gamma}$ and just now

$$\|\eta_{d,2j}\|_{H_{d,\alpha}} = \left(\frac{2^d \varrho_{d,\alpha}(2j)}{\gamma_{d,[d]}}\right)^{1/2} = 2^{d/2} \varrho_{d,\alpha,\gamma}^{1/2}(2j).$$

Hence,

$$\tilde{\eta}_{d,2j,\gamma} = 2^{-d/2}\varrho_{d,\alpha,\gamma}^{-1/2}(2j)\eta_{d,2j} \quad \text{for all } j \in \mathbb{N}^d$$

are orthonormal in $H_{d,\alpha,\gamma}$ and orthogonal in L_2.

This suggests that we should define $\tilde{H}_{d,\gamma}$ as $\tilde{H}_d$ with $\tilde{\eta}_{d,2j}$ replaced by $\tilde{\eta}_{d,2j,\gamma}$. Observe that again for $\alpha \geq 2$, we have $\tilde{H}_{d,\gamma} \subseteq H_d$ so that the linear Schrödinger problem for $\tilde{H}_{d,\gamma}$ is well defined. The operator W_d of the approximation problem over $\tilde{H}_{d,\gamma}$ has the eigenvalues

$$\lambda_{d,\gamma,j} = \gamma_{d,[d]} \prod_{k=1}^{d} \frac{\beta_2}{2^{2\alpha}j_k^{2\alpha}} \quad \text{for all } j \in \mathbb{N}^d.$$

Obviously, for the normalized error criterion the presence of positive $\gamma_{d,[d]}$ is irrelevant and therefore we have the same tractability conditions as for the unweighted case. For the normalized error criterion we thus have:

- the linear Schrödinger problem for $\tilde{H}_{d,\gamma}$ is weakly tractable for all weights γ for which $\gamma_{d,[d]} > 0$,

- the linear Schrödinger problem for $\tilde{H}_{d,\gamma}$ is polynomially intractable for all weights γ for which $\gamma_{d,[d]} > 0$.

5.4 Other Ways of Obtaining Linear Weighted Problems

The purpose of this section is to show that the construction of linear weighted problems presented in the previous section does not cover all interesting cases. As we shall see, we may have other natural linear weighted problems defined over standard spaces for which the tractability analysis of the previous sections cannot be applied. More importantly, the tractability results may be quite different.

The construction of linear weighted problems in the previous section was based on tensor products. As we know, tractability depends on the behavior of the eigenvalues of the operator W_d. The construction was done in such a way that the eigenvalues of the operator W_d were given by weighted products of univariate eigenvalues. Furthermore, the weight $\gamma_{d,\emptyset}$ was related to a 1-dimensional subspace. As we shall see in this section, this property is important. If it is violated, then we may even have intractability, no matter how other weights are defined.

In this section, we return to the problem that we discussed at the beginning of the book, namely, the problem of the relationship between smoothness and tractability. Intuitively, one might expect that more smoothness makes the problem easier. This is definitely true if we consider the asymptotic error estimates which go to 0 faster for smoother problems. How about tractability? Does smoothness also help to obtain at least weak tractability? It turns out that it is *not* the case in general. We provide

examples for which smoothness *hurts* in terms of tractability. More specifically, this holds for multivariate weighted approximation problems defined on variants of the Sobolev spaces of smooth functions, which are *intractable* for all choices of weights as long as the smoothness parameter $m \geq 2$, see [280]. Hence, even if we take all zero weights except $\gamma_{d,\emptyset} = 1$, we have the curse of dimensionality. The reason is that the weight $\gamma_{d,\emptyset}$ will now correspond to a subspace of dimension m^d, which is exponential in d iff $m \geq 2$. Hence, the only case for which we do not have this exponential explosion of dimension is the smallest smoothness $m = 1$. In this case, the situation is quite different. We have weak tractability for practically all weights, and for some weights we may even have strong polynomial tractability. Hence, the problem is tractable only if we have the smallest smoothness $m = 1$.

One of the major results of the previous section was that many multivariate problems are polynomially tractable for *any* finite-order weights. In this section we present a multivariate weighted problem that is still polynomially tractable for *bounded* finite-order weights, but may be *intractable* for unbounded finite-order weights. This result is made possible by defining the weighted problem differently than before and by showing that the eigenvalues of the operator W_d are no longer given by weighted products. Instead they are given by weighted *sums* of univariate eigenvalues. In this case we need a different analysis which shows polynomial tractability for bounded finite-order weights, and intractability for some unbounded finite-order weights.

These intriguing properties will be presented in the next section for a specific example of multivariate approximation defined on a weighted Sobolev space of smooth functions. We then generalize this example and study general linear weighted multivariate problems defined differently than in the previous section. We present several tractability results that are different from those obtained in the previous sections.

5.4.1 Weighted Sobolev Space of Smooth Functions

This section is based on [280]. We first define the Sobolev space

$$H_{d,\gamma} = H_{d,m,\gamma}$$

with $m \geq 1$ being the smoothness parameter. For $m = 1$, the space $H_{d,1,\gamma}$ is defined in Section A.2.1 of Appendix A, and was studied by Thomas-Agnan [236]. For all $m \geq 1$, it will be instructive to start with $d = 1$. For $\gamma > 0$, the space $H_{1,m,\gamma}$ consists of real functions defined on $[0, 1]$ whose $(m - 1)$st derivatives are absolutely continuous and whose mth derivatives belong to $L_2([0, 1])$, with the inner product

$$\langle f, g \rangle_1 = \int_0^1 f(x)g(x)\,\mathrm{d}x \; + \; \gamma^{-1} \int_0^1 f^{(m)}(x)g^{(m)}(x)\,\mathrm{d}x$$

for all $f, g \in H_{1,m,\gamma}$.

For $d \geq 2$ and a product weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ with $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ and positive $\gamma_{d,j}$, the space $H_{d,m,\gamma}$ is the d-fold tensor product of $H_{1,m,\gamma_{d,j}}$ and the inner

product is

$$\langle f, g \rangle_{H_{d,m,\gamma}} = \sum_{\mathfrak{u} \subseteq [d]} \frac{1}{\prod_{j \in \mathfrak{u}} \gamma_{d,j}} \int_{[0,1]^d} \frac{\partial^{m|\mathfrak{u}|} f}{\partial^m x_{\mathfrak{u}}}(x) \frac{\partial^{m|\mathfrak{u}|} g}{\partial^m x_{\mathfrak{u}}}(x) \, dx$$

for all $f, g \in H_{d,m,\gamma}$. That is, for each term we differentiate functions m times with respect to all variables in $\mathfrak{u}$, and then integrate over all d variables. For $\mathfrak{u} = \emptyset$ we have $\gamma_{d,\emptyset} = 1$, and the integrand is simply $f(x)g(x)$.

For general weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, we replace $\prod_{j \in \mathfrak{u}} \gamma_{d,j}$ by $\gamma_{d,\mathfrak{u}}$, and obtain the inner product

$$\langle f, g \rangle_{H_{d,m,\gamma}} = \sum_{\mathfrak{u} \subseteq [d]} \frac{1}{\gamma_{d,\mathfrak{u}}} \int_{[0,1]^d} \frac{\partial^{m|\mathfrak{u}|} f}{\partial^m x_{\mathfrak{u}}}(x) \frac{\partial^{m|\mathfrak{u}|} g}{\partial^m x_{\mathfrak{u}}}(x) \, dx$$

for all $f, g \in H_{d,m,\gamma}$. We assume that $\gamma_{d,\emptyset} = 1$ for all $d \in \mathbb{N}$.

If $\gamma_{d,\mathfrak{u}} = 0$ for some $\mathfrak{u} \subseteq [d]$ then we assume that

$$\frac{\partial^{m|\mathfrak{u}|} f}{\partial^m x_{\mathfrak{u}}} = 0,$$

and interpret $0/0 = 0$. Then obviously

$$\frac{\partial^{m|\mathfrak{v}|} f}{\partial^m x_{\mathfrak{v}}} = 0 \quad \text{for all } \mathfrak{v} \text{ such that } \mathfrak{u} \subseteq \mathfrak{v},$$

and we can take $\gamma_{d,\mathfrak{v}} = 0$.

We define weighted multivariate approximation $APP_\gamma = \{APP_{d,\gamma}\}$, where

$$APP_{d,\gamma} \colon H_{d,m,\gamma} \to L_2 := L_2([0,1]^d)$$

with $APP_{d,\gamma} f = f$. We consider this problem in the worst case setting for the class Λ^{all}.

We first note that the weighted problem APP_γ is *not* of the form studied in Section 5.3. Indeed, take the zero weights $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u} \neq \emptyset$. Then the space $H_{d,m,\gamma}$ is the space of polynomials of degree at most $m - 1$ in each variable, and it has dimension m^d. For $m \geq 2$, the weight $\gamma_{d,\emptyset} = 1$ corresponds to this space, and therefore does not correspond to a 1-dimensional space as in Section 5.3. We prove that the weighted problem APP_γ is intractable in this case. This should be contrasted with its counterpart problems in Section 5.3 which are all trivial in this case. Indeed, these earlier problems are all 1-dimensional, and hence they can be solved exactly using one information operation. As we shall see in a moment, for $m = 1$ and general weights $\gamma_{d,\mathfrak{u}}$, the weighted problem APP_γ is still different from the weighted problems of Section 5.3.

We now analyze APP_γ. It is easy to see that $\|f\|_{L_2} \leq \|f\|_{H_{d,m,\gamma}}$ with equality for $f = 1$. This implies that the initial error is 1, and the absolute and normalized error criteria coincide. It is well known that the nth minimal error $e^{\text{wor}}(n; m, \gamma)$ satisfies

$$e^{\text{wor}}(n; m, \gamma) = o\left(n^{-r}\right) \quad \text{for all } r < m \text{ as } n \to \infty.$$

Furthermore, if at least one of the weights $\gamma_{d,\mathfrak{u}}$ is positive for non-empty $\mathfrak{u}$, the last estimate is sharp, i.e., r cannot be larger than m. This shows how smoothness increases the speed of convergence, and implies that the information complexity

$$n(\varepsilon, d; m, \gamma) = o\left(\varepsilon^{-p}\right) \quad \text{for all } p > 1/m \text{ as } n \to \infty.$$

We now turn to tractability and prove the following theorem.

Theorem 5.9. *We have*

$$n(\varepsilon, d; m, \gamma) \geq m^d$$

for arbitrary weights $\gamma_{d,\mathfrak{u}}$ with $|\mathfrak{u}| > 0$ and for arbitrary $\varepsilon \in (0, 1)$.

Hence, multivariate approximation APP_γ suffers from the curse of dimensionality and is intractable for $m \geq 2$.

Proof. We could prove this theorem based on results already established. Instead, we supply a direct and short proof, which basically was also already used in Chapter 3 for some specific multivariate problems. Let

$$X = \{f \mid f \text{ is a polynomial of degree at most } m - 1 \text{ for each variable}\}.$$

Clearly, $\dim(X) = m^d$, $X \subseteq H_{d,m,\gamma}$, and

$$\|f\|_{H_{d,m,\gamma}} = \|f\|_{L_2} \quad \text{for all } f \in X.$$

We stress that the last equality holds for arbitrary weights $\gamma_{d,\mathfrak{u}}$, including the zero weights $\gamma_{d,\mathfrak{u}} = 0$ for all $|\mathfrak{u}| > 0$. For the zero weights, we obviously have $X = H_{d,m,\gamma}$. For general weights, X is an m^d-dimensional subspace of $H_{d,m,\gamma}$, for which the norms in spaces $H_{d,m,\gamma}$ and L_2 are the same.

Suppose that $n < m^d$. Consider an arbitrary algorithm

$$A_{n,d}(f) = \varphi(L_1(f), L_2(f), \ldots, L_n(f)),$$

where $L_j = L_j(\cdot; L_1(f), \ldots, L_{j-1}(f))$ are linear adaptive functionals, $L_j \in H_{d,m,\gamma}^*$. Choose now a non-zero $g \in X$ such that

$$L_1(g) = 0, \; L_2(g; 0) = 0, \ldots, L_n(g; 0, \ldots, 0) = 0.$$

Such a function g exists since we have a system of n homogeneous linear equations with at least $n + 1$ unknowns. We can normalize the function g by taking $\|g\|_{H_{d,m,\gamma}} = 1$. Let $a = \varphi(0, \ldots, 0) \in L_2$. Then

$$e^{\mathrm{wor}}(A_{n,d}) \geq \max_{c \in \{-1,1\}} \|cg - a\|_{L_2}$$

$$\geq \tfrac{1}{2} \left(\|g - a\|_{L_2} + \|g + a\|_{L_2}\right) \geq \|g\|_{L_2} = 1.$$

Since $A_{n,d}$ is arbitrary, we conclude that the nth minimal error $e^{\mathrm{wor}}(n) \geq 1$ for $n < m^d$. Hence, to obtain an algorithm with worst case error $\varepsilon < 1$, we must take at least m^d information operations. This means that $n(\varepsilon, d; m, \gamma) \geq m^d$, as claimed. $\qquad\square$

The last result shows that we cannot possibly have any kind of tractability unless $m = 1$. We now analyze this case. We need to find the eigenpairs of the operator $W_d = \mathrm{APP}_d^* \mathrm{APP}_d : H_{d,1,\gamma} \to H_{d,1,\gamma}$. We know from Section A.2.1 of Appendix A that $\{e_k\}_{k \in \mathbb{N}^d}$ given by (A.7) is an orthogonal basis of $H_{d,1,\gamma}$. Observe that

$$\langle f, g \rangle_{L_2} = \langle \mathrm{APP}_d f, \mathrm{APP}_d g \rangle_{L_2} = \langle f, W_d g \rangle_{H_{d,1,\gamma}} \quad \text{for all } f, g \in H_{d,1,\gamma}.$$

Taking $f = e_j$ and $g = e_k$ for arbitrary $j, k \in \mathbb{N}^d$ and remembering that $\{e_k\}$ is also orthogonal in L_2, we conclude that $W_d e_k$ is orthogonal to e_j for all $j \neq k$. Hence, e_k is an eigenfunction of W_d and its eigenvalue $\lambda_{d,k}$ is given by $\lambda_{d,k} = \langle e_k, e_k \rangle_{L_2} / \langle e_k, e_k \rangle_{H_{d,1,\gamma}}$. From (A.10) we thus have

$$\lambda_{d,k} = \frac{1}{1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1) \right]^2}.$$

Hence, for general weights, the eigenvalues $\lambda_{d,\mathfrak{u}}$ are *not* weighted products of the univariate eigenvalues, unlike the case in the previous section. This proves that APP_γ differs from the linear weighted problems studied in Section 5.3.

For the information complexity $n(\varepsilon, d) := n(\varepsilon, d; 1, \gamma)$ we have

$$n(\varepsilon, d) = |\{k \mid \lambda_{d,k} > \varepsilon^2\}| = |\{k \mid \lambda_{d,k}^{-1} < \varepsilon^{-2}\}|.$$

Hence,

$$n(\varepsilon, d) = \left| \left\{ k \in \mathbb{N}^d \mid 1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1) \right]^2 < \varepsilon^{-2} \right\} \right|. \quad (5.30)$$

We now consider tractability for several specific choices of the weights.

- **Equal weights $\gamma_{d,\mathfrak{u}} = 1$ for all $\mathfrak{u} \subseteq [d]$.** In this case, $H_{d,1,\gamma}$ is a tensor product space and the eigenvalues $\lambda_{d,k}$ are of product form

$$\lambda_{d,k} = \prod_{j=1}^{d} \frac{1}{1 + \left[\pi(k_j - 1) \right]^2} \quad \text{for all } k \in \mathbb{N}^d.$$

For these weights, the eigenvalues are products of the eigenvalues of the univariate case, as in the previous sections. Therefore we can use the tractability results already established.

Note that for the univariate case we have $\lambda_1 = 1$ and $\lambda_2 = 1/(1 + \pi^2)$, so that the largest eigenvalue λ_1 is simple. Obviously, $\lambda_n = \Theta(n^{-2})$, and Theorem 5.6 implies that we have weak tractability, but not polynomial tractability. Still even in this case of equal weights, we break the curse of dimensionality present for $m \geq 2$ and obtain weak tractability. Much more can be said about $n(\varepsilon, d)$. In particular, we may use the results of Chapter 8 to conclude that for any positive numbers α_1 and α_2 satisfying $\alpha_1 \alpha_2 \geq 1$, there exist two positive numbers C and t such that

$$n(\varepsilon, d) \leq C \exp \left(t \left[\ln^{1+\alpha_1}(1 + \varepsilon^{-1}) + \ln^{1+\alpha_2}(1 + d) \right] \right)$$

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

- **Product weights $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$.** In this case, $H_{d,1,\gamma}$ is also a tensor product space and the eigenvalues $\lambda_{d,k}$ are of product form

$$\lambda_{d,k} = \prod_{j=1}^{d} \frac{\gamma_{d,j}}{\gamma_{d,j} + \left[\pi(k_j - 1)\right]^2} \quad \text{for all } k \in \mathbb{N}^d.$$

Hence, this case is also covered by the previous analysis. It is easy to check that

$$\sup_{d \in \mathbb{N}} \max_{j \in [d]} \gamma_{d,j} < \infty$$

implies weak tractability. In general, the last inequality is needed. Namely, there exists a sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ with $\sup_{d \in \mathbb{N}} \max_{j \in [d]} \gamma_{d,j} = \infty$ for which APP_γ is intractable. One such example is given by the case $\gamma_{d,j} = d$. Indeed, the second largest eigenvalue for the d-dimensional case is $1/(1 + \pi^2/d)$, and if we take

$$\varepsilon = \varepsilon_d = \frac{1}{2} \frac{1}{(1 + \pi^2/d)^{d/2}} = \tfrac{1}{2} \exp(-\pi^2/2)(1 + o(1)),$$

then

$$n(\varepsilon_d, d) \geq 2^d.$$

So we have the curse of dimensionality and intractability, as claimed.

From our general results, it is easy to check strong polynomial tractability and polynomial tractability of APP_γ for bounded product weights. Namely, we have strong polynomial tractability iff there is a positive number τ for which

$$\sup_{d \in \mathbb{N}} \sum_{j=1}^{d} \min\left(1, \gamma_{d,j}^\tau\right) < \infty.$$

Furthermore, the exponent of strong polynomial tractability is

$$p = \max(1, 2\tau^*),$$

where τ^* is the infimum of τ satisfying the last inequality. In particular, for $\gamma_{d,j} = j^{-\beta}$ we have $\tau^* = 1/\beta$, and for $\beta \geq 2$ the exponent of strong tractability is 1, just as for the univariate case. So in this case, we have strong polynomial tractability for $m = 1$ with the smallest possible exponent.

Polynomial tractability holds iff there is a positive number τ for which

$$\limsup_{d \to \infty} \frac{\sum_{j=1}^{d} \min\left(1, \gamma_{d,j}^\tau\right)}{\ln d} < \infty.$$

Obviously, for bounded product weights, we may replace $\min(1, \gamma_{d,j}^\tau)$ simply by $\gamma_{d,j}^\tau$.

Finite-order weights

In this subsection we consider multivariate approximation APP_γ for the space $H_{d,1,\gamma}$, i.e., $m = 1$, equipped with finite-order weights. This case requires a different proof technique than before since the eigenvalues $\lambda_{d,k}$ have a different form. We now establish polynomial tractactability for *bounded* finite-order weights.

Theorem 5.10. APP_γ *is polynomially tractable for bounded finite-order weights, i.e.,*

$$\gamma_{d,\mathfrak{u}} = 0 \quad \text{for all } \mathfrak{u} \text{ with } |\mathfrak{u}| > \omega^*,$$
$$M := \sup_d \max_{\mathfrak{u} \subseteq [d], |\mathfrak{u}| \leq \omega^*} \gamma_{d,\mathfrak{u}} < \infty.$$

In this case, for every $\tau > 1$ there exists a positive number C_{τ,ω^*} independent of M, d and ε such that

$$n(\varepsilon, d) \leq C_{\tau,\omega^*} M^{\tau/2} d^{\omega^*} \varepsilon^{-\tau} \quad \text{for all } \varepsilon \in (0, 1], \ d \in \mathbb{N}.$$

Proof. It is enough to consider $d > \omega^*$. We have

$$\frac{1}{\lambda_{d,k}} = 1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2 \quad \text{for all } k \in \mathbb{N}^d, \tag{5.31}$$

with the convention that for $\gamma_{d,\mathfrak{u}} = 0$ there exists $k_j = 1$ for $j \in \mathfrak{u}$, and we interpret $0/0 = 0$.

We now show that the sum in (5.31) consists of at most $2^{\omega^*} - 1$ positive terms. Indeed, we have $W_d e_k = \lambda_{d,k} e_k$. Let

$$\mathfrak{u}_k = \{j \in [d] \mid k_j > 1\} = \{k_{j_1}, k_{j_2}, \ldots, k_{j_s}\},$$

where $s = |\mathfrak{u}_k|$. Note that if $|\mathfrak{u}_k| > \omega^*$ then $\gamma_{d,\mathfrak{u}_k} = 0$, so that the norm $\|e_k\|_{H_{d,1,\gamma}}$ given in Appendix A would be infinite. Hence, k may have at most ω^* components greater than 1. Equivalently, at least $d - \omega^*$ components of k are 1. So we cannot have a positive term in the sum (5.31) unless $\mathfrak{u} \subseteq \mathfrak{u}_k$. Since $|\mathfrak{u}_k| \leq \omega^*$, the number of its non-empty subsets is at most $2^{\omega^*} - 1$. Hence we can have at most $2^{\omega^*} - 1$ terms in the sum (5.31), as claimed. Furthermore, the number of such sets $\mathfrak{u}_k$ is at most

$$\sum_{j=1}^{\omega^*} \binom{d}{j} = \Theta(d^{\omega^*}).$$

We can rewrite the expression for $\lambda_{d,k}^{-1}$ as

$$\frac{1}{\lambda_{d,k}} = 1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq \mathfrak{u}_k} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2.$$

Since $\gamma_{d,\mathfrak{u}} \leq M$, we have

$$\frac{1}{\lambda_{d,k}} \geq 1 + \frac{1}{M} \sum_{\emptyset \neq \mathfrak{u} \subseteq \mathfrak{u}_k} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2$$

$$= 1 + \frac{1}{M}\left[\prod_{i=1}^{s}(1 + \pi^2(k_{j_i} - 1)^2) - 1\right]$$

$$\geq 1 + \frac{1}{M}\left[\pi^{2s}\prod_{i=1}^{s}(k_{j_i} - 1)^2 - 1\right].$$

Since $\lambda_{d,k}^{-1} < \varepsilon^{-2}$, this implies that

$$\prod_{i=1}^{s}(k_{j_i} - 1)^2 < \frac{M(\varepsilon^{-2} - 1) + 1}{\pi^{2s}},$$

or (equivalently) that

$$\sum_{i=1}^{s}\ln(k_{j_j} - 1) < x := \ln\left(\frac{M(\varepsilon^{-2} - 1) + 1}{\pi^{2s}}\right)^{1/2}.$$

Let $\ell = [\ell_1, \ell_2, \ldots, \ell_s]$ with $\ell_i = k_{j_i} - 1 \geq 1$. Due to (5.30), we conclude that

$$n(\varepsilon, d) = \mathcal{O}\left(d^{\omega^*}\left|\{\ell \mid \sum_{i=1}^{s}\ln \ell_i < x\}\right|\right).$$

From (8.20) we can estimate the cardinality of the last set. We find that for every number $\tau > 1$ there exists a number $C(\tau, s)$ such that

$$n(\varepsilon, d) = \mathcal{O}\left(d^{\omega^*}C(\tau, s)\exp(\tau x)\right) = \mathcal{O}\left(d^{\omega^*}M^{\tau/2}\varepsilon^{-\tau}\right).$$

The factor in the big $\mathcal{O}$ notation is independent of M, d and ε, and may only depend on τ and s, or since $s \leq \omega^*$ it may depend on τ and ω^*. This completes the proof. $\qquad\square$

We now show that the assumption that finite-order weights are bounded is essential, and APP$_\gamma$ may be intractable for unbounded finite-order weights. This can happen even for a very simple case, in which we have only two non-zero weights. Namely, take $\gamma_{d,\emptyset} = 1$, $\gamma_{d,\{1\}} = 2^d$, with $\gamma_{d,\mathfrak{u}} = 0$ for the remaining $\mathfrak{u}$. We now have $\omega^* = 1$ but $M = \infty$. Then APP$_\gamma$ reduces to the univariate approximation problem and

$$n(\varepsilon, d) = |\{\ell \in \mathbb{N} \mid \ell^2 < 2^d(\pi\varepsilon^{-1})^2\}| = \pi 2^{d/2}\varepsilon^{-1}(1 + o(1)).$$

Clearly, we have the curse of dimensionality and the problem is intractable.

Remark 5.11. We stress that the results presented in this section that large smoothness $m \geq 2$ implies the curse of dimensionality and that the smallest smoothness $m = 1$ may lead even to strong tractability, are the consequence of the choice of norm in the Sobolev space $H_{d,m,\gamma}$. For some other choices of the norm, increasing smoothness may help. Indeed, let us switch to a more standard norm which for $d = 1$ is of the form

$$\|f\|^2 = \int_0^1 \left[f(x)\right]^2 \mathrm{d}x + \int_0^1 \left[f'(x)\right]^2 \mathrm{d}x + \cdots + \int_0^1 \left[f^{(m)}(x)\right]^2 \mathrm{d}x$$

and corresponds to the standard Sobolev space $H^m([0,1])$. Then for $d > 1$ we take the d-fold tensor product of $H^m([0,1])$ and obtain a space $H^m([0,1]^d)$.

Let $e(n,m,d)$ denote the nth minimal worst case error of L_2-multivariate approximation for the space $H^m([0,1])^d$. Then the initial errors are the same, i.e,

$$e(0,m,d) = 1 \quad \text{for all } m, d \in \mathbb{N}.$$

Clearly, the problem for $m+1$ is easier than the problem for m since we have

$$e(n, m+1, d) \leq e(n,m,d) \quad \text{for all } n, m, d \in \mathbb{N},$$
$$e(n,m,d) = \mathcal{O}\left(n^{-m+\delta}\right) \quad \text{for all } \delta > 0.$$

It is easy to check that the largest eigenvalue of W_1 has multiplicity one, and therefore Theorem 5.6 implies that L_2-multivariate approximation is weakly tractable, but not polynomially tractable, for any m.

The tractability of L_2-multivariate approximation for the weighted version of $H^m([0,1]^d)$ has not yet been studied. There are many different ways to introduce weighted spaces $H_\gamma^m([0,1]^d)$. One way is to define the norm for $d = 1$ as

$$\|f\|^2 = \int_0^1 \left[f(x)\right]^2 \mathrm{d}x + \frac{1}{\gamma} \left(\int_0^1 \left[f'(x)\right]^2 \mathrm{d}x + \cdots + \int_0^1 \left[f^{(m)}(x)\right]^2 \mathrm{d}x\right),$$

so that only one term remains unweighted. Let us denote this space by $H_\gamma^m([0,1])$. For $d \geq 1$, we define

$$H_\gamma^m([0,1]^d) = H_{\gamma_{d,1}}^m([0,1]) \otimes H_{\gamma_{d,2}}^m([0,1]) \otimes \cdots \otimes H_{\gamma_{d,d}}^m([0,1]).$$

In this way we have the weighted space $H_\gamma^m([0,1]^d)$ with product weights. Similarly as we did for the weighted space $H_{d,1,\gamma}$, one can obtain the weighted space $H_\gamma^m([0,1]^d)$ for an arbitrary sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ of weights.

Note that the initial errors for L_2-multivariate approximation do not depend on the weights and are always 1. For $m = 1$ we have $H_\gamma^1([0,1]^d) = H_{d,1,\gamma}$ and obviously $H_\gamma^m([0,1]^d) \subseteq H_{d,1,\gamma}$. Therefore, the same conditions on the weights which we presented in this section for polynomial or strong polynomial tractability of L_2-multivariate approximation over the space $H_{d,1,\gamma}$ imply polynomial or strong polynomial

tractability for L_2-multivariate approximation over the space $H^m([0, 1]^d)$ for any m. Obviously, for the weighted space $H^m_\gamma([0, 1]^d)$, it is needed to check if these conditions are also necessary and find better estimates on the exponents of polynomial and strong polynomial tractability.

5.4.2 General Linear Weighted Problems

In this section we present a generalization of the multivariate problem APP_γ studied in the previous section, and obtain weighted linear problems different from those considered in Section 5.3.

The first step is the same as in Section 5.2. We take a linear tensor product problem $S = \{S_d\}$ defined for

$$S_d = S_1 \otimes \cdots \otimes S_1 \colon H_d = H_1 \otimes \cdots \otimes H_1 \to G_d = G_1 \otimes \cdots \otimes G_1,$$

where H_1 and G_1 are separable Hilbert spaces, and S_1 is a compact linear operator.

To define linear weighted tensor product problems differently than in Section 5.3, we assume that the space H_1 of the univariate case is constructed as follows. Let F_1 and F_2 be two Hilbert spaces such that $F_1 \cap F_2 = \{0\}$. Their inner products are denoted by $\langle \cdot, \cdot \rangle_{F_i}$. We define the Hilbert space H_1 as $H_1 = F_1 \oplus F_2$ with the inner product

$$\langle f, g \rangle_{H_1} = \langle f_1, g_1 \rangle_{F_1} + \langle f_2, g_2 \rangle_{F_2},$$

where $f, g \in H_1$ have the unique representation $f = f_1 + f_2$ and $g = g_1 + g_2$, with $f_1, g_1 \in F_1$ and $f_2, g_2 \in F_2$.

Then $H_d = (F_1 \oplus F_2) \otimes \cdots \otimes (F_1 \oplus F_2)$ and every $f \in H_d$ has a unique decomposition

$$f = \sum_{b \in \{1,2\}^d} f_b \quad \text{with } f_b \in F_b := F_{b_1} \otimes F_{b_2} \otimes \cdots \otimes F_{b_d}.$$

Here $b = [b_1, b_2, \ldots, b_d]$ is a vector whose components take value 1 or 2. Clearly, F_b is also a separable Hilbert space, namely, the d-fold tensor product of F_1 or F_2 depending on the values of b_j. We have

$$\langle f, g \rangle_{H_d} = \sum_{b \in \{1,2\}^d} \langle f_b, g_b \rangle_{F_b} \quad \text{for all } f, g \in H_d.$$

Let $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ be an arbitrary sequence of positive weights. Define the weighted space $H_{d,\gamma}$ as the space H_d with the inner product

$$\langle f, g \rangle_{H_{d,\gamma}} = \sum_{b \in \{1,2\}^d} \gamma^{-1}_{d,\mathfrak{u}_b} \langle f_b, g_b \rangle_{F_b} \quad \text{for all } f, g \in H_d.$$

Here $\mathfrak{u}_b = \{j \in [d] \mid b_j = 2\}$. There is a one-to-one correspondence between $b \in \{1,2\}^d$ and $\mathfrak{u} \subseteq [d]$. Indeed, if for $b, c \in \{1,2\}^d$, we have $\mathfrak{u}_b = \mathfrak{u}_c$ then $b = c$.

Similarly, if for a given $\mathfrak{u} \subseteq [d]$ we define $b_{\mathfrak{u}}$ by $(b_{\mathfrak{u}})_j = 2$ for $j \in \mathfrak{u}$, and $(b_{\mathfrak{u}})_j = 1$ for $j \notin \mathfrak{u}$, then $b_{\mathfrak{u}} = b_{\mathfrak{v}}$ implies that $\mathfrak{u} = \mathfrak{v}$.

The space $H_{d,\gamma}$ is a Hilbert space that is algebraically the same as the space H_d, and the norms of $H_{d,\gamma}$ and H_d are equivalent.

We also consider the case for which some weights $\gamma_{d,\mathfrak{u}}$ may be 0. If $\gamma_{d,\mathfrak{u}} = 0$ then we assume that $f_{b_{\mathfrak{u}}} = 0$ for all $f_{b_{\mathfrak{u}}}$ in F_b. If one of the weights is 0 then the Hilbert space $H_{d,\gamma}$ becomes a proper subspace of H_d.

Obviously, the linear operators S_d are well defined on $H_{d,\gamma}$ for any choice of γ. Note that although the values of $S_d f$ do not depend on γ, its adjoint S_d^* as well its norm $\|S_d\|_{H_{d,\gamma}}$ *do* depend on γ. This explains why both the information complexity and the tractability results depend on γ. To stress the dependence on γ, as in Section 5.3.3, we define $S_{d,\gamma} \colon H_{d,\gamma} \to G_d$ as $S_{d,\gamma} f = S_d f$, which makes sense since $H_{d,\gamma} \subseteq H_d$ and S_d is defined over H_d. Letting $S_\gamma = \{S_{d,\gamma}\}$, we have now defined the weighted problem that we will study in this section.

We assume that $\gamma_{d,\emptyset} = 1$. The weight $\gamma_{d,\emptyset}$ corresponds to $b_\emptyset = \vec{1} = [1, 1, \ldots, 1]$ and to the subspace

$$F_{\vec{1}} = F_1 \otimes F_1 \otimes \cdots \otimes F_1.$$

We stress that the dimension of F_1 can be arbitrary. As we shall see, its dimension will play a crucial role. This is the main difference between the constructions in this section and in Section 5.3.1. In the latter section, the construction was such that the weight $\gamma_{d,\emptyset}$ corresponded to a 1-dimensional space. Hence, both constructions may coincide only when $\dim(F_1) = 1$. Note that they do coincide if $F_1 = \mathrm{span}(e_1)$ and e_1 is an eigenfunction of $W_1 = S_{1,\gamma}^* S_{1,\gamma}$.

Obviously, some linear operators $S_{d,\gamma}$ are trivially tractable, independently of the weights $\gamma_{d,\mathfrak{u}}$. This holds for $\dim(S_1(H_1)) \leq 1$, i.e., when S_1 is a continuous linear operator of rank at most 1. In this case, $S_1 = \langle \cdot, h \rangle_{H_1} g$ for some $h \in H_1$ and some $g \in G_1$, so that

$$S_d = \langle \cdot, h \rangle_{H_d} g_d$$

with the d-fold tensor product elements $h_d = h \otimes \cdots \otimes h$ and $g_d = g \otimes \cdots \otimes g$. Since we can compute $\langle f, h_d \rangle_{H_d}$ for the class Λ^{all}, we solve the problem S_d exactly with at most one information operation. More precisely, if $h_d = 0$ then S_d is 0 and no information operation is needed, whereas if $h_d \neq 0$ then one information operation is sufficient. Hence, without loss of generality we may assume that $\dim(S_1(H_1)) \geq 2$.

Let $n(\varepsilon, d) = n^{\mathrm{wor}}(\varepsilon, S_{d,\gamma}, \Lambda^{\mathrm{all}})$ denote the information complexity of $S_{d,\gamma}$ for the absolute or normalized error criterion in the worst case setting and for the class Λ^{all}. The norm of $S_{d,\gamma}$ is obtained for some f from the unit ball of $H_{d,\gamma}$, i.e., $\|S_{d,\gamma}\| = \|S_{d,\gamma} f\|_{G_d}$. In what follows, we sometimes assume that we can choose f as an element of the unit ball of $F_{\vec{1}}$. That is, the norm of $S_{d,\gamma}$ is the same when we restrict this operator to the subspace $F_{\vec{1}}$ of $H_{d,\gamma}$.

We already explained why we need to assume that $\dim(S_1(H_1)) \geq 2$. As we shall see in a moment, if the last assumption is strengthened to $\dim(S_1(F_1)) \geq 2$ then polynomial tractability cannot hold. Since $H_1 = F_1 \oplus F_2$, this means that if S_1 restricted to F_1 is *not* a continuous linear functional, then polynomial tractability

cannot occur. Note that $\dim(S_1(F_1)) \geq 2$ implies that $\dim(F_1) \geq 2$. Later on, we shall see examples of multivariate problems for which these assumptions hold for arbitrary weights, including the case of all zero weights $\gamma_{d,\mathfrak{u}} = 0$ for $\mathfrak{u} \neq \emptyset$.

Theorem 5.12. *Consider the weighted problem S_γ defined as in this section in the worst case setting and for the class Λ^{all}.*

For the absolute error criterion we have:

- *If*
$$\left\| S_1 \big|_{F_1} \right\| > 1 \quad \text{and} \quad \dim(S_1(F_1)) \geq 2,$$
then S_γ is intractable for arbitrary weights $\gamma_{d,\mathfrak{u}}$ for $|\mathfrak{u}| > 0$.

- *If*
$$\left\| S_1 \big|_{F_1} \right\| = 1 \quad \text{and} \quad \dim(S_1(F_1)) \geq 2,$$
then S_γ is polynomially intractable for arbitrary weights $\gamma_{d,\mathfrak{u}}$ for $|\mathfrak{u}| > 0$.

  If additionally the largest eigenvalue of $S_1\big|_{F_1}^{*} S_1\big|_{F_1}$ is of multiplicity $k \geq 2$, then S_γ is intractable and
$$n(\varepsilon, d) \geq k^d$$
for arbitrary weights $\gamma_{d,\mathfrak{u}}$ for $|\mathfrak{u}| > 0$.

For the normalized error criterion we have:

- *If there exists a linear subspace X of F_1 of dimension at least 2 such that*
$$\left\| S_{d,\gamma} \right\| = \left\| S_d \big|_{X_d} \right\| > 0 \quad \text{with } X_d = X \otimes \cdots \otimes X \ (d \text{ times}),$$
then S_γ is polynomially intractable.

  If additionally there exists a positive number α such that
$$\| S_1 f \|_{G_1} = \alpha \| f \|_{F_1} \quad \text{for all } f \in X$$
then S_γ is intractable.

- *Consider product weights $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ with*
$$M := \sup_d \max_{j \in [d]} \gamma_{d,j} < \infty.$$

  If
$$\dim(S_1(F_1)) \geq 2$$
then S_γ is polynomially intractable.

  If additionally there exists a linear subspace X of F_1 of dimension at least 2 such that $S_1(X)$ is orthogonal to $S_1(F_2)$, and there exists a positive number α such that
$$\| S_1 f \|_{G_1} = \alpha \| f \|_{F_1} \quad \text{for all } f \in X \text{ and } M \| S_1 \big|_{F_2} \|^2 \leq \alpha^2,$$
then S_γ is intractable.

Proof. We first consider the absolute error criterion. Observe that the original problem $S_{d,\gamma}$ defined over $H_{d,\gamma}$ is not easier than the problem $S_{d,\gamma}$ defined over the subspace $F_{\bar{1}}$. Note that $S_{d,\gamma}$ over $F_{\bar{1}}$ is an unweighted tensor product problem that does not depend on the weights $\gamma_{d,\mathfrak{u}}$. That is, $S_{d,\gamma}f = S_d f$ for $f \in F_{\bar{1}} = F_1 \otimes \cdots \otimes F_1$ with the norm of $f = f_1 \otimes f_2 \otimes \cdots \otimes f_d$, for $f_j \in F_1$, given by $\|f\|_{F_d} = \prod_{j=1}^{d} \|f_j\|_{F_1}$. As always, the information complexity of $S_{d,\gamma}$ over $F_{\bar{1}}$ depends on the eigenvalues of the operator $W_{d,F_{\bar{1}}} = S_d\big|_{F_{\bar{1}}}^{*} S_d\big|_{F_{\bar{1}}}$. Due to the tensor product structure of $S_{d,\gamma}$ and $F_{\bar{1}}$, these eigenvalues are products $\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d}$, where $j_k \in \mathbb{N}$ and $\{\lambda_j\}$ are the ordered eigenvalues of W_1. Since $\dim(S_1(F_1)) \geq 2$ we know that $\lambda_1 \geq \lambda_2 > 0$. Since $\|S_1\big|_{F_1}\| \geq 1$, we know that

$$\|S_{d,\gamma}\big|_{F_{\bar{1}}}\| = \lambda_1^{d/2} \geq 1.$$

Hence

$$n(\varepsilon, d) \geq \left|\left\{j \in \mathbb{N}^d \mid \lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d} > \varepsilon^2\right\}\right|.$$

Applying Theorem 5.4, we conclude that S_γ is intractable for $\lambda_1 = \|S_1\big|_{F_1}\| > 1$, and polynomially intractable for $\lambda_1 = \|S_1\big|_{F_1}\| = 1$. Furthermore if we have $\lambda_2 = 1$ then the same theorem implies that S_γ is intractable. Note that all these results hold for arbitrary weights $\gamma_{d,\mathfrak{u}}$ for $|\mathfrak{u}| > 0$.

We now turn to the normalized error criterion. Consider the first part. Since $\|S_{d,\gamma}\| = \|S_d\big|_{X_d}\|$, the both problems $S_{d,\gamma}$ and the unweighed problem S_d over $X_d \subseteq F_{\bar{1}}$ have the same initial error. The information complexity of S_d over X_d depends on the eigenvalues of $W_{d,X_d} = S_d\big|_{X_d}^{*} S_d\big|_{X_d}$. Since S_d and X_d have tensor product structure, the eigenvalues are again of the form $\lambda_{j_1,X}\lambda_{j_2,X}\cdots\lambda_{j_d,X}$, where $j_k \in \mathbb{N}$ and $\{\lambda_{j,X}\}$ are the ordered eigenvalues of $W_{1,X}$. Then $\dim(S_1(X)) \geq 2$ implies that $\lambda_1 \geq \lambda_2 > 0$ and $\|S_{d,\gamma}\| = \|S_d\big|_{X_d}\| = \lambda_1^{d/2}$. Hence

$$n(\varepsilon, d) \geq \left|\left\{j \in \mathbb{N}^d \mid \lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_d} > \varepsilon^2\lambda_1^d\right\}\right|.$$

Applying Theorem 5.5, we conclude that S_γ is polynomially intractable.

Assume now that $\|S_1 f\|_{G_1} = \alpha\|f\|_{F_1}$ for all $f \in X$. This implies that $W_{1,X} f = \alpha f$ for all $f \in X$ and therefore all eigenvalues of $W_{1,X}$ are α. The largest eigenvalue is $\lambda_1 = \alpha$, and since $\dim(X) \geq 2$, its multiplicity is at least 2. Again Theorem 5.5 implies that S_γ is intractable, as claimed.

We proceed to the second part for product weights. We do not assume that $S_{d,\gamma}$ and $S_d\big|_{X_d}$ have the same initial error, and so the previous reasoning does not apply.

We first consider the case $d = 1$. As always, let $\{\lambda_{1,j}\}$ denote the ordered eigenvalues of $W_1 = S_{1,\gamma}^{*} S_{1,\gamma}: H_{1,\gamma} \to H_{1,\gamma}$. Here $\gamma \in (0, M]$. The largest $\lambda_{1,\gamma}$ is obviously equal to the square of the norm $\|S_1\|$. For $f = f_1 + f_2$, where $f_i \in F_i$, we have

$\|f\|^2_{H_{1,\gamma}} = \|f_1\|^2_{F_1} + \gamma^{-1}\|f_1\|^2_{F_2}$ and

$$\|S_1 f\|_{G_1} \|S_1\big|_{F_1}\| \|f_1\|_{F_1} + \gamma^{1/2}\|S_1\big|_{F_2}\|\gamma^{-1/2}\|f_2\|_{F_2}$$
$$\leq \left(\|S_1\big|_{F_1}\|^2 + \gamma\|S_1\big|_{F_2}\|^2\right)^{1/2}\|f\|_{H_{1,\gamma}}.$$

This proves that

$$\lambda_{1,\gamma} \leq \|S_1\big|_{F_1}\|^2 + \gamma\|S_1\big|_{F_2}\|^2 \leq \|S_1\big|_{F_1}\|^2 + M\|S_1\big|_{F_2}\|^2.$$

Since the last estimate is also true for $\gamma = 0$, it holds for all $\gamma \in [0, M]$.

We need a lower bound estimate of $\lambda_{2,\gamma}$. Recall that

$$\lambda_{2,\gamma} = \inf_{h \in H_{1,\gamma}} \sup_{f \in H_{1,\gamma}, \langle f,h \rangle_{H_{1,\gamma}} = 0} \frac{\langle W_1 f, f \rangle_{H_{1,\gamma}}}{\langle f, f \rangle_{H_{1,\gamma}}}.$$

If we replace $H_{1,\gamma}$ in the supremum by F_1, then we obtain a lower bound on $\lambda_{2,\gamma}$. For $f \in F_1$, we have

$$\langle W_1 f, f \rangle_{H_{1,\gamma}} = \|S_1 f\|^2_{G_1} = \langle V f, f \rangle_{F_1},$$

where $V = S_1\big|^*_{F_1} S_1\big|_{F_1} : F_1 \to F_1$. Since $\langle f, h \rangle_{H_{1,\gamma}} = \langle f, h_1 \rangle_{F_1}$, the last infimum over $h \in H_{1,\gamma}$ is the same as the infimum over $h \in F_1$. Therefore we have

$$\lambda_{2,\gamma} \geq \inf_{h \in F_1} \sup_{f \in F_1, \langle f,h \rangle_{F_1} = 0} \frac{\langle V f, f \rangle_{F_1}}{\langle f, f \rangle_{F_1}} = \lambda_2(V),$$

where $\lambda_2(V)$ is the second largest eigenvalue of V. Since $\dim(V(F_1)) = \dim(S_1(F_1))$ is at least 2 by hypothesis, we conclude that $\lambda_2(V)$ is positive. Hence,

$$0 < \lambda_2(V) \leq \lambda_1(V) = \|S_1\big|_{F_1}\|^2.$$

We now turn to $d \geq 2$. Let $\{\lambda_{d,j,\gamma}\}$ denote the ordered eigenvalues of $W_d = S_d^* S_d : H_{d,\gamma} \to H_{d,\gamma}$. For product weights, we have

$$H_{d,\gamma} = H_{1,\gamma_{d,1}} \otimes \cdots \otimes H_{1,\gamma_{d,d}}.$$

From the tensor product structure, we have

$$\{\lambda_{d,j,\gamma}\} = \left\{\textstyle\prod_{k=1}^d \lambda_{j_k,\gamma_{d,k}} \mid j_k = 1, 2, \ldots\right\},$$

where $\gamma_{d,j} \in [0, M]$.

The square of the norm of $S_{d,\gamma}$ is the largest eigenvalue $\lambda_{d,1,\gamma}$, and so

$$\|S_{d,\gamma}\|^2_{H_{d,\gamma}} = \lambda_{d,1,\gamma} = \prod_{j=1}^d \lambda_{1,\gamma_j} \leq \prod_{j=1}^d \left(\|S_1\big|_{F_1}\|^2 + M\|S_1\big|_{F_2}\|^2\right).$$

We now find a lower bound on

$$n(\varepsilon, d) = \min\{n \mid \lambda_{d,n+1,\gamma} \leq \varepsilon^2 \lambda_{d,1,\gamma}\},$$

for a specific ε. Take an arbitrary integer k, and fix ε as

$$\varepsilon = \frac{1}{2} \left(\frac{\lambda_2(V)}{\|S_1|_{F_1}\|^2 + M\|S_1|_{F_2}\|^2} \right)^{k/2} \in (0, 1).$$

For $d > k$, consider the vectors $i = [i_1, i_2, \ldots, i_d]$ with $i_j \in \{1, 2\}$. Take k indices i_j equal to 2 and $d - k$ indices i_j equal to 1. We have $\binom{d}{k}$ such vectors, for which the eigenvalues satisfy

$$\prod_{j=1}^{d} \lambda_{i_j, \gamma_{d,j}} = \prod_{j:i_j=1} \lambda_{1, \gamma_{d,j}} \prod_{j:i_j=2} \lambda_{2, \gamma_{d,j}} = \prod_{j:i_j=2} \frac{\lambda_{2, \gamma_{d,j}}}{\lambda_{1, \gamma_{d,j}}} \prod_{j=1}^{d} \lambda_{1, \gamma_{d,j}}.$$

Since

$$\frac{\lambda_{2, \gamma_{d,j}}}{\lambda_{1, \gamma_{d,j}}} \geq \frac{\lambda_2(V)}{\|S_1|_{F_1}\|^2 + M\|S_1|_{F_2}\|^2},$$

we conclude that

$$\prod_{j=1}^{d} \lambda_{i_j, \gamma_{d,j}} \geq 4\varepsilon^2 \lambda_{d,1,\gamma}.$$

This proves that

$$n(\varepsilon, d) \geq \binom{d}{k} = \Theta(d^k) \quad \text{as } d \to \infty.$$

Since k can be arbitrarily large, this means that S_γ is polynomially intractable, as claimed.

We proceed to the last part of the theorem. Since $S(f_1)$ is orthogonal to $S(f_2)$ for all $f_1 \in X \subseteq F_1$ and for all $f_2 \in F_2$, we can improve our upper bound on $\lambda_{1,\gamma} = \|S_1\|^2$. For $f \in F_1$ we have

$$\begin{aligned}
\|S_1 f\|_{G_1}^2 &= \|S_1 f_1 + S_1 f_2\|_{G_1}^2 = \|S_1 f_1\|_{G_1}^2 + \|S_1 f_2\|_{G_1}^2 \\
&\leq \alpha^2 \|f_1\|_{F_1}^2 + \gamma \|S_1|_{F_2}\|^2 \gamma^{-1} \|f_2\|_{F_2}^2 \\
&\leq \max\left(\alpha^2, \gamma \|S_1|_{F_2}\|^2\right) \|f\|_{H_{1,\gamma}}^2.
\end{aligned}$$

Hence,

$$\lambda_{1,\gamma} = \|S_1\|^2 \leq \max\left(\alpha^2, \gamma \|S_1|_{F_2}\|^2\right).$$

For $\gamma \leq M$, the last assumption $M\|S_1|_{F_2}^2 \leq \alpha^2$ implies that $\lambda_{1,\gamma} \leq \alpha^2$. Since $\|S_1 f\|_{G_1} = \alpha \|f\|_{F_1}$ for all $f \in X$, then $\|S_1\| \geq \alpha$. Hence

$$\lambda_{1,\gamma} = \alpha \quad \text{for all } \gamma \in [0, M].$$

For $d \geq 2$ and $\gamma_{d,j} \in [0, M]$, we therefore have

$$\lambda_{d,1,\gamma} = \prod_{j=1}^{d} \lambda_{1,\gamma_{d,j}} = \alpha^d.$$

This means that $\|S_d\| = \|S_d|_{X_d}\|$, and S_γ is intractable due to the previous point. This completes the proof. $\square$

We now comment on the assumptions of Theorem 5.12. First, we discuss the absolute error criterion and show that $\|S_1|_{F_1}\| \geq 1$ is needed. Indeed, if $\|S_1|_{F_1}\| < 1$ and $\dim(S_1(F_1)) \geq 2$, we may have polynomial tractability. Indeed, take all $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > 0$. Then $S_d = S_d|_{F_1}$ and $\lambda_1 = \|S_1\|^2 < 1$. If we assume that $\lambda_n = \mathcal{O}(n^{-r})$ for some positive r then Theorem 5.4 states that S_γ is even strongly polynomially tractable. On the other hand, $\lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2})$ implies weak tractability of S_γ.

For the normalized error criterion, the assumption $\|S_{d,\gamma}\| = \|S_d|_{X_d}\| > 0$ implies polynomial intractability, but does not imply intractability in general. Indeed, let $\gamma_{d,\mathfrak{u}} = 1$ for all $\mathfrak{u}$ and let

$$\lambda_2 < \lambda_1 \quad \text{and} \quad \lambda_n = o((\ln n)^{-2}(\ln \ln n)^{-2}).$$

Then Theorem 5.5 implies that S_γ is weakly tractable.

For product weights, the assumption $M\|S_1|_{F_2}\|^2 \leq \alpha^2$ is needed in general. Indeed, take $\gamma_{d,j} = 1$ for all j. Define $F_1 = \mathrm{span}(e_1, e_2)$, $F_2 = \mathrm{span}(e_3, e_4, \dots)$ for orthonormal $\{e_j\}_{j\in\mathbb{N}}$. Let $G_1 = H_1 = F_1 \oplus F_2$, and

$$S_1 e_1 = e_1, \quad S_1 e_2 = e_2, \quad S_1 e_j = \beta_j e_j \text{ for } j \geq 3$$

with $\beta_j = \beta n^{-r}$ for some positive β and r. For $d = 1$, it is easy to check that the eigenvalues of W_1 are

$$1, 1, \beta, \beta 2^{-r}, \dots, \beta n^{-r}, \dots.$$

Then $M = \alpha = 1$ and $\|S_1|_{F_2}\|^2 = \beta$. For $\beta \leq 1$, the last point of Theorem 5.3 applies, and we have intractability of S_γ. So, let us assume that $\beta > 1$. Then $\lambda_1 = \beta$, $\lambda_2 = \max(1, \beta/2^r) < \lambda_1$ and $\lambda_n = \Theta(n^{-r})$. From Theorem 5.5 we know that S_γ is now weakly tractable.

We illustrate Theorem 5.12 by three examples of multivariate approximation defined for several standard Sobolev spaces.

Example 5.13. We show that Theorem 5.12 applies for the multivariate approximation problem studied in Section 5.4.1. Take the space $H_{1,m} = H_{1,m,\gamma}$ for $\gamma = 1$ from Section 5.4.1. We show that $H_{1,m} = F_1 \cap F_2$ for appropriately defined F_i.

Let F_1 be the space of univariate polynomials of degree at most $m - 1$ equipped with the $L_2 = L_2([0, 1])$-norm. Define

$$F_2 = \{f \in H_{1,m} \mid \langle f, p\rangle_{L_2} = 0 \text{ for all } p \in F_1\}$$

and equip F_2 with the norm of $H_{1,m}$. Clearly, $F_1 \cap F_2 = \{0\}$.

For $f \in H_{1,m}$, let $f_1 = \sum_{j=1}^{m} \langle f, p_j \rangle_{L_2} p_j$ for orthonormal polynomials $p_j \in F_1$. Then obviously $f_1 \in F_1$, and $f_2 = f - f_1 \in F_2$. It is easy to see that the decomposition $f = f_1 + f_2$ is unique, and

$$\|f\|_{H_{1,m}}^2 = \|f_1\|_{H_{1,m}}^2 + \|f_2\|_{H_{1,m}}^2.$$

Using the weighted tensor construction it is easy to see that the space $H_{d,m,\gamma}$ of Section 5.4.1 is the same as the space $H_{d,\gamma}$ of Section 5.4.2. Finally, we let $G_d = L_2([0,1]^d)$, and $S_d f = f$ for all $f \in H_{d,\gamma}$. This corresponds to the multivariate approximation problem APP_γ of Section 5.4.1.

Then $S_1 f = f$ and $\|S_1|_{F_1}\| = \|S_1\| = 1$ and we can take $X = F_1$, which implies that $X_d = F_{\bar{1}}$. In this case we have $k = m$ and the statements of Theorems 5.12 and 5.9 coincide.

Example 5.14. The purpose of this example is to demonstrate a different weighted problem with the space F_1 of arbitrarily large dimension, which can lead to negative tractability results.

Let $r \geq 1$. Define $F_1 = \mathrm{span}(1, x, \dots, x^{r-1})$ as the r-dimensional space of polynomials of degree at most $r - 1$ restricted to the interval $[0,1]$, with the inner product

$$\langle f, g \rangle_{F_1} = \sum_{j=0}^{r-1} f^{(j)}(0) g^{(j)}(0).$$

Let F_2 be the space of functions f defined over $[0,1]$ for which $f^{(r-1)}$ is absolutely continuous, $f^{(r)}$ belongs to $L_2([0,1])$, and $f^{(j)}(0) = 0$ for $j = 0, 1, \dots, r - 1$. The inner product in F_2 is

$$\langle f, g \rangle_{F_2} = \int_0^1 f^{(r)}(t) g^{(r)}(t) \, dt.$$

Obviously, $F_1 \cap F_2 = \{0\}$, as required in our analysis. We thus have

$$H_{1,\gamma} = \{f : [0,1] \to \mathbb{R} \mid f^{(r-1)} \text{ absolutely continuous}, \ f^{(r)} \in L_2([0,1])\}$$

with inner product

$$\langle f, g \rangle_{H_{1,\gamma}} = \sum_{j=0}^{r-1} f^{(j)}(0) g^{(j)}(0) + \gamma^{-1} \int_0^1 f^{(r)}(t) g^{(r)}(t) \, dt.$$

It is well known that the Hilbert space $H_{1,\gamma}$ has a reproducing kernel of the form

$$K_{1,\gamma}(x,t) = \sum_{j=0}^{r-1} \frac{x^j}{j!} \frac{t^j}{j!} + \gamma \int_0^1 \frac{(x-u)_+^{r-1}}{(r-1)!} \frac{(t-u)_+^{r-1}}{(r-1)!} \, dt,$$

where as always $u_+ = \max(u, 0)$, see e.g. [176].

For the approximation problem $S_d f = f$ for all $f \in H_{d,\gamma}$ with $G_d = L_2([0,1]^d)$, we have

$$\|S_1\big|_{F_1}\| > 1 \quad \text{and} \quad \dim(S_1(F_1)) = r.$$

The first inequality can be easily checked by taking, for example, the function $f(x) = (1+x)/\sqrt{2}$. Then $f \in F_1$ for all $r \geq 1$, and $\|f\|_{F_1} = 1$ whereas $\|S_1 f\|_{G_1} = \sqrt{7/6}$.

For $r \geq 2$, Theorem 5.12 states that this approximation problem is intractable for the absolute error criterion with arbitrary weights $\gamma_{d,\mathfrak{u}}$ for $|\mathfrak{u}| > 0$, and polynomially intractable for the normalized error criterion and arbitrary product weighs.

We leave the reader the task of checking tractability for the normalized error criterion with general weights.

For $r \geq 1$, it is known that the eigenvalues λ_j of the operator $S_1\big|_{H_2}^* S_1\big|_{H_2}$ are proportional to j^{-2r}. Hence for $r = 1$, we have $F_1 = \mathrm{span}(e_1)$ with $e_1(x) \equiv 1$. Furthermore e_1 is an eigenfunction of W_1. Therefore the construction of this section coincides with the construction of Section 5.3.1. Therefore we can use Theorem 5.6 and conclude that for product weights, the approximation problem is strongly polynomially tractable iff $p_{\gamma^*} < \infty$, and then the exponent of strong polynomial tractability is $\max(2p_{\gamma^*}, 1)$.

Example 5.15. The purpose of this example is to show that the splitting between the unweighted and weighted parts for the univariate case may change tractability results.

We consider spaces similar to those from the previous example, but with a different split between the unweighted and weighted parts. For $r \geq 1$ take an integer $k \in [1, r]$. Define $F_{1,k} = \mathrm{span}(1, x, \ldots, x^{k-1})$ and

$$F_{2,k} = \mathrm{span}(x^k, x^{k+1}, \ldots, x^{r-1}) \oplus F_2,$$

with F_2 as before. Then we have $H_{1,k,\gamma} = F_{1,k} \oplus F_{2,k}$ with the kernel

$$K_{1,k,\gamma}(x,t) = 1 + \sum_{j=1}^{k-1} \frac{x^j}{j!}\frac{t^j}{j!} + \gamma\left(\sum_{j=k}^{r-1} \frac{x^j}{j!}\frac{t^j}{j!} + \int_0^1 \frac{(x-u)_+^{r-1}}{(r-1)!}\frac{(t-u)_+^{r-1}}{(r-1)!}\,\mathrm{d}t\right).$$

This corresponds to the inner product

$$\langle f, g\rangle_{H_{1,k,\gamma}} = \sum_{j=0}^{k-1} f^{(j)}(0)g^{(j)}(0)$$

$$+ \gamma^{-1}\left(\sum_{j=k}^{r-1} f^{(j)}(0)g^{(j)}(0) + \int_0^1 f^{(r)}(t)g^{(r)}(t)\,\mathrm{d}t\right).$$

For the approximation problem of the previous example, we have

$$\dim(S_1(F_{1,k})) = k.$$

Hence, for $k \geq 2$ we have the same intractability results as before.

It is known that the eigenvalues λ_j of $S_1\big|^*_{F_{2,k}} S_1\big|_{F_{2,k}}$ are still of order j^{-2r}. Hence, for $k = 1$ and product weights the approximation problem is strongly polynomially tractable iff $p_{\gamma^*} < \infty$ with the exponent of strong polynomial tractability $\max(2p_{\gamma^*}, r^{-1})$. Note that the exponent of strong polynomial tractability may now be much smaller than in the previous example.

5.5 Notes and Remarks

NR 5:1. The results presented in this chapter on weak tractability are new. The results on polynomial tractability can be found in many papers and we will try to identify them in the successive notes.

The first papers on tractability were written for the worst case setting and for the absolute error criterion, later the emphasis shifted to the normalized error criterion. As already briefly mentioned in Chapter 2, the absolute error criterion has the nice property that the sum of two polynomially or weakly tractable problems is still polynomially or weakly tractable. The last property is generally not true for the normalized error criterion, see [287] or Example 4.45. On the other hand, tractability results for the absolute error results are highly dependent on scaling. Probably the most visible example is for linear (unweighted) tensor product problems for which the initial error in the d-dimensional case is $\lambda_1^{d/2}$, where $\lambda_1^{1/2}$ is the largest singular value for the univariate problem. Hence, the problem is properly scaled only for $\lambda_1 = 1$, whereas for $\lambda_1 < 1$ it is exponentially small in d, and for $\lambda_1 > 1$ is exponentially large in d. Hence, for $\lambda_1 < 1$ we approximate a linear operator with an exponentially small norm, and for $\lambda_1 > 1$, we approximate a linear operator with an exponentially large norm, both cases equally unsatisfying. We believe that a proper scaling can be achieved by a proper choice of weights. In any case, we study both the absolute and normalized error criteria in the book, for both the unweighted and weighted cases.

NR 5.1:1. Theorem 5.1 on polynomial tractability for the absolute error criterion is a variant of Theorem 4.1 of [287]. In the latter paper the condition (5.1) is replaced by an equivalent condition (5.3). The reader is also referred to [269] for more information. It is interesting to notice that in this first tractability paper, there is also a notion of tractability in ε^{-1} and d. Tractability in ε^{-1} means that we do not care about the dependence on d and want to have only a polynomial dependence in ε^{-1}. Similarly, tractability in d means that we do not care about the dependence on ε^{-1} and want to have only a polynomial dependence in d. In this book we do not follow these concepts and always insist on at least non-exponential dependence on both ε^{-1} and d. However, when we study generalized tractability in Chapter 8, we may restrict the domain either of ε^{-1} to the interval $[\varepsilon_0, 1]$ or d to the interval $[1, d^*]$ and then the dependence on ε^{-1} or d do not play a role. This is similar in spirit to the notions of tractability in ε^{-1} and d.

NR 5.1:2. Theorem 5.2 on polynomial tractability for the normalized error criterion cannot be formally found in the literature, although similar reasoning has been used in many papers. The idea of using the sum of some powers of the eigenvalues to deduce polynomial tractability conditions was first used for the average case setting in [94].

NR 5.2:1. We define a linear tensor product problem as d tensor copies of a linear problem defined on univariate functions. As already mentioned in the footnote of this section (see p. 174), it would be possible to start with any compact operator $S_1 \colon H_1 \to G_1$ between arbitrary separable Hilbert spaces and have d tensor copies of it. For example, H_1 could be a space of m-variate functions. This would correspond formally to a measurable subset D of $\mathbb{R}^m$ and S_1 would be an m-variate linear operator. The operators S_d would be linear operators defined on spaces of dm-variate functions, etc. All theorems would look the same with the interpretation that the eigenpairs (λ_i, e_i) of W_1 correspond to the m-variate case. This approach was present in [93], [273] since for some problems it is more natural to have the first basic step with $m > 1$, see Kuo and Sloan [120], Kwas [127], Kwas and Li [128], and Li [134]. More precisely, Kuo and Sloan studied integration over tensor products of spheres and in this case m is at least 2, whereas Kwas and Li studied, in particular, multivariate (m-variate) Feynman–Kac path integration for an arbitrarily large m. Then approximation of path integration requires us to work with tensor products of dm-variate functions.

We decide to simplify the analysis and to deal with one less parameter and opt for $m = 1$. The reader should, however, remember that this is done only for simplicity, and that everything goes through for a general Hilbert space H_1.

NR 5.2:2. Polynomial tractability as presented in Theorem 5.5 is an elaborated version of Theorem 3.1 of [288]. In the latter theorem, conditions on polynomial tractability in ε^{-1} and d are also given.

NR 5.3:1. This is a major section of the book introducing weighted spaces and weighted multivariate problems. We restricted ourselves to Hilbert spaces and linear problems, although it should be clear how to generalize these ideas to more general spaces and problems. The restriction to Hilbert spaces was partially done for simplicity to reduce technical difficulties in presenting the new concepts. But the main reason was that we have the explicit formula for the information complexity only in Hilbert spaces and for linear problems. For more general weighted spaces, the analysis would probably require more specific assumptions about multivariate problems. For example, if we drop the assumption that S_d is linear, then the analysis depends on the specific form of S_d. This point will be illustrated in Volumes II and III, where we discuss a number of non-linear problems.

NR 5.3:2. The construction of weighted Hilbert spaces in this section is similar to the construction done in [271], [272], see also [126].

NR 5.3:3. We mentioned in the text who first used each specific type of weights. We believe that new types of weights will be proposed in the near future and they will be

modeling different applications of multivariate problems. We wish to remind the reader that the concept of weights is relative new since their first use was proposed in [216] in 1998, which is just 10 years ago. It is worth adding that finite-diameter weights were proposed by Creutzig [33] in 2007 at the time when we were working on this book.

NR 5.3:4. Theorem 5.7 in its full generality is new. However, the case of product weights has been studied in [271], and the case of finite-order weights in [272], [273]. The latter papers also study the class Λ^{std}, and we return to these papers in Volumes II and III.

NR 5.4:1. As already mentioned, Subsection 5.4.1 is based on [280]. Subsection 5.4.2 is in turn partially based on [292], where the same problem is analyzed for product weights.

Chapter 6

Average Case Setting

In the previous chapter we studied polynomial and weak tractability for linear multivariate problems in the worst case setting and for the class Λ^{all}. In this chapter we keep most of the assumptions from the previous chapter but change the worst case setting to the average one.

In Section 6.1, we consider a sequence of linear multivariate problems $\{S_d\}$ for which S_{d+1} is not necessarily related to S_d. Polynomial and weak tractability conditions are expressed, similarly to those in the worst case setting, in terms of summability conditions of the singular values of the linear multivariate problems. In the worst case setting everything depends on the behaviour of the largest singular values, whereas in the average case setting everything depends on the truncated traces of the singular values. As before, we study the absolute and normalized error criteria and show that, in general, polynomial and weak tractabilities for these two error criteria are not related.

In Section 6.2, we generalize the construction of linear tensor product problems presented in the worst case setting. In the average case setting only the target space is assumed to be a Hilbert space; moreover, only this space needs to be a tensor product space. The source space may be more general; we only need to assume that it is a separable Banach space. It is relatively easy to show that non-trivial linear tensor product problems are intractable for the normalized error criterion, and they are also intractable for the absolute error criterion if the initial error is at least one. For the absolute error criterion with initial error less than one, we show that polynomial tractability is equivalent to strong polynomial tractability and they both hold if the singular eigenvalues decay to 0 as $\mathcal{O}(n^{-p})$ with $p > 1$.

Section 6.3 deals with weighted tensor product problems. We study necessary and sufficient conditions on the weights to get polynomial and weak tractability for the normalized error criterion. In particular, strong polynomial tractability holds for product weights iff both the sequences of the singular values and weights are summable with some power less than one. For finite-order weights, polynomial tractability holds iff the sequence of the singular values is summable with some power less than one independently of the specific values of finite-order weights. These conditions are significantly relaxed when weak tractability is considered.

We illustrate the results of this chapter by several linear multivariate problems. In particular, we study multivariate approximation for weighted Korobov spaces which has been studied before in the worst case setting. This will allow us to compare tractability results for this problem in both the worst case and average case settings.

6.1 Linear Problems

In this section, we consider linear multivariate problems defined from separable Banach spaces into Hilbert spaces and equipped with zero-mean Gaussian measures. We study their polynomial and weak tractability for the absolute and normalized error criteria in the average case setting and for the class Λ^{all}.

As in Section 4.4 of Chapter 4, we consider the problem $S = \{S_d\}$, where $S_d : F_d \to G_d$ is a linear operator and F_d is a separable Banach space whereas G_d is a Hilbert space. The space F_d is equipped with the zero-mean Gaussian measure μ_d, so that $\nu_d = \mu_d S_d^{-1}$ is the zero-mean Gaussian measure defined on the space of the solution elements $S_d(f)$. Its correlation operator C_{ν_d} has the eigenpairs

$$C_{\nu_d} \eta_{d,j} = \lambda_{d,j} \eta_{d,j},$$

where $\langle \eta_{d,j}, \eta_{d,i} \rangle_{G_d} = \delta_{i,j}$ and $\lambda_{d,1} \geq \lambda_{d,2} \geq \cdots \geq 0$, with

$$\text{trace}(C_{\nu_d}) = \sum_{j=1}^{\infty} \lambda_{d,j} < \infty.$$

We first consider tractability of S for the absolute error criterion. We know from Section 4.3.1 of Chapter 4 that the nth minimal error of the linear operator S_d in the average case setting for the class Λ^{all} is the truncated trace

$$\left(\sum_{j=n+1}^{\infty} \lambda_{d,j} \right)^{1/2}$$

and the information complexity $n(\varepsilon, d) := n^{\text{avg}}(\varepsilon, S_d, \Lambda^{\text{all}})$ is

$$n(\varepsilon, d) = \min \left\{ n \mid \sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \right\}.$$

To omit the trivial case, we assume that S is non-zero, i.e., no S_d is the zero operator, which implies that $\lambda_{d,1} > 0$ for all $d \in \mathbb{N}$. We first study polynomial tractability.

Theorem 6.1. *Consider the non-zero problem $S = \{S_d\}$ for a linear S_d defined from a separable Banach space F_d into a Hilbert space G_d. We study the problem S for the absolute error criterion in the average case setting and for the class Λ^{all}.*

- *S is polynomially tractable iff there exist a positive C_1, non-negative q_1, q_2 and $\tau \in (0, 1)$ such that*

$$C_2 := \sup_d \left(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} d^{-q_2} < \infty. \tag{6.1}$$

- *If* (6.1) *holds then*

$$n(\varepsilon, d) \leq \left(C_1 + \left(\frac{\tau C_2}{1 - \tau} \right)^{\tau/(1-\tau)} + 1 \right) d^{\max(q_1, q_2 \tau/(1-\tau))} \varepsilon^{-2\tau/(1-\tau)}$$

  for all $\varepsilon \in (0, 1]$ *and* $d = 1, 2, \ldots$.

- *If S is polynomially tractable, so that* $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ *for some positive C and p, and* $q \geq 0$, *then* (6.1) *holds with*

$$C_1 = 2C + 1, \quad q_1 = q, \quad q_2 = 2qp^{-1},$$

  and for any τ *such that* $\tau \in ((1 + 2/p)^{-1}, 1)$. *Then*

$$C_2 \leq 2(4C)^{2/p} \left[\zeta \left(\tau \left(1 + \frac{2}{p} \right) \right) \right]^{1/\tau},$$

  where ζ *is the Riemann zeta function.*

- *S is strongly polynomially tractable iff* (6.1) *holds with* $q_1 = q_2 = 0$. *The exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = \inf \left\{ \frac{2\tau}{1 - \tau} \mid \tau \text{ satisfies } (6.1) \text{ with } q_1 = q_2 = 0 \right\}.$$

Proof. The proof is similar to the proof of Theorem 5.1 in the worst case setting, which allows us to proceed a little faster.

Assume first that we have polynomial tractability with $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some positive C and p, and $q \geq 0$. This means that $\sum_{j=n(\varepsilon, d)+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2$. Since the eigenvalues $\lambda_{d,j}$ are non-increasing, we have

$$\sum_{j=\lfloor C d^q \varepsilon^{-p} \rfloor + 1}^{\infty} \lambda_{d,j} \leq \varepsilon^2.$$

Without loss of generality we assume that $C \geq 1$. Let $k = \lfloor C d^q \varepsilon^{-p} \rfloor + 1 \geq 2$. If we vary $\varepsilon \in (0, 1]$ then k takes the values $k = \lfloor C d^q \rfloor + 1, \lfloor C d^q \rfloor + 2, \ldots$. We also have $k \leq C d^q \varepsilon^{-p} + 1$, which is equivalent to $\varepsilon^2 \leq (C d^q/(k-1))^{2/p}$. Hence

$$\sum_{j=k}^{\infty} \lambda_{d,j} \leq \left(\frac{C d^q}{k - 1} \right)^{2/p} \quad \text{for all } k \geq \lfloor C d^q \rfloor + 1.$$

Observe that

$$k \lambda_{d, 2k-1} \leq \sum_{j=k}^{\infty} \lambda_{d,j} \leq (C d^q)^{2/p} \left(\frac{2k-1}{k-1} \right)^{2/p} \left(\frac{1}{2k-1} \right)^{2/p},$$

$$(k + 1) \lambda_{d, 2k} \leq \sum_{j=k}^{\infty} \lambda_{d,j} \leq (C d^q)^{2/p} \left(\frac{2k}{k-1} \right)^{2/p} \left(\frac{1}{2k} \right)^{2/p}.$$

This yields

$$\lambda_{d,2k-1} \le (Cd^q)^{2/p} \frac{2k-1}{k} \left(\frac{2k-1}{k-1}\right)^{2/p} \left(\frac{1}{2k-1}\right)^{1+2/p},$$

$$\lambda_{d,2k} \le (Cd^q)^{2/p} \frac{2k}{k+1} \left(\frac{2k}{k-1}\right)^{2/p} \left(\frac{1}{2k}\right)^{1+2/p}.$$

Since $k \ge 2$ we have $2k/(k-1) \le 4$, and the last estimates can be simplified to

$$\lambda_{d,j} \le 2(4Cd^q)^{2/p} j^{-(1+2/p)} \quad \text{for } j \ge \lceil (2C+1)d^q \rceil \ge 2 (\lfloor Cd^q \rfloor + 1) - 1.$$

Take $\tau \in ((1+2/p)^{-1}, 1)$. Then

$$\left(\sum_{j=\lceil (2C+1)d^q \rceil}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} \le 2(4Cd^q)^{2/p} \zeta\left(\tau(1+2/p)\right)^{1/\tau}. \tag{6.2}$$

Hence, (6.1) holds with $C_1 = 2C + 1$, $q_1 = q$, $q_2 = 2q/p$, and any τ such that $\tau \in ((1+2/p)^{-1}, 1)$ and $C_2 \le 2(4C)^{2/p}\zeta(\tau(1+2/p))^{1/\tau}$. This also proves the third point of the theorem.

Assume now that (6.1) holds. Since $\lambda_{d,j}$'s are ordered, we have as in (5.2),

$$\left(n - \lceil C_1 d^{q_1} \rceil + 1\right)^{1/\tau} \lambda_{d,n} \le \left(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} \le C_2 d^{q_2} \tag{6.3}$$

for $n = \lceil C_1 d^{q_1} \rceil, \lceil C_1 d^{q_1} \rceil + 1, \dots$. For such n, and $\alpha = \lceil C_1 d^{q_1} \rceil - 1$ we then have

$$\sum_{j=n+1}^{\infty} \lambda_{d,j} \le C_2 d^{q_2} \sum_{j=n+1}^{\infty} \frac{1}{(j-\alpha)^{1/\tau}} \le C_2 d^{q_2} \int_n^{\infty} \frac{dx}{(x-\alpha)^{1/\tau}}$$

$$= \frac{\tau C_2 d^{q_2}}{1-\tau} \frac{1}{(n - \lceil C_1 d^{q_1} \rceil + 1)^{(1-\tau)/\tau}}.$$

Hence, $\sum_{j=n+1}^{\infty} \lambda_{d,j} \le \varepsilon^2$ for

$$n = \left\lceil \left(\frac{\tau C_2 d^{q_2}}{(1-\tau)\varepsilon^2} \right)^{\tau/(1-\tau)} \right\rceil + \lceil C_1 d^{q_1} \rceil - 1.$$

This proves that

$$n(\varepsilon, d) \le \left(C_1 + \left(\frac{\tau C_2}{1-\tau} \right)^{\tau/(1-\tau)} + 1 \right) d^{\max(q_1, q_2\tau/(1-\tau))} \varepsilon^{-2\tau/(1-\tau)}.$$

Thus, S is polynomially tractable, proving the second point of the theorem.

Strong polynomial tractability of S is proven similarly by taking $q = 0$ in the first part of the proof, and $q_1 = q_2 = 0$ in the second part. The formula for the exponent of strong polynomial tractability follows from the second and third points of the theorem. This completes the proof. $\qquad \square$

It is interesting to compare Theorems 5.1 and 6.1, which present necessary and sufficient conditions for polynomial tractability in the worst and average case settings. First of all we stress that they address polynomial tractability of two different and, in general, unrelated multivariate problems. The worst case is defined for the unit ball of a Hilbert space H_d whereas the average case setting is defined for a whole Banach space F_d. The spaces H_d and F_d are, in general, not related. Therefore the eigenvalues $\lambda_{d,j}$ for the worst case may be quite different than the eigenvalues $\lambda_{d,j}$ for the average case setting. Following the usual notation, we use the same symbol for both sets of eigenvalues although we admit that this may sometimes be confusing.

The polynomial tractability conditions are quite similar for the two settings, since we have a similar dependence of the information complexity on the sequence of the eigenvalues; in the worst case the information complexity is the smallest n for which the $(n + 1)$st largest eigenvalue is at most ε^2, whereas in the average case the information complexity is the smallest n for which the truncated trace starting from the $(n + 1)$st largest eigenvalue is at most ε^2. Hence in both settings, the first polynomially many eigenvalues do not effect polynomial tractability, whereas the sum of the rest of their positive powers must be uniformly bounded in d. The only difference is that the exponent τ in the worst case setting can be arbitrarily large, whereas in the average case setting τ must be smaller than one. Also observe the difference between the exponents of strong polynomial tractability. In the worst case setting, this exponent is 2τ and in the average case setting it is $2\tau/(1 - \tau)$, choosing the smallest possible τ satisfying the condition (5.1) or (6.1), respectively.

As in Chapter 5, it is easy to check that (6.1) holds iff there exist non-negative C_1, C_2, q_1, q_2 and $r > 1$ such that

$$\lambda_{d,n} \leq C_2 d^{q_2} \big(n - \lceil C_1 d^{q_1} \rceil + 1\big)^{-r} \quad \text{for all } n \geq \lceil C_1 d^{q_1} \rceil. \tag{6.4}$$

Indeed, if (6.1) holds then (6.3) implies (6.4) with the same C_1, C_2, q_1, q_2 and with $r = 1/\tau > 1$. On the other hand, if (6.4) holds then for any $\tau \in (1/r, 1)$ we have

$$\Big(\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} \Big)^{1/\tau} \leq C_2 d^{q_2} \zeta(\tau r)^{1/\tau},$$

and (6.1) holds.

As an example observe that for $\lambda_{d,j} = e^{\alpha\sqrt{d}} j^{-\beta}$, with $\beta > 1$, we do *not* have polynomial tractability for $\alpha > 0$, whereas we have strong polynomial tractability for $\alpha \leq 0$, in which case the exponent of strong polynomial tractability is $2/(\beta - 1)$.

We turn to the normalized error criterion for the same class Λ^{all}. We now have

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

Theorem 6.2. *Consider the non-zero problem $S = \{S_d\}$ for a linear S_d defined from a separable Banach space F_d into a Hilbert space G_d. We study the problem S for the normalized error criterion in the average case setting and for the class Λ^{all}.*

- *S is polynomially tractable iff there exist $q_2 \geq 0$ and $\tau \in (0,1)$ such that*

$$C_2 := \sup_d \frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty} \lambda_{d,j}} d^{-q_2} \; < \; \infty. \tag{6.5}$$

- *If* (6.5) *holds then*

$$n(\varepsilon, d) \leq \left(\left(\frac{\tau C_2}{1-\tau}\right)^{\tau/(1-\tau)} + 1\right) d^{q_2\tau/(1-\tau)} \varepsilon^{-2\tau/(1-\tau)}$$

for all $\varepsilon \in (0,1]$ and $d = 1,2,\ldots$.

- *If S is polynomially tractable, so that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some positive C and p, and $q \geq 0$, then* (6.5) *holds for $\tau \in ((1+2/p)^{-1}, 1)$ with $q_2 = q\max(1/\tau, 2/p)$ and*

$$C_2 \leq \left(2C + 1 + \left(2(4C)^{2/p}\zeta(\tau(1+2/p))^{1/\tau}\right)^{\tau}\right)^{1/\tau}.$$

- *S is strongly polynomially tractable iff* (6.5) *holds with $q_2 = 0$. The exponent of strong polynomial tractability is*

$$p^{\text{str-wor}} = \inf\left\{\frac{2\tau}{1-\tau} \mid \tau \text{ satisfies (6.5) with } q_2 = 0\right\}.$$

Proof. Since the proof is similar to the previous one, we only sketch the differences between them. Assuming that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ we now know that

$$\sum_{j=\lfloor C d^q \varepsilon^{-p}\rfloor + 1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j}.$$

From (6.2) and for $\tau \in ((1+2/p)^{-1}, 1)$ we obtain

$$\left(\sum_{j=\lceil (2C+1)d^q\rceil}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau} \leq 2(4Cd^q)^{2/p}\zeta\left(\tau(1+2/p)\right)^{1/\tau} \sum_{j=1}^{\infty} \lambda_{d,j}.$$

This yields

$$\left(\frac{\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}}{\left(\sum_{j=1}^{\infty} \lambda_{d,j}\right)^{\tau}}\right)^{1/\tau} \leq \left(\frac{\lambda_{d,1}^{\tau}(\lceil(2C+1)d^q\rceil - 1) + \sum_{j=\lceil(2C+1)d^q\rceil}^{\infty} \lambda_{d,j}^{\tau}}{\left(\sum_{j=1}^{\infty} \lambda_{d,j}\right)^{\tau}}\right)^{1/\tau}$$

$$\leq \left((2C+1)d^q + \left(2(4Cd^q)^{2/p}\zeta(\tau(1+2/p))^{1/\tau}\right)^{\tau}\right)^{1/\tau}$$

$$\leq C_2 d^{q\max(1/\tau,2/p)},$$

where $C_2 = \left(2C + 1 + \left(2(4C)^{2/p}\zeta(\tau(1+2/p))^{1/\tau}\right)^{\tau}\right)^{1/\tau}$. This proves (6.5), as well as the third point of the theorem.

Assuming (6.5), we conclude that

$$n^{1/\tau}\lambda_{d,n} \leq \Big(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \Big)^{1/\tau} \leq C_2 d^{q_2} \sum_{j=1}^{\infty} \lambda_{d,j} \quad \text{for all } n = 1, 2, \dots.$$

Then

$$\sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \frac{\tau C_2 d^{q_2}}{1-\tau} \frac{1}{n^{(1-\tau)/\tau}} \sum_{j=1}^{\infty} \lambda_{d,j}.$$

Hence, $\sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j}$ holds for

$$n \leq \Big(\Big(\frac{\tau C_2 d^{q_2}}{1-\tau} \Big)^{\tau/(1-\tau)} + 1 \Big) \varepsilon^{-2\tau/(1-\tau)}.$$

This proves polynomial tractability of S and the second point of the theorem. Strong polynomial tractability follows as before. □

As in the worst case setting, the main difference between Theorem 6.1 and Theorem 6.2 is that for the absolute error criterion, polynomially many largest eigenvalues of $\{\lambda_{d,j}\}$ do not count, whereas for the normalized error criterion, the whole sequence of normalized eigenvalues $\{\lambda_{d,j} / \sum_{k=1}^{\infty} \lambda_{d,k}\}$ counts. The reason is that although polynomially many initial eigenvalues can be arbitrarily large, for the normalized error criterion we consider the ratios $\lambda_{d,j} / \sum_{k=1}^{\infty} \lambda_{d,k}$, which are always at most one. Hence, for the normalized error criterion there is no need to drop the initial polynomial part of the sequence, which was necessary for the absolute error criterion.

As in the worst case setting, it is natural to ask whether polynomial tractabilities for the absolute and normalized error criteria in the average case setting are related. It is easy to see that they are *not*. That is, it may happen that we have polynomial tractability for the absolute error criterion but not for the normalized error criterion or vice versa. Indeed, similarly to Chapter 5, consider the eigenvalues $\{\lambda_{d,j}\}$ such that

$$\{\lambda_{d,j}\} = \big\{ (j_1 + \alpha)^{-\beta} (j_2 + \alpha)^{-\beta} \cdots (j_d + \alpha)^{-\beta} \big\}_{j_1, j_2, \dots, j_d = 1}^{\infty} \tag{6.6}$$

for some $\alpha \geq 0$ and $\beta > 1$. Proceeding exactly as in Chapter 5, it is easy to check that for sufficiently large α, we have strong polynomial tractability for the absolute error criterion. More precisely, this holds for α so large that the positive τ^* defined by the condition

$$\sum_{j=1}^{\infty} (j + \alpha)^{-\tau^*} = 1$$

satisfies $\tau^* < \beta$. Then the exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \frac{2\tau^*}{\beta - \tau^*}.$$

For the normalized error criterion with the same sequence of eigenvalues (6.6), we have for large α and $\tau \in (\beta^{-1}, 1)$,

$$\frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty} \lambda_{d,j}} = \frac{\left(\sum_{j=1}^{\infty}(j+\alpha)^{-\beta\tau}\right)^{d/\tau}}{\left(\sum_{j=1}^{\infty}(j+\alpha)^{-\beta}\right)^d} = \left[\frac{\beta-1}{(\beta\tau-1)^{1/\tau}}\alpha^{(1-\tau)/\tau}(1+o(1))\right]^d.$$

It can be checked that it goes exponentially fast to infinity with d for large α and $\beta > 1$. Due to Theorem 6.1, this implies that the problem S is *polynomially intractable* for large positive α and $\beta > 1$.

Hence for large positive α, we have polynomial tractability for the absolute error criterion and polynomial intractability for the normalized error criterion.

The opposite case of polynomial intractability for the absolute error criterion and polynomial tractability for the normalized error criterion in the average case setting can be obtained for the sequence of eigenvalues that we considered before, namely for $\lambda_{d,j} = e^{\alpha\sqrt{d}} j^{-\beta}$ for positive α and $\beta > 1$. The lack of polynomial tractability for the absolute error criterion was previously discussed, whereas for the normalized error criterion we have

$$\frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty} \lambda_{d,j}} = \frac{\zeta(\beta\tau)^{1/\tau}}{\zeta(\beta)},$$

so that we even obtain strong polynomial tractability with the exponent $2/(\beta-1)$.

We now analyze weak tractability of S. As in the worst case setting, the conditions for weak tractability can be presented simultaneously for both the absolute and normalized error criteria by defining

$$\mathrm{CRI}_d = 1 \qquad \text{for the absolute error criterion,}$$

$$\mathrm{CRI}_d = \sum_{j=1}^{\infty} \lambda_{d,j} \quad \text{for the normalized error criterion.}$$

We are ready to prove the following theorem.

Theorem 6.3. *Consider the non-zero problem $S = \{S_d\}$ for a linear S_d defined from a separable Banach space F_d into a Hilbert space G_d. We study the problem S for the absolute or normalized error criterion in the average case setting and for the class Λ^{all}. Let*

$$t_{d,j} := \sum_{k=j}^{\infty} \lambda_{d,j}.$$

S is weakly tractable iff

- *we have*

$$\lim_{j \to \infty} \frac{t_{d,j}}{\mathrm{CRI}_d} \ln^2 j = 0 \quad \text{for all } d, \text{ and}$$

- *there exists a function* $f : (0, \frac{1}{2}] \to \mathbb{N}_+$ *such that*

$$M := \sup_{\beta \in (0, \frac{1}{2}]} \frac{1}{\beta^2} \sup_{d \geq f(\beta)} \sup_{j \geq \lceil \exp(d\sqrt{\beta}) \rceil + 1} \frac{t_{d,j}}{\mathrm{CRI}_d} \ln^2 j \; < \; \infty.$$

Proof. Observe that in the average case setting we have $\sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \mathrm{CRI}_d$, which is equivalent to $t_{d,n+1} \leq \varepsilon^2 \mathrm{CRI}_d$. The condition on $t_{d,j}$ is exactly what we studied in the worst case setting, and therefore Theorem 5.3 applies and directly implies Theorem 6.3. $\qquad\qquad\square$

We now translate the conditions on $t_{d,j}$ into conditions on the specific eigenvalues $\lambda_{d,j}$ that are given by

$$\frac{\lambda_{d,j}}{\mathrm{CRI}_d} = \Theta\left(\frac{C_d}{j^{p_1} \ln^{p_2}(j+1)} \right) \quad \text{for all } j, d \in \mathbb{N}, \tag{6.7}$$

where the factors in the big Θ notation are independent of j and d. Since $\lambda_{d,j}$ must be summable we need to assume that $p_1 > 1$ or that $p_1 = 1$ and $p_2 > 1$. Based on Theorem 6.3 it is easy to obtain the following corollary.

Corollary 6.4. *Consider the non-zero problem* $S = \{S_d\}$ *for a linear* S_d *defined from a separable Banach space* F_d *into a Hilbert space* G_d *with the eigenvalues* $\lambda_{d,j}$ *satisfying* (6.7). *We study the problem S for the absolute or normalized error criterion in the average case setting and for the class* Λ^{all}.

- *Let* $p_1 > 1$. *S is weakly tractable iff* $C_d = \exp(o(d))$.

- *Let* $p_1 = 1$. *S is weakly tractable iff* $p_2 > 3$ *and* $C_d = o(d^{p_2 - 3})$.

Proof. Observe that

$$\sum_{k=j}^{\infty} \lambda_{d,j} = \Theta\left(C_d \int_j^{\infty} \frac{\mathrm{d}x}{x^{p_1} \ln^{p_2}(x)} \right) \mathrm{CRI}_d.$$

Assume first that $p_1 > 1$. Modulo a power of the logarithm, the last sum is of order $C_d j^{-(p_1-1)}$. This leads to

$$\ln n(\varepsilon, d) = \Theta\left(\ln C_d + \ln \varepsilon^{-1} \right).$$

Then $\lim_{\varepsilon^{-1}+d \to \infty} \ln n(\varepsilon, d)/(\varepsilon^{-1} + d) = 0$ iff $\ln C_d = o(d)$, as claimed.

Assume now that $p_1 = 1$. Then the last sum is of order $C_d \ln^{-(p_2-1)} j$ and

$$\ln n(\varepsilon, d) = \Theta\left(C_d^{1/(p_2-1)} \varepsilon^{-2/(p_2-1)} \right).$$

Suppose that we have weak tractability, so that

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} = 0.$$

Taking $\varepsilon^{-1} \to \infty$ and a fixed d we see that $2/(p_2 - 1) < 1$, i.e., that $p_2 > 3$. On the other hand, if $\varepsilon^{-1} = d$ and $d \to \infty$ then $C_d = o(d^{p_2-3})$. Finally, if $p_2 > 3$ and $C_d = o(d^{p_2-3})$, we let $z = \max(\varepsilon^{-1}, d)$ to find

$$\ln n(\varepsilon, d) = o\big(z^{(p_2-3)/(p_2-1)} z^{2/(p_2-1)}\big) = o(z),$$

which implies weak tractability, and completes the proof.   □

The case of $p_1 > 1$ in Corollary 6.4 may even yield polynomial tractability, whereas the case $p_1 = 1$ contradicts polynomial tractability since $\sum_{k=j}^\infty \lambda_{d,j}^\tau = \infty$ for all j and $\tau \in (0, 1)$, see Theorems 6.1 and 6.2. Therefore the case $p_1 = 1$ is more interesting when we consider weak tractability. We still can get weak tractability iff $p_2 > 3$ and $C_d = o(d^{p_2-3})$. This is the case when the eigenvalues $\lambda_{d,j}$ decay with j as $j^{-1} \ln^{-p_2} j$, and depend at most polynomially on d with the exponent at most $p_2 - 3$ as a function of d.

We now discuss conditions needed for weak tractability. We might intuitively hope that as long as the eigenvalues $\lambda_{d,j}$ depend sub-exponentially on d then weak tractability holds. This hope was confirmed in the worst case setting, where in Chapter 5 we studied the eigenvalues $\lambda_{d,j} = \exp(\alpha_1 d^{\alpha_2}) j^{-\alpha_3}$ for positive α_i and weak tractability holds for $\alpha_2 < 1$. In the average case setting, the situation is different since weak tractability depends on the truncated sums of eigenvalues which may behave differently than the eigenvalues alone. Indeed, Corollary 6.4 states that as long as the eigenvalues behave polynomially, roughly like j^{-p_1} with $p_1 > 1$ then the truncated sums from the jth largest eigenvalue also behave polynomially, roughly as $j^{-(p_1-1)}$, and weak tractability holds with $C_d = \exp(o(d))$ depending sub-exponentially on d. The case $p_1 = 1$ is different, since now the truncated sums from the jth largest eigenvalues depend on a power of the logarithm of j and weak tractability allows C_d to be at most polynomially dependent on d.

As in the worst case setting, it may be difficult to check the second condition of Theorem 6.3 to establish weak tractability. We now generalize Lemma 5.4 from the worst case to the average case setting.

Lemma 6.5. *Consider the non-zero problem $S = \{S_d\}$ for a linear S_d defined from a separable Banach space F_d into a Hilbert space G_d. We study the problem S for the absolute or normalized error criterion in the average case setting and for the class Λ^{all}.*

If there exists a positive $\tau \in (0, 1)$ such that

$$\lim_{d \to \infty} \frac{\ln \big(\big(\sum_{j=1}^\infty \lambda_{d,j}^\tau\big)^{1/\tau} / \mathrm{CRI}_d\big)}{d} = 0 \tag{6.8}$$

then S is weakly tractable.

Proof. Since $n^{1/\tau} \lambda_{d,n} \le \big(\sum_{j=1}^\infty \lambda_{d,j}^\tau\big)^{1/\tau}$ then

$$\sum_{j=n+1}^\infty \lambda_{d,j} \le \frac{\tau}{1-\tau} \Big(\sum_{j=1}^\infty \lambda_{d,j}^\tau\Big)^{1/\tau} \frac{1}{n^{(1-\tau)/\tau}}.$$

This yields

$$n(\varepsilon, d) \leq \left\lceil \left[\frac{\tau}{1-\tau} \frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau}}{\mathrm{CRI_d}} \right]^{\tau/(1-\tau)} \varepsilon^{-2\tau/(1-\tau)} \right\rceil.$$

Then the assumption (6.8) easily leads to

$$\lim_{\varepsilon^{-1}+d \to \infty} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1}+d} = 0.$$

Hence, S is weakly tractable.   □

Clearly, (6.8) is only a sufficient condition for weak tractability. For example, take

$$\lambda_{d,j} = \frac{1}{j \ln^4(j+1)} \quad \text{for all } j, d \in \mathbb{N}.$$

Then $\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \right)^{1/\tau} = \infty$ for any $\tau \in (0,1)$, and Lemma 6.5 does not apply, however, the problem is weakly tractable.

Example: Trade-offs of the Exponents (Continued)

We return to trade-offs between the exponents of d and ε^{-1} for polynomial tractability, this time in the average case setting. For simplicity we only consider the normalized error criterion. As in the worst case setting, we study the following sequence of the eigenvalues

$$\lambda_{d,j} = \prod_{k=1}^{g(d)} j_k^{-\alpha} \quad \text{for all } j \in \mathbb{N}^d,$$

this time for $\alpha > 1$ so that $\lambda_{d,j}$ are summable. As before, $g : \mathbb{N} \to \mathbb{N}$ and $g(d) \leq d$.

It is easy to check from Theorem 6.2 that strong polynomial tractability holds iff $\limsup_{d \to \infty} g(d) < \infty$, and polynomial tractability iff

$$A_g := \limsup_{d \to \infty} \frac{g(d)}{\ln d} < \infty.$$

Furthermore, if $A_g = 0$ then the exponent of d can be arbitrarily small.

Assume now that $g(d) = \min(d, \lceil \ln(d+1) \rceil)$. Then it is easy to check that

$$n(\varepsilon, d) = \mathcal{O}\left(d^{q_\tau} \varepsilon^{-p_\tau} \right),$$

where $\tau \in (1/\alpha, 1)$ and

$$q_\tau = \frac{\tau}{1-\tau} \ln \frac{\zeta(\alpha\tau)^{1/\tau}}{\zeta(\alpha)} \quad \text{and} \quad p_\tau = \frac{2\tau}{1-\tau}.$$

We first show that these exponents are optimal in the following sense. Indeed, from the case $d = 1$ we conclude that the exponent of ε^{-1} is at least $2/(\alpha - 1)$, and from the case $d = 2$ that it must be larger than $2/(\alpha - 1)$. Note that for $\tau \in (1/\alpha, 1)$ we have $2\tau/(1 - \tau) > 2/(\alpha - 1)$ and if τ tends to $1/\alpha$ then $2\tau/(1 - \tau)$ approaches $2/(\alpha - 1)$. If we fix the exponent $2\tau/(1 - \tau)$ of ε^{-1} then we claim that the exponent of d must be at least q_τ. Indeed, $n(\varepsilon, d) = \mathcal{O}(d^q \varepsilon^{-2\tau/(1-\tau)})$ implies that

$$\sum_{j=n+1}^{\infty} \lambda_{d,j} = \mathcal{O}\big(d^{q(1-\tau)/\tau + \ln \zeta(\alpha)} n^{-(1-\tau)/\tau}\big),$$

$$\lambda_{d,j} = \mathcal{O}\big(d^{q(1-\tau)/\tau + \zeta(\alpha)} j^{-1/\tau}\big).$$

Therefore for $\eta \in (\tau, 1)$ we have

$$\frac{\left(\sum_{j=1}^{\infty} \lambda_{d,j}^{\eta}\right)^{1/\eta}}{\sum_{j=1}^{\infty} \lambda_{d,j}} = \mathcal{O}\big(d^{q(1-\tau)/\tau} \zeta(\eta/\tau)\big).$$

On the other hand, the left hand side is of order $d^{\ln \zeta(\alpha\eta)^{1/\eta}/\zeta(\alpha)}$. Letting η go to τ we conclude that

$$q \geq \frac{\tau}{1 - \tau} \ln \frac{\zeta(\alpha\tau)^{1/\tau}}{\zeta(\alpha)},$$

as claimed.

If we now vary $\tau \in (1/\alpha, 1)$ then the exponents of d and ε^{-1} also vary. For τ going to $1/\alpha$ we minimize the exponent of ε^{-1} at the expense of the exponent of d which goes to infinity. On the other hand, if τ goes to 1 then the exponent of d goes to

$$-\frac{\alpha \zeta'(\alpha)}{\zeta(\alpha)} + \ln \zeta(\alpha)$$

at the expense of the exponent of ε^{-1} which goes to infinity.

If we assume that $d = d_\varepsilon = \varepsilon^{-\beta}$ then

$$n(\varepsilon, d_\varepsilon) = \mathcal{O}\left(\varepsilon^{-\left(\frac{2\tau}{1-\tau} + \frac{\tau\beta}{1-\tau} \ln \frac{\zeta(\alpha\tau)^{1/\tau}}{\zeta(\alpha)}\right)}\right).$$

We now can choose τ to minimize the exponent. Then τ is the solution of

$$2 + \beta\left[(1 - \tau)\left(\frac{\alpha \zeta'(\alpha\tau)}{\zeta(\alpha\tau)} - \ln \zeta(\alpha)\right) + \ln \frac{\zeta(\alpha\tau)}{\zeta(\alpha)^\tau}\right] = 0.$$

For $\alpha = \beta = 2$ we obtain $\tau = 0.846587$, whereas for $\alpha = 2$ and $\beta = 1$ we obtain $\tau = 0.778559$.

Example: Schrödinger Equation (Continued)

The linear Schrödinger equation was defined in Chapter 5 with the Hilbert space H_d given by (5.10) as its domain. Let

$$\tilde{\eta}_{d,j} = \beta_{d,j} \eta_{d,j} \quad \text{for } j \in \mathbb{N}^d, \tag{6.9}$$

where $(\beta_{d,j}, \eta_{d,j})$'s are the eigenpairs of the compact operator $(-\Delta + q_0)^{-1}$. Then $\{\tilde{\eta}_{d,j}\}$ is an orthonormal basis of H_d. We equip the space H_d with a zero-mean Gaussian measure μ_d such that its covariance operator C_{μ_d} is given by

$$C_{\mu_d} \tilde{\eta}_{d,j} = \alpha_{d,j} \tilde{\eta}_{d,j} \quad \text{for } j \in \mathbb{N}^d.$$

Here, the eigenvalues $\alpha_{d,j}$ are positive and $\text{trace}(C_{\mu_d}) = \sum_{j \in \mathbb{N}^d} \alpha_{d,j} < \infty$.

As explained in Chapter 5, we approximate $S_d f$ by an algorithm of the form $S_d A_{n,d} f$, where $A_{n,d}$ is an algorithm for approximating f from H_d. Due to isometry of S_d, we have

$$\int_{H_d} \|S_d f - S_d A_{n,d} f\|_{L_2}^2 \mu(\mathrm{d}f) = \int_{H_d} \|f - A_{n,d} f\|_{L_2}^2 \mu(\mathrm{d}f),$$

where $L_2 = L_2([0,1]^d)$.

Hence, also in the average case setting, the linear Schrödinger problem reduces to multivariate approximation, $\text{APP}_d f = f$ for $f \in H_d$.

Let $\nu_d = \mu_d \text{APP}_d^{-1}$. Then ν_d is a zero-mean Gaussian measure on L_2. To find its covariance operator C_{ν_d}, note that $\langle f, \tilde{\eta}_{d,j} \rangle_{H_d} = \beta_{d,j}^{-1} \langle f, \eta_{d,j} \rangle_{L_2}$. Therefore

$$\begin{aligned}
\alpha_{d,j} \delta_{j,k} &= \int_{H_d} \langle f, \tilde{\eta}_{d,j} \rangle_{H_d} \langle f, \tilde{\eta}_{d,k} \rangle_{H_d} \mu_d(\mathrm{d}f) \\
&= \frac{1}{\beta_{d,j} \beta_{d,k}} \int_{H_d} \langle f, \eta_{d,j} \rangle_{L_2} \langle f, \eta_{d,k} \rangle_{L_2} \mu_d(\mathrm{d}f) \\
&= \frac{1}{\beta_{d,j} \beta_{d,k}} \int_{H_d} \langle f, \eta_{d,j} \rangle_{L_2} \langle f, \eta_{d,k} \rangle_{L_2} \nu_d(\mathrm{d}f) = \frac{\langle C_{\nu_d} \eta_{d,j}, \eta_{d,k} \rangle_{L_2}}{\beta_{d,j} \beta_{d,k}}.
\end{aligned}$$

Hence, $C_{\nu_d} \eta_{d,j} = \alpha_{d,j} \beta_{d,j}^2 \eta_{d,j}$ for $j \in \mathbb{N}^d$, and the eigenvalues of C_{ν_d} are

$$\lambda_{d,j} = \alpha_{d,j} \beta_{d,j}^2 \quad \text{for } j \in \mathbb{N}^d.$$

Formally we can now apply the results of this section for the sequence $\{\lambda_{d,j}\}_{j \in \mathbb{N}^d}$, although, in general, it is difficult to check the summability conditions for an arbitrary sequence of $\alpha_{d,j}$.

To simplify further calculations, we take

$$\alpha_{d,j} = \left(q_0 + \sum_{k=1}^d \pi^2 j_k^2 \right)^2 \left(\frac{\delta_{j(1)}}{j_1^\alpha} + \frac{\delta_{j(2)}}{j_2^\alpha} + \cdots + \frac{\delta_{j(d)}}{j_d^\alpha} \right),$$

where $\delta_{j(k)} = 0$ if there exists an index $i \neq k$ such that $j_i > 2$, and $\delta_{j(k)} = 1$ otherwise. That is, for $d = 1$, we have

$$\alpha_{d,j} = \frac{(q_0 + \pi^2 j^2)^2}{j^\alpha},$$

and for $d > 1$, we have $\alpha_{d,j} = 0$ iff j has at least two components larger than 1. To guarantee that $\sum_{j \in \mathbb{N}^d} \alpha_{d,j} < \infty$ we need to assume that $\alpha > 5$. Then

$$\lambda_{d,j} = \frac{\delta_{j(1)}}{j_1^\alpha} + \frac{\delta_{j(2)}}{j_2^\alpha} + \cdots + \frac{\delta_{j(d)}}{j_d^\alpha}.$$

We first check tractability of the linear Schrödinger equation for the absolute error criterion. Note that for $\tau \in (1/\alpha, 1)$ we have

$$\left(\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau \right)^{1/\tau} = d^{1/\tau} \left(\zeta(\tau\alpha) - 1 + d^{-1} \right)^{1/\tau}$$

with the Riemann zeta function ζ.

Due to Theorem 6.1, we see that strong polynomial tractability does *not* hold, whereas polynomial tractability does with, e.g., $q_1 = 0$, $q_2 = 1/\tau$ and $\tau \in (1/\alpha, 1)$. We then have

$$n(\varepsilon, d) = \mathcal{O}\left(d^{1/(1-\tau)} \varepsilon^{2\tau/(1-\tau)} \right)$$

with the factor in the big $\mathcal{O}$ notation independent of ε^{-1} and d. Note that the exponents of d and ε^{-1} can be both arbitrarily close to $\alpha/(\alpha - 1) < 1.25$ and $2/(\alpha - 1) < 0.5$, respectively.

We now turn to the normalized error criterion. We now have

$$\frac{\left(\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau \right)^{1/\tau}}{\sum_{j \in \mathbb{N}^d} \lambda_{d,j}} = d^{(1-\tau)/\tau} \frac{\zeta(\alpha\tau) - 1 + d^{-1}}{\zeta(\alpha) - 1 + d^{-1}}.$$

Due to Theorem 6.2, strong polynomial tractable does not hold, whereas polynomial tractability does with $q_2 = (1 - \tau)/\tau$ and $\tau \in (1/\alpha, 1)$. We then have

$$n(\varepsilon, d) = \mathcal{O}\left(d \varepsilon^{2\tau/(1-\tau)} \right)$$

with the factor in the big $\mathcal{O}$ notation independent of ε^{-1} and d. Note that the exponent of d is now better than for the absolute error criterion, whereas the exponent of ε^{-1} can be arbitrarily close to $2/(\alpha - 1) < 0.5$, as for the absolute error criterion.

We summarize the analysis of this example. The linear Schrödinger problem in the average case setting with the zero-mean Gaussian measure considered here is

- not strongly polynomially tractable for the absolute and normalized error criteria,

- polynomially tractable for the absolute and normalized error criteria.

We remind the reader that the linear Schrödinger problem for the space H_d is intractable in the worst case setting. Hence, intractability in the worst case setting is broken by switching to the average case setting for both error criteria.

6.2 Linear Tensor Product Problems

So far, we have studied linear operators S_d without assuming any relations between them. In this section, similarly to Section 5.2 of Chapter 5 we define all the S_d in terms of a univariate problem. In Chapter 5, where we considered the worst case setting, we defined S_d as tensor product problems generated by the univariate case. In this chapter we consider the average case setting, for which we can relax the assumptions of Section 5.2. It is enough to assume that the Hilbert space G_d is a tensor product of d copies of an infinite dimensional separable Hilbert space G, i.e., $G_d = \otimes_{k=1}^d G$. This means that G_d is a separable Hilbert space spanned by $\otimes_{k=1}^d g_k$ for $g_k \in G$, and the inner product in G_d is defined such that

$$\big\langle \otimes_{k=1}^d g_k, \otimes_{k=1}^d h_k \big\rangle_{G_d} = \prod_{k=1}^d \langle g_k, h_k \rangle_G \quad \text{for } g_k, f_k \in G.$$

Let $\{\eta_i\}$ be a complete orthonormal system of G. Then for $d \geq 1$ and a multi-index $j = [j_1, j_2, \ldots, j_d]$ with $j_i \geq 1$, the system $\{\eta_{d,j}\}$ with

$$\eta_{d,j} = \otimes_{k=1}^d \eta_{j_k} \tag{6.10}$$

is a complete orthonormal system of G_d and

$$S_d f = \sum_{j \in \mathbb{N}^d} \big\langle S_d f, \eta_{d,j} \big\rangle_{G_d} \eta_{d,j} \quad \text{for } f \in F_d,$$

where, as before, F_d is a separable Banach space.

For $f \in F_d$ define $L_j(f) = \big\langle S_d f, \eta_{d,j} \big\rangle_{G_d}$. Since μ_d is a zero-mean Gaussian measure we know that

$$\int_{F_d} L_j(f) \mu_d(\mathrm{d}f) = \int_{G_d} \big\langle g, \eta_{d,j} \big\rangle_{G_d} \nu_d(\mathrm{d}g) = 0 \quad \text{for all } j \in \mathbb{N}^d.$$

We assume that the linear functionals L_j are orthogonal, i.e., for all $j, i \in \mathbb{N}^d$ we have

$$\int_{F_d} L_j(f) L_i(f) \mu_d(\mathrm{d}f) = \int_{G_d} \big\langle g, \eta_{d,j} \big\rangle_{G_d} \big\langle g, \eta_{d,i} \big\rangle_{G_d} \nu_d(\mathrm{d}g) = \lambda_{d,j} \delta_{j,i}.$$

Hence

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

To preserve the tensor product structure of the space G_d and its orthonormal system $\{\eta_{d,j}\}$, we assume that the eigenvalues $\lambda_{d,j}$ are given as follows. For $d = 1$, we assume that $\lambda_{1,j} = \lambda_j$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and

$$\sum_{j=1}^\infty \lambda_j = \text{trace}(C_{\nu_1}) < \infty.$$

For $d \geq 1$, we assume that

$$\lambda_{d,j} = \prod_{k=1}^{d} \lambda_{j_k} \quad \text{for all } j = [j_1, j_2, \ldots, j_d] \in \mathbb{N}^d. \tag{6.11}$$

Observe that

$$\text{trace}(C_{\nu_d}) = \sum_{j \in \mathbb{N}^d} \lambda_{d,j} = \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d.$$

Similarly to the worst case, we call the multivariate problem $S = \{S_d\}$ with the eigenpairs of the correlation operators C_{ν_d} given by (6.10) and (6.11) a *linear tensor product problem in the average case setting*.

Let us order the sequence of the eigenvalues $\{\lambda_{d,j}\}_{j \in \mathbb{N}^d} = \{\lambda_{d,j}\}_{j \in \mathbb{N}}$ such that $\lambda_{d,1} \geq \lambda_{d,2} \geq \cdots \geq 0$. Clearly, $\lambda_{d,1} = \lambda_1^d$, $\lambda_{d,2} = \lambda_1^{d-1}\lambda_2$ etc. Note that for $\lambda_2 = 0$, the operator S_d is equivalent to a continuous linear functional. Then S_d can be solved exactly with one information operation from the class Λ^{all}, and so the problem S is trivial in this case. In the sequel, we will always assume that $\lambda_2 > 0$, so that we have at least 2^d positive eigenvalues. Indeed, we have $\binom{d}{k}$ positive eigenvalues $\lambda_1^{d-k}\lambda_2^k$ for $k = 0, 1, \ldots, d$. In this case it is meaningful to study tractability even if the rest of the eigenvalues is 0, i.e., $\lambda_j = 0$ for $j \geq 3$. Observe that

$$\sum_{j=n+1}^{\infty} \lambda_{d,j} = \sum_{j=1}^{\infty} \lambda_{d,j} - \sum_{j=1}^{n} \lambda_{d,j} = \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d - \sum_{j=1}^{n} \lambda_{d,j} \geq \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d - n\lambda_1^d.$$

Hence, the information complexity

$$n(\varepsilon, d) := n^{\text{avg}}(\varepsilon, d, \Lambda^{\text{all}}) = \min \Big\{ n \mid \sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \text{CRI}_d \Big\}$$

can be bounded from below by

$$n(\varepsilon, d) \geq \min \Big\{ n \mid \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d - n\lambda_1^d \leq \varepsilon^2 \text{CRI}_d \Big\} \geq \frac{\big(\sum_{j=1}^{\infty} \lambda_j \big)^d - \varepsilon^2 \text{CRI}_d}{\lambda_1^d},$$

where, as always, $\text{CRI}_d = 1$ for the absolute error criterion, and

$$\text{CRI}_d = \sum_{j=1}^{\infty} \lambda_{d,j} = \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d$$

for the normalized error criterion.

We are ready to study polynomial tractability of linear tensor product problems. We will address both the absolute and normalized error criteria.

Theorem 6.6. *Consider the linear tensor product problem in the average case setting $S = \{S_d\}$ with $\lambda_2 > 0$. We study this problem for the class Λ^{all}.*

- *Consider the normalized error criterion. Then S is intractable since*

$$n(\varepsilon, d) \geq (1 - \varepsilon^2) \Big(\sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_1} \Big)^d \geq (1 - \varepsilon^2) \Big(1 + \frac{\lambda_2}{\lambda_1} \Big)^d \quad \text{for all } \varepsilon \in [0, 1)$$

  is exponential in d and we have the curse of dimensionality.

- *Consider the absolute error criterion. Let $\sum_{j=1}^{\infty} \lambda_j \geq 1$. Then S is intractable since*

$$n(\varepsilon, d) \geq \frac{1}{2} \Big(\sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_1} \Big)^d \geq \frac{1}{2} \Big(1 + \frac{\lambda_2}{\lambda_1} \Big)^d \quad \text{for all } \varepsilon \in [0, \sqrt{2}/2]$$

  is exponential in d and we have the curse of dimensionality.

- *Consider the absolute error criterion. Let $\sum_{j=1}^{\infty} \lambda_j < 1$. Then S is polynomially tractable iff S is strongly polynomially tractable iff there is $\tau \in (0, 1)$ such that*

$$\sum_{j=1}^{\infty} \lambda_j^{\tau} \leq 1. \tag{6.12}$$

  If (6.12) holds then

$$n(\varepsilon, d) \leq \frac{1}{1 - \tau} \varepsilon^{-2\tau/(1-\tau)},$$

  and the exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \inf \big\{ \tfrac{2\tau}{1-\tau} \mid \tau \text{ satisfies } (6.12) \big\}.$$

Proof. Consider first the normalized error. For $n = n(\varepsilon, d)$ we have

$$\Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d - \lambda_1^d n \leq \varepsilon^2 \Big(\sum_{j=1}^{\infty} \lambda_j \Big)^d$$

which gives the bound needed.

Consider now the absolute error. Assume first that $a := \sum_{j=1}^{\infty} \lambda_j \geq 1$. Then we have $a^d - \varepsilon^2 \geq \frac{1}{2} a^d$ for $\varepsilon \in [0, \sqrt{2}/2]$ and

$$\varepsilon^2 \geq \sum_{j=n(\varepsilon,d)+1}^{\infty} \lambda_{d,j} \geq a^d - n(\varepsilon, d) \lambda_1^d.$$

This yields

$$n(\varepsilon, d) \geq \frac{a^d - \varepsilon^2}{\lambda_1^d} \geq \frac{a^d}{2\lambda_1^d} = \frac{1}{2} \sum_{j=1}^{\infty} \Big(\frac{\lambda_j}{\lambda_1} \Big)^d \geq \frac{1}{2} \Big(1 + \frac{\lambda_2}{\lambda_1} \Big)^d$$

which is the bound needed.

Assume finally that $\sum_{j=1}^{\infty} \lambda_j < 1$ and consider polynomial tractability of S. We will use Theorem 6.1. Take $\tau \in (0,1)$ and consider

$$\sum_{j=\lceil C_1 d^{q_1} \rceil}^{\infty} \lambda_{d,j}^{\tau} = \Big(\sum_{j=1}^{\infty} \lambda_j^{\tau} \Big)^d - \sum_{j=1}^{\lceil C_1 d^{q_1} \rceil - 1} \lambda_{d,j}^{\tau} \ge \Big(\sum_{j=1}^{\infty} \lambda_j^{\tau} \Big)^d - \lambda_1^d C_1 d^{q_1}$$

$$= \Big(\sum_{j=1}^{\infty} \lambda_j^{\tau} \Big)^d \Big(1 - \Big(\frac{\lambda_1^{\tau}}{\sum_{j=1}^{\infty} \lambda_j^{\tau}} \Big)^d C_1 d^{q_1} \Big).$$

Then $\sum_{j=1}^{\infty} \lambda_j^{\tau} > 1$ implies that C_2 in (6.1) is infinite. If $\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1$ then

$$\sum_{j=1}^{\infty} \lambda_{d,j}^{\tau} \le \Big[\sum_{j=1}^{\infty} \lambda_j^{\tau} \Big]^d \le 1.$$

Hence, we can take $C_1 = 1$, $q_1 = q_2 = 0$ and C_2 in (6.1) is bounded by 1. Hence, polynomial tractability of S is equivalent to strong polynomial tractability of S and holds iff $\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1$. The rest follows from Theorem 6.1. This completes the proof. $\square$

We now comment on Theorem 6.6. It it interesting to notice that we always have intractability for the normalized error criterion. On the other hand, polynomial tractability is equivalent to strong polynomial tractability for the absolute error criterion, holding iff $\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1$ for some positive $\tau < 1$. We now show that for $\lambda_2 > 0$ we have

$$\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1 \text{ for } \tau \in (0,1) \quad \text{iff} \quad \sum_{j=1}^{\infty} \lambda_j < 1 \text{ and } \lambda_j = \mathcal{O}\left(j^{-p}\right) \text{ for } p > 1.$$

Indeed, if $\sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1$ for some $\tau \in (0,1)$, then for any $s \in [\tau,1]$ by using Jensen's inequality[1] twice we have

$$\sum_{j=1}^{\infty} \lambda_j \le f(s) := \Big(\sum_{j=1}^{\infty} \lambda_j^s \Big)^{1/s} \le \Big(\sum_{j=1}^{\infty} \lambda_j^{\tau} \Big)^{1/\tau} \le 1.$$

Hence, if $\sum_{j=1}^{\infty} \lambda_j = 1$ then $f(s) \equiv 1$. This can only happen when $\lambda_j^s = \text{constant}$, which implies that $\lambda_j \in \{0,1\}$. Hence $\lambda_2 = 1$ and $\lambda_1 \ge 1$. This yields that $\sum_{j=1}^{\infty} \lambda_j \ge 2$, which is a contradiction. Thus $\sum_{j=1}^{\infty} \lambda_j < 1$. We also have

$$n \lambda_n^{\tau} \le \sum_{j=1}^{\infty} \lambda_j^{\tau} \le 1,$$

---

[1] Jensen's inequality states that $\sum_{j=1}^{\infty} a_j \le \left(\sum_{j=1}^{\infty} a_j^r \right)^{1/r}$ for $a_j \ge 0$ and $r \in (0,1]$. In our case, it is enough to take $a_j = \lambda_j$ and $r = s$ for the first use of this inequality, and $r = \tau/s$ for the second time.

and therefore $\lambda_n \le n^{-1/\tau}$. Hence $\lambda_j = \mathcal{O}(j^{-p})$ with $p = 1/\tau > 1$.

Assume that $\sum_{j=1}^{\infty} \lambda_j < 1$ and $\lambda_j = \mathcal{O}(j^{-p})$ with $p > 1$. This clearly implies that $\lambda_j < 1$. Consider the function $g(s) = f^s(s)$ for $s \in (1/p, 1]$. The function g is well defined, since $\lambda_j^s = \mathcal{O}(j^{-sp})$ and the series $\sum_{j=1}^{\infty} j^{-sp}$ is convergent for $sp > 1$. Furthermore,

$$g'(s) = -\sum_{j=1}^{\infty} \lambda_j^s \ln \lambda_j^{-1}$$

is also well defined and negative. Take a positive δ such that $\delta + 1/p < 1$. Then the function g' is bounded over $[\delta + 1/p, 1]$, and therefore for $\tau \in [\delta + 1/p, 1]$ we have

$$g(\tau) = g(1) + \mathcal{O}(1 - \tau).$$

Since $g(1) < 1$ there exists a number $\tau < 1$ for which $g(\tau) \le 1$. This implies $f(\tau) \le 1$, as claimed.

We stress that polynomial tractability holds only if the initial error $\left(\sum_{j=1}^{\infty} \lambda_j\right)^{d/2}$ is exponentially small in d. Then the problem is obviously trivial for all

$$\varepsilon^2 \ge \left(\sum_{j=1}^{\infty} \lambda_j\right)^d$$

since $n(\varepsilon, d) = 0$ for all such ε. Hence we may have polynomial tractability for the absolute error criterion only if the problem is badly scaled.

As in the comment after Theorem 5.5, we cannot in general claim that the exponent of strong polynomial tractability is obtained for τ such that $\sum_{j=1}^{\infty} \lambda_j^\tau = 1$. Indeed, take $\lambda_j = \alpha j^{-2} \ln^{-p_2}(j+1)$. Then $\sum_{j=1}^{\infty} \lambda_j^{1/2} < 1$ for $p_2 > 2$ and small α, whereas $\sum_{j=1}^{\infty} \lambda_j^\tau = \infty$ for all $\tau < 1/2$. This shows that the exponent of strong polynomial tractability is 2, although $\sum_{j=1}^{\infty} \lambda_j^{1/2}$ can be arbitrarily small.

We now turn to weak tractability. Obviously, weak tractability does not hold for the normalized error criterion. Consider then the absolute error criterion. Observe that since $\{\lambda_j\}$ is summable, i.e., $\sum_{j=1}^{\infty} \lambda_j < \infty$, then $\lambda_j = \mathcal{O}(j^{-p})$ for some $p \ge 1$. For $p > 1$, we know that we have strong polynomial tractability iff $\sum_{j=1}^{\infty} \lambda_j < 1$. On the other hand we know that $\sum_{j=1}^{\infty} \lambda_j \ge 1$ implies intractability. Hence, for $p > 1$, weak tractability implies that $\sum_{j=1}^{\infty} \lambda_j < 1$, and is equivalent to strong polynomial tractability. We summarize this observation in the following theorem.

Theorem 6.7. *Consider the linear tensor product problem in the average case setting $S = \{S_d\}$ with $\lambda_2 > 0$ and let $\lambda_j = \mathcal{O}(j^{-p})$ with $p > 1$. We study this problem for the absolute error criterion and the class Λ^{all}. Then the following statements are equivalent:*

- *S is weakly tractable,*

- *S is polynomially tractable,*

- *S is strongly polynomially tractable,*

- $\sum_{j=1}^{\infty} \lambda_j < 1$.

Hence, the only case for which weak tractability may hold, and for which polynomial tractability does not hold, is when $p = 1$. This corresponds, in particular, to the case for which $\lambda_j = \mathcal{O}(j^{-1} \ln^{-p_2}(j + 1))$ with $p_2 > 1$. We leave the question whether weak tractability holds for such sequences as our next open problem.

Open Problem 28.

- Consider the linear tensor product problem in the average case setting $S = \{S_d\}$ with $\sum_{j=1}^{\infty} \lambda_j < 1$ and $\lambda_2 > 0$. Study this problem for the absolute error criterion and for the class Λ^{all}. Verify whether there are eigenvalues λ_j for which we have weak tractability but not polynomial tractability. If so, characterize all such $\{\lambda_j\}$. In particular, characterize the p_2 for which we have weak tractability for

$$\lambda_j = \Theta\left(\frac{1}{j \ln^{p_2}(j + 1)}\right).$$

Example: Approximation for Continuous Functions

We illustrate the results of this section for approximation defined for the space $F_d = C([0, 1]^d)$ of real continuous functions. The Banach space F_d is equipped with a zero-mean Gaussian measure μ_d whose covariance kernel is given by the Korobov kernel. That is,

$$\int_{C([0,1]^d)} f(x) f(y) \mu_d(\mathrm{d}f) = K_{d,\alpha}(x, y) \quad \text{for all } x, y \in [0, 1]^d,$$

where $K_{d,\alpha}$ is given by (A.2) in Appendix A with $\alpha > 1/2$, and it is the reproducing kernel of the Korobov space $H_{d,\alpha}$.

The approximation problem is defined as APP $= \{\text{APP}_d\}$, where

$$\text{APP}_d : F_d = C([0, 1]^d) \to G_d = L_2([0, 1]^d) \quad \text{with APP}_d f = f.$$

The measure $\nu_d = \mu_d \text{APP}_d^{-1}$ is now a zero-mean Gaussian measure with the covariance operator $C_{\nu_d} : G_d \to G_d$ given by

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

We now discuss the eigenpairs $(\lambda_{d,h}, \eta_{d,h})_{h \in \mathbb{Z}^d}$ of C_{ν_d}. The eigenvalues are given by

$$\lambda_{d,h} = \varrho_{d,\alpha}^{-1}(h) = \prod_{k=1}^{d} \lambda_{h_k}$$

with

$$\lambda_{h_k} = \beta_1 \delta_{0,h_k} + \beta_2 (1 - \delta_{0,j_k}) |h_k|^{-2\alpha},$$

with positive β_1 and β_2.

For $h = 0$, we have $\lambda_{d,0} = 1$ and the corresponding eigenfunction is $\eta_{d,0} = 1$. For $h \neq 0$, the eigenvalues $\lambda_{d,h}$ and $\lambda_{d,-h}$ are equal, and the corresponding eigenfunctions $\eta_{d,h}$ and $\eta_{d,-h}$ can be taken as $\cos(2\pi h \cdot x)$ and $\sin(2\pi h \cdot x)$.

Hence, the eigenpairs of C_{ν_d} are the same as the eigenpairs of the operator $W_d = \mathrm{APP}_d \mathrm{APP}_d^* : H_{d,\alpha} \to H_{d,\alpha}$ studied for approximation in the worst case setting for the Korobov space, see Chapter 5. Note that

$$\sum_{h \in \mathbb{Z}} \lambda_h = \beta_1 + 2\beta_2 \zeta(2\alpha).$$

From Theorems 6.6 and 6.7 we conclude that

- APP is intractable for the normalized error criterion.

- Let $\beta_1 + 2\beta_2 \zeta(2\alpha) \geq 1$. Then APP is intractable for the absolute error criterion.

- Let $\beta_1 + 2\beta_2 \zeta(2\alpha) < 1$. Then APP is strongly polynomially tractable for the absolute error criterion. The exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \frac{2\tau}{1 - \tau},$$

where $\tau \in (0, 1)$ is the unique solution of

$$\beta_1^\tau + 2\beta_2^\tau \zeta(2\alpha\tau) = 1.$$

We add that similar results also hold in the worst case setting for approximation defined for the Korobov space $H_{d,\alpha}$. Observe that strong polynomial tractability for both approximation problems holds for the absolute error criterion under the same condition. However, the strong polynomial tractability exponent in the worst case setting is smaller since it is 2τ instead of $2\tau/(1 - \tau)$ which is its counterpart in the average case setting. The difference between them can be arbitrarily large if τ is close to one. In this case, the exponent in the worst case is close to 2, whereas the exponent in the average case approaches infinity.

The reader may be surprised by the last remark that multivariate approximation in the worst case seems easier than in the average case. The point is that we compare two *different* approximation problems defined for two *different* spaces and in two *different* settings. The space $C([0, 1]^d)$ considered in the average case setting is much larger than the Korobov space $H_{d,\alpha}$ considered in the worst case setting. Indeed, $H_{d,\alpha}$ is obviously a subset of $C([0, 1]^d)$ and $\mu_d(H_{d,\alpha}) = 0$ due to the Kolmogorov principle, see Shilov and Fan Dyk Tin [211], it can be also found in [242, p. 308]. Hence, multivariate approximation for the Korobov space $H_{d,\alpha}$ equipped with the Gaussian measure μ_d is trivial in the average case setting. We need a much larger space X, such

as $C([0, 1]^d)$, to guarantee that $\mu_d(X)) = 1$. Of course, it is also of interest to select a different Gaussian measure for which the Korobov space $H_{d,\alpha}$ has measure 1, and then compare the worst and average case tractability results for the same multivariate approximation problem defined over $H_{d,\alpha}$. This is done in our next example.

Example: Approximation for Korobov Space

We consider multivariate approximation for the Korobov space $H_{d,\alpha}$ as it was done in Chapter 5 for the worst case setting. We now consider the average case setting in which we equip $H_{d,\alpha}$ with a zero-mean Gaussian measure μ_d as follows. Recall that $\{e_h\}_{h \in \mathbb{Z}_d}$, defined as in Appendix A, is an orthonormal system of $H_{d,\alpha}$. We choose the covariance operator of C_{μ_d} such that

$$C_{\mu_d} e_h = \alpha_{d,h} e_h \quad \text{for all } h \in \mathbb{Z}^d,$$

for positive $\alpha_{d,h}$ such that $\sum_{h \in \mathbb{Z}^d} \alpha_{d,h} < \infty$. To preserve the tensor product structure, we assume that

$$\alpha_{d,h} = \prod_{j=1}^{d} \alpha_{h_j} \quad \text{with } \sum_{h \in \mathbb{Z}} \alpha_h < \infty$$

for some positive sequence $\{\alpha_h\}_{h \in \mathbb{Z}}$.

Proceeding as for the linear Schrödinger problem in the average case setting, we check that $\nu_d = \mu_d \text{APP}_d^{-1}$ is a zero-mean Gaussian measure defined on $L_2 = L_2([0, 1]^d)$ with the covariance operator C_{ν_d} given by $C_{\nu_d} \eta_{d,h} = \lambda_{d,h} \eta_{d,h}$, where

$$\lambda_{d,h} = \alpha_{d,h} \varrho_{d,\alpha}(h) = \prod_{j=1}^{d} \alpha_{h_j} \left(\beta_1 \delta_{0,h_j} + \beta_2(1 - \delta_{0,h_j})|h_j|^{-2\alpha} \right) \quad \text{for all } h \in \mathbb{Z}^d,$$

and $\eta_{d,h}(x) = \exp(2\pi i h \cdot x)$. To simplify further calculations, take

$$\alpha_h = \eta_1 \delta_{0,h} + \eta_2(1 - \delta_{0,h})|h|^{-2\beta} \quad \text{for all } h \in \mathbb{Z},$$

with positive η_1 and η_2, and $\beta > 1/2$, so that $\sum_{h \in \mathbb{Z}} \alpha_h = \eta_1 + 2\zeta(2\beta) < \infty$. Then

$$\lambda_{d,h} = \prod_{j=1}^{d} \left(\eta_1 \beta_1 \delta_{0,h_j} + \eta_2 \beta_2(1 - \delta_{0,h_j})|h_j|^{-2(\alpha+\beta)} \right) \quad \text{for all } h \in \mathbb{Z}^d.$$

From Theorems 6.6 and 6.7 we conclude that

- APP is intractable for the normalized error criterion.

- Let $\eta_1 \beta_1 + 2\eta_2 \beta_2 \zeta(2(\alpha + \beta)) \geq 1$. Then APP is intractable for the absolute error criterion.

- Let $\eta_1\beta_1 + 2\eta_2\beta_2\zeta(2(\alpha+\beta)) < 1$. Then APP is strongly polynomially tractable for the absolute error criterion. The exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \frac{2\tau}{1-\tau},$$

where $\tau \in (0,1)$ is the unique solution of

$$(\eta_1\beta_1)^\tau + 2(\eta_2\beta_2)^\tau\zeta(2(\alpha+\beta)\tau) = 1. \tag{6.13}$$

We now compare the worst and average case tractability results for multivariate approximation for the absolute error criterion. In the worst case setting, strong polynomial tractability holds iff $\max(\beta_1,\beta_2) < 1$. This may be a weaker condition than $\eta_1\beta_1 + 2\eta_2\beta_2\zeta(2(\alpha+\beta)) < 1$ which is a necessary and sufficient condition in the average case setting.

The reader may think that it sounds like a contradiction since we may have strong tractability in the worst case and intractability in the average case. There is no contradiction since we are still comparing two *different*, although similar multivariate approximation problems. The worst case is defined for the *unit ball* of the space $H_{d,\alpha}$, whereas the average case is considered for the *whole* space $H_{d,\alpha}$. It is known that the average case over the unit ball is essentially the same as for the whole space only if the trace of C_{μ_d} is relatively small, see [286] where this result was originally proved, as well as Section 5.8 of [242] where this result is also reported. In our case,

$$\text{trace}(C_{\mu_d}) = [\eta_1 + 2\eta_2\zeta(2\beta)]^d,$$

and only for $\eta_1 + 2\eta_2\zeta(2\beta) < 1$, the measure of the unit ball is practically 1 for large d. Hence, the last inequality is also needed to make the average case for the unit ball of $H_{d,\alpha}$ roughly the same as the average case for the whole space. If so then the exponents of strong polynomial tractability behave properly, i.e.,

$$p^{\text{str-avg}} = \frac{2\tau^{\text{avg}}}{1-\tau^{\text{av}}} < p^{\text{str-wor}} = 2\tau^{\text{wor}},$$

where τ^{avg} satisfies (6.13), and τ^{wor} satisfies (5.14).

Indeed, this inequality can be checked directly. We use Hölder's inequality for $p = 1 + \tau^{\text{wor}}$ and $q = p/(p-1) = 1 + (\tau^{\text{wor}})^{-1}$, and obtain

$$(\eta_1\beta_1)^{\tau^{\text{wor}}/(1+\tau^{\text{wor}})} + 2(\eta_2\beta_2)^{\tau^{\text{wor}}/(1+\tau^{\text{wor}})}\zeta\left(2(\alpha+\beta)\tau^{\text{wor}}/(1+\tau^{\text{wor}})\right)$$

$$\leq \left(\beta^{\tau^{\text{wor}}} + 2\beta^{\tau^{\text{wor}}}\zeta(2\alpha\tau^{\text{wor}})\right)^{1/p} \left(\eta_1 + 2\eta_2\zeta(2\beta)\right)^{1/q} < 1.$$

Hence, $\tau^{\text{avg}} < \tau^{\text{wor}}/(1 + \tau^{\text{wor}})$, and $p^{\text{str-avg}} < p^{\text{str-wor}}$, as claimed.

We stress that the exponents $p^{\text{str-wor}}$ and $p^{\text{str-avg}}$ can be quite different. Indeed, assume that α and β are approaching $1/2$, and β_1 and β_2 are approaching 1. Then $p^{\text{str-wor}}$ goes to infinity as already discussed in Chapter 5, whereas $p^{\text{str-avg}}$ goes to $2\tau/(1-\tau)$, where τ is the unique solution of

$$\eta_1^\tau + 2\eta_2^\tau\zeta(2\tau) = 1.$$

Example: Schrödinger Equation (Continued)

As before we equip the space H_d given by (5.10) with a zero-mean Gaussian measure whose covariance operator C_{μ_d} has the eigenpairs $(\alpha_{d,j}, \tilde{\eta}_{j,d})$, where $j \in \mathbb{N}^d$, with $\tilde{\eta}_{j,d}$ given by (6.9). To simplify calculations, we now take

$$\alpha_{d,j} = \frac{\left(q_0 + \sum_{k=1}^d \pi^2 j_k^2\right)^2}{(j_1 + 1)^\alpha (j_2 + 1)^\alpha \cdots (j_d + 1)^\alpha}.$$

To guarantee that $\sum_{j \in \mathbb{N}^d} \alpha_{d,j} < \infty$ we need to assume that $\alpha > 5$. Then the eigenvalues of the covariance operator C_{ν_d} are

$$\lambda_{d,j} = \frac{1}{(j_1 + 1)^\alpha (j_2 + 1)^\alpha \cdots (j_d + 1)^\alpha} \quad \text{for all } j \in \mathbb{N}^d.$$

Hence, they are of product form with $\lambda_{1,j} = \lambda_j = (j + 1)^{-\alpha}$ for $j \in \mathbb{N}$.

We first check tractability of the linear Schrödinger equation for the absolute error criterion. Note that for $\tau > 1/\alpha$ we have

$$\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau = \left(\sum_{j \in \mathbb{N}} \lambda_j^\tau\right)^d = (\zeta(\alpha\tau) - 1)^d,$$

with the Riemann zeta function ζ. Define a number α^* such that

$$\zeta(\alpha^*) = 2.$$

Then $\alpha^* = 1.72865\dots$. Clearly, if $\alpha\tau \geq \alpha^*$ then $\zeta(\alpha\tau) - 1 \leq 1$, and $\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau$ is uniformly bounded in d. On the other hand, if $\alpha\tau < \alpha^*$ then $\zeta(\alpha\tau) - 1 > 1$ and $\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau$ is exponentially large in d.

Due to Theorem 6.1, see also Theorem 6.6, the linear Schrödinger equation is strongly polynomially tractable. The exponent of strong polynomial tractability is obtained by taking $\tau = \alpha^*/\alpha$, and is equal to

$$\frac{2\alpha^*}{\alpha - \alpha^*} \leq \frac{2\alpha^*}{5 - \alpha^*} \approx 1.0568\dots.$$

We now turn to the normalized error criterion. We now have

$$\frac{\left(\sum_{j \in \mathbb{N}^d} \lambda_{d,j}^\tau\right)^{1/\tau}}{\sum_{j \in \mathbb{N}^d} \lambda_{d,j}} = \left[\frac{(\zeta(\alpha\tau) - 1)^{1/\tau}}{\zeta(\alpha) - 1}\right]^d.$$

For $\tau > 1/\alpha$, Jensen's inequality implies $(\zeta(\alpha\tau) - 1)^{1/\tau}/(\zeta(\alpha) - 1) > 1$, and we have exponential dependence on d. Due to Theorem 6.2, see also Theorem 6.6, the problem is *not* polynomially tractable.

Next we check weak tractability. Let $\{\lambda_{d,j}\}_{j \in \mathbb{N}}$ be the ordered sequence of $\{\lambda_{d,j}\}_{j \in \mathbb{N}^d}$. We have $\lambda_{d,1} = 2^{-d\alpha}$ and therefore $\sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j}$ implies

$$(\zeta(\alpha) - 1)^d - n2^{-d\alpha} \leq \varepsilon^2 (\zeta(\alpha) - 1)^d.$$

Hence,

$$n(\varepsilon, d) \geq (1 - \varepsilon^2) \left[2^{\alpha} (\zeta(\alpha) - 1) \right]^d.$$

Since $2^{\alpha}(\zeta(\alpha) - 1) > 1 + (2/3)^{\alpha} > 1$, the information complexity is exponential in d and weak tractability cannot hold.

We summarize the analysis of this example. The linear Schrödinger problem in the average case setting with the zero-mean Gaussian measure considered here is

- intractable for the normalized error criterion,

- is strongly polynomially tractable with the exponent at most $1.0568\ldots$ for the absolute error criterion.

Hence, intractability of the linear Schrödinger problem for the space H_d in the worst case setting is now broken by switching to the average case setting but only for the absolute error criterion.

6.3 Linear Weighted Tensor Product Problems

As in Section 5.3 of Chapter 5, we now consider a sequence of weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ for all $d \in \mathbb{N}$ and subsets $\mathfrak{u}$ of $[d] = \{1, 2, \ldots, d\}$. For a linear tensor product problem $S = \{S_d\}$, we assumed in the previous section that the linear functionals $L_j = \langle S_d f, \eta_{d,j} \rangle_{G_d}$ for $j \in \mathbb{N}^d$ were orthogonal and $\eta_{d,j}$'s were the eigenelements of the correlation operator C_{ν_d} with the corresponding eigenvalues

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

For the weighted case, we keep everything as it was before except that the eigenvalues of C_{ν_d} may now depend on the weight sequence γ. More precisely, we assume as before that we have a sequence of $\{\lambda_j\}$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq 0 \quad \text{and} \quad \sum_{j=1}^{\infty} \lambda_j < \infty.$$

Without loss of generality we assume that $\lambda_2 > 0$.

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

$$\lambda_{d,\gamma,j} = \gamma_{d,\mathfrak{u}(j)} \prod_{k=1}^{d} \lambda_{j_k} = \gamma_{d,\mathfrak{u}(j)} \lambda_1^{d - |\mathfrak{u}(j)|} \prod_{k \in \mathfrak{u}(j)} \lambda_{j_k}.$$

We say that $S_\gamma = \{S_{d,\gamma}\}$ is a *linear weighted tensor product problem in the average case setting*, or simply a *weighted* problem, if the $S_{d,\gamma} = S_d$ are defined as in the previous section and the correlation operator C_{ν_d} of the zero-mean Gaussian measure ν_d has the eigenpairs

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

Hence in the average case setting, the only difference between linear tensor and linear weighted tensor product problems is in the distribution of the linear functionals L_j. For the weighted problem, the distribution depends on the weight sequence γ. By considering different weights, we can model different a priori knowledge about the distributions of the solution elements $S_d f$, or more precisely, about the distributions of their inner products $\langle S_d f, \eta_{d,j} \rangle_{G_d}$ with respect to the elements $\eta_{d,j}$. Obviously, if $\gamma_{d,\mathfrak{u}} \equiv 1$ then a linear weighted tensor product problem reduces to the linear tensor product problem studied in the previous section. The weights only matter if they do not all equal to one. For example, for finite-order weights with order ω^*, we know a priori that $\langle S_d f, \eta_{d,j} \rangle_{G_d} = 0$ (with probability 1) for all $j \in \mathbb{N}^d$ with $|\mathfrak{u}(j)| > \omega^*$.

We are ready to study polynomial and weak tractability of the weighted problem $S_\gamma = \{S_{d,\gamma}\}$ for the class Λ^{all}. As with the worst case setting, we leave as an open problem the case of the absolute error criterion.

Open Problem 29.

- Consider the linear weighted tensor product problem in the average case setting $S_\gamma = \{S_{d,\gamma}\}$ with $\lambda_2 > 0$ as defined in this section. Find necessary and sufficient conditions for polynomial and weak tractability of S_γ for the absolute error criterion and for the class Λ^{all}.

We now consider the normalized error criterion. For $\tau \in (0, 1]$, define

$$\alpha_\tau = \sum_{j=2}^{\infty} \left(\frac{\lambda_j}{\lambda_1} \right)^\tau \tag{6.14}$$

with formally $\alpha_\tau = \infty$ if the last series is not convergent. From Jensen's inequality we have $\alpha_1^\tau \leq \alpha_\tau$, and obviously $\alpha_1 < \alpha_\tau$.

The sum-exponent is defined in (5.19). It is easy to see that the sum-exponent p_λ for $\lambda = \{\lambda_j / \lambda_1\}$ is the same as the sum exponent p_{λ^*} for $\lambda^* = \{\lambda_{j+1}/\lambda_2\}$ defined before Theorem 5.7 in Chapter 5, and is now given by

$$p_\lambda = \inf \{\tau \in (0, 1] \mid \alpha_\tau < \infty\}.$$

Note that p_λ is always well defined, since the last series is convergent for $\tau = 1$.

We are ready to present a theorem on polynomial tractability.

Theorem 6.8. *Consider the linear weighted tensor product problem in the average case setting $S_\gamma = \{S_{d,\gamma}\}$ with $\lambda_2 > 0$. We assume that for each d there is at least one non-empty $\mathfrak{u}$ such that $\gamma_{d,\mathfrak{u}} > 0$. We study the problem S_γ for the normalized error criterion and for the class Λ^{all}.*

- S_γ is polynomially tractable iff $p_\lambda < 1$ and there exist $q_2 \geq 0$ and $\tau \in (p_\lambda, 1)$ such that

$$C_2 := \sup_d \frac{\left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^\tau \alpha_\tau^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|}} d^{-q_2} \; < \; \infty. \tag{6.15}$$

- If (6.15) holds then

$$n(\varepsilon, d) \leq \left(\left(\frac{\tau C_2}{1-\tau}\right)^{\tau/(1-\tau)} + 1\right) d^{q_2\tau/(1-\tau)} \varepsilon^{-2\tau/(1-\tau)}$$

for all $\varepsilon \in (0, 1]$ and $d = 1, 2, \ldots$.

- If S_γ is polynomially tractable, so that $n(\varepsilon, d) \leq C d^q \varepsilon^{-p}$ for some $C > 0$, $p > 0$ and $q \geq 0$, then $p_\lambda \leq (1 + 2/p)^{-1}$ and (6.15) holds for $\tau \in ((1 + 2/p)^{-1}, 1)$ with $q_2 = q \max(1/\tau, 2/p)$ and

$$C_2 \leq \left(2C + 1 + \left(2(4C)^{2/p} \zeta(\tau(1 + 2/p))^{1/\tau})^\tau\right)^{1/\tau}.$$

- S_γ is strongly polynomially tractable iff (6.15) holds with $q_2 = 0$. The exponent of strong polynomial tractability is

$$p^{\text{str-avg}} = \inf \left\{ \tfrac{2\tau}{1-\tau} \mid \tau \in (p_\lambda, 1) \text{ and satisfies (6.15) with } q_2 = 0\right\}.$$

- For product weights $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ with $\gamma_{d,j+1} \leq \gamma_{d,j}$ for $j \in [d-1]$ and $\sup_d \gamma_{d,1} < \infty$, we have

  - S_γ is polynomially tractable iff $p_\lambda < 1$ and there exists $\tau \in (p_\lambda, 1)$ such that

$$\limsup_{d \to \infty} \frac{\sum_{j=1}^d \gamma_{d,j}^\tau}{\ln d} < \infty. \tag{6.16}$$

    If (6.16) 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

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

  - S_γ is strongly polynomially tractable iff $p_\lambda < 1$ and $p_\gamma < 1$. Then 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 order-dependent weights $\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$, we have*

  - S_γ *is polynomially tractable iff $p_\lambda < 1$ and there exist $q_2 \geq 0$ and $\tau \in (p_\lambda, 1)$ such that*

  $$C_2 := \sup_d \frac{\left(\sum_{k=0}^d \binom{d}{k} \left[\Gamma_{d,k}\right]^\tau \alpha_\tau^k\right)^{1/\tau}}{\sum_{k=0}^d \binom{d}{k} \Gamma_{d,k} \alpha_1^k} d^{-q_2}. \tag{6.17}$$

  If (6.17) *holds then*

  $$n(\varepsilon, d) \leq \left(\left(\frac{\tau C_2}{1-\tau}\right)^{\tau/(1-\tau)} + 1\right) d^{q_2\tau/(1-\tau)} \varepsilon^{-2\tau/(1-\tau)}$$

  for all $\varepsilon \in (0, 1]$ and $d = 1, 2, \ldots$.

  - S_γ *is strongly polynomially tractable iff $p_\lambda < 1$ and* (6.17) *holds with $q_2 = 0$. Then the exponent of strong polynomial tractability is*

  $$p^{\text{str-avg}} = \inf\left\{\tfrac{2\tau}{1-\tau} \mid \tau \in (p_\lambda, 1) \text{ and satisfies } (6.17) \text{ with } q_2 = 0\right\}.$$

- *For finite-order weights $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > \omega^*$ with order ω^*, we have*

  - S_γ *is polynomially tractable iff $p_\lambda < 1$. Then for any $\tau \in (p_\lambda, 1)$, we have*

  $$n(\varepsilon, d) \leq C \left|\{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \neq 0\}\right| \varepsilon^{-2\tau/(1-\tau)}, \tag{6.18}$$

  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

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

  whereas for finite-diameter weights of order q^* with $q^* < d$ we have

  $$n(\varepsilon, d) \leq 2^{q^*-1} C (d - q^* + 2) \varepsilon^{-2\tau/(1-\tau)},$$

  - S_γ *is strongly polynomially tractable iff $p_\lambda < 1$ and there exists $\tau \in (p_\lambda, 1)$ such that*

  $$\sup_d \frac{\left(\sum_{\mathfrak{u} \subseteq [d], |\mathfrak{u}| \leq \omega^*} \gamma_{d,\mathfrak{u}}^\tau \alpha_\tau^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u} \subseteq [d], |\mathfrak{u}| \leq \omega^*} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|}} < \infty. \tag{6.19}$$

  The exponent of strong polynomial tractability is

  $$p^{\text{str-avg}} = \inf\left\{\tfrac{2\tau}{1-\tau} \mid \tau \in (p_\lambda, 1) \text{ and satisfies } (6.19)\right\}.$$

Proof. The proof easily follows from Theorem 6.2. Indeed, let $\{\lambda_{d,j}\}_{j\in\mathbb{N}}$ be the ordered sequence of the eigenvalues of C_{ν_d}. We have

$$\{\lambda_{d,j}\}_{j\in\mathbb{N}^d} = \left\{\gamma_{d,\mathfrak{u}(j)}\lambda_1^{d-|\mathfrak{u}(j)|}\prod_{k\in\mathfrak{u}(j)}\lambda_{j_k}\right\}_{j\in\mathbb{N}^d}.$$

Then

$$\sum_{j=1}^{\infty}\lambda_{d,j}^\tau = \sum_{j\in\mathbb{N}^d}\gamma_{d,\mathfrak{u}(j)}^\tau\prod_{k=1}^d\lambda_{j_k}^\tau = \gamma_{d,\emptyset}^\tau\lambda_1^{d\tau} + \sum_{\mathfrak{u}\subseteq[d],\mathfrak{u}\neq\emptyset}\gamma_{d,\mathfrak{u}}^\tau\lambda_1^{(d-|\mathfrak{u}|)\tau}\Big(\sum_{j=2}^\infty\lambda_j^\tau\Big)^{|\mathfrak{u}|}$$

$$= \lambda_1^{d\tau}\bigg(\gamma_{d,\emptyset}^\tau + \sum_{\mathfrak{u}\subseteq[d],\mathfrak{u}\neq\emptyset}\gamma_{d,\mathfrak{u}}^\tau\Big(\sum_{j=2}^\infty\Big(\frac{\lambda_j}{\lambda_1}\Big)^\tau\Big)^{|\mathfrak{u}|}\bigg) = \lambda_1^{d\tau}\sum_{\mathfrak{u}\subseteq[d]}\gamma_{d,\mathfrak{u}}^\tau\alpha_\tau^{|\mathfrak{u}|}.$$

Since for each d there is a non-empty $\mathfrak{u}$ such that $\gamma_{d,\mathfrak{u}} > 0$, the sum $\sum_{j=1}^\infty\lambda_{d,j}^\tau$ is finite for some $\tau \in (0,1)$ iff α_τ is finite. This means that $p_\lambda < 1$ if (6.5) holds. Note that

$$\frac{\big(\sum_{j=1}^\infty\lambda_{d,j}^\tau\big)^{1/\tau}}{\sum_{j=1}^\infty\lambda_{d,j}} = \frac{\big(\sum_{\mathfrak{u}\subseteq[d]}\gamma_{d,\mathfrak{u}}^\tau\alpha_\tau^{|\mathfrak{u}|}\big)^{1/\tau}}{\sum_{\mathfrak{u}\subseteq[d]}\gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}}.$$

This yields the formula (6.15) for C_2.

Then the first four points of Theorem 6.8 are equivalent to the first four points of Theorem 6.2 and are repeated for completeness. This is true modulo the bound, $p_\lambda \leq (1+2/p)^{-1}$. This follows from the first part of the proof of Theorem 6.2 which states that $\sum_{j=1}^\infty\lambda_{d,j}^\tau$ is finite for all $\tau \in ((1+2/p)^{-1},1)$ which implies that $\sum_{j=1}^\infty\lambda_j^\tau$ is also finite. Hence $p_\lambda \leq (1+2/p)^{-1}$ as claimed.

Consider now the product weights. For $\tau \in (0,1)$, we have

$$\frac{\big(\sum_{j=1}^\infty\lambda_{d,j}^\tau\big)^{1/\tau}}{\sum_{j=1}^\infty\lambda_{d,j}} = \prod_{j=1}^d\frac{(1+\gamma_{d,j}^\tau\alpha_\tau)^{1/\tau}}{1+\gamma_{d,j}\alpha_1}.$$

Let $\alpha = \sup_d\gamma_{d,j} = \sup_d\gamma_{d,1}$. We know that $\alpha \in (0,\infty)$. For $x \in [0,\alpha]$ define

$$g(x) = 1 + x^\tau\alpha_\tau - (1+cx^\tau\alpha_\tau)(1+x\alpha_1)^\tau$$

for some (small) positive $c < 1$. Obviously $g(0) = 0$, and $g(x) = (1-c)x^\tau\alpha_\tau(1+o(1))$ for small x. Hence g is positive for arguments close to 0. Using the fact that $\alpha_\tau \geq \alpha_1^\tau$ it is easy to check that $g'(x) \geq 0$ for all $x \in (0,\alpha]$ if c is sufficiently small. This means that $g(x) \geq 0$ for all $x \in [0,\alpha]$, and proves that

$$(1+cx^\tau\alpha_\tau)^{1/\tau} \leq \frac{(1+x^\tau\alpha_\tau)^{1/\tau}}{1+x\alpha_1} \leq (1+x^\tau\alpha_\tau)^{1/\tau}.$$

Hence

$$\prod_{j=1}^{d}(1 + c\gamma_{d,j}^{\tau}\alpha_{\tau})^{1/\tau} \le \prod_{j=1}^{d}\frac{(1 + \gamma_{d,j}^{\tau}\alpha_{\tau})^{1/\tau}}{1 + \gamma_{d,j}\alpha_1} \le \prod_{j=1}^{d}(1 + \gamma_{d,j}^{\tau}\alpha_{\tau})^{1/\tau}.$$

Using (5.26) it is now clear that

$$d^{-q}\prod_{j=1}^{d}\frac{(1 + \gamma_{d,j}^{\tau}\alpha_{\tau})^{1/\tau}}{1 + \gamma_{d,j}\alpha_1}$$

is uniformly bounded in d iff

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

The rest of this point follows easily. Strong polynomial tractability follows from the fourth point of the theorem.

The point for order-dependent weights follows from the first point, since C_2 given in (6.17) is now the same as C_2 in (6.15).

For finite-order weights, we have

$$C_2 = \sup_{d}\frac{\left(\sum_{\mathfrak{u}\subseteq[d],|\mathfrak{u}|\le\omega^*}\gamma_{d,\mathfrak{u}}^{\tau}\alpha_{\tau}^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u}\subseteq[d],|\mathfrak{u}|\le\omega^*}\gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}}d^{-q_2}.$$

Note that

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

Since $\alpha_{\tau}^{1/\tau} \ge \alpha_1$, we obtain

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

Using Hölder's inequality, we estimate

$$\left[\sum_{\mathfrak{u}\subseteq[d],|\mathfrak{u}|\le\omega^*}\left(\gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}\right)^{\tau}\right]^{1/\tau} \le \left|\{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \ne 0\}\right|^{(1-\tau)/\tau}\sum_{\mathfrak{u}\subseteq[d],|\mathfrak{u}|\le\omega^*}\gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}.$$

From the proof of Theorem 6.2 we easily conclude that

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

where C is given in Theorem 6.8. The rest is easy since the cardinality of the set $\{\mathfrak{u} \mid \gamma_{d,\mathfrak{u}} \ne 0\}$ is at most $2d^{\omega^*}$ for arbitrary finite-order weights, and it is at most $2^{q^*-1}(d - q^* + 2)$ for finite-diameter weights. Hence we can take $q_2 = \omega^*$ or $q_2 = 1$, respectively. Strong tractability is clear since (6.19) is the same as (6.15) for finite-order weights. This completes the proof. $\qquad\square$

We now comment on Theorems 5.7 and 6.8 which describe the necessary and sufficient conditions on polynomial tractability in the worst case and average case settings. First of all, note that in the worst case setting, tractability conditions are given for the normalized weights and eigenvalues, whereas in the average case setting, they are expressed in terms of the original weights and normalized (somehow differently) eigenvalues. The reason is that the initial errors and the nth minimal errors are different in these two settings. In the worst case setting, the initial error is the square root of the largest eigenvalue and the nth minimal error is the square root of the $(n + 1)$st largest eigenvalue, whereas in the average case setting the initial error is the square root of the sum of all eigenvalues and the nth minimal error is the square root of the truncated sum of all eigenvalues, starting from the $(n + 1)$st largest one. This causes the difference in tractability conditions. In particular, to obtain polynomial tractability we need to assume that the sum-exponent of the normalized eigenvalues is finite in the worst case setting and is less than 1 in the average case setting. We stress that both the sum-exponents are the same, $p_{\lambda^*} = p_\lambda$, although the sequences λ^* and λ of the normalized eigenvalues are slightly different.

The conditions on the weights look formally similar. Again, in the worst case setting, the parameter τ must be greater than the sum-exponent but otherwise can be arbitrarily large, whereas in the average case setting it must be also smaller than one. There is also the difference in the exponents of d and ε^{-1}.

Despite all these technical differences, the essence of Theorems 5.7 and 6.8 is the same. We can guarantee polynomial or strong polynomial tractability only for eigenvalues and weights that decay fast enough. In particular, for product weights we obtain polynomial tractability if the sum of some power of the weights grows no faster than $\ln d$, see (5.21) and (6.16). Again, the power of the weights in the worst case must be larger than the sum-exponent, whereas in the average case setting it must be also less than one. For arbitrary finite-order weights, we obtain polynomial tractability assuming only that the sum-exponent of the univariate eigenvalues is finite in the worst case setting, whereas it must be less than 1 in the average case setting. The dependence on d in both settings is similar, and depends on the cardinality of the non-zero weights. For general finite-order weights this cardinality is of order d^{ω^*}, and it is linear in d for finite-diameter weights.

We turn to weak tractability of the weighted problem S_γ. As in Chapter 5, to omit the scaling problem, we assume that

$$\gamma_{d,\emptyset} = 1 \text{ and } \gamma_{d,\mathfrak{u}} \in [0, 1] \text{ for all non-empty } \mathfrak{u} \subseteq [d]. \tag{6.20}$$

Then the largest eigenvalue of the correlation operator C_{ν_d} is $\lambda_{d,1} = \lambda_1^d$.

We first present necessary conditions for weak tractability. Clearly, the univariate eigenvalues λ_j must go to 0 sufficiently quickly to guarantee that $n(\varepsilon, 1)$ is not exponential in ε^{-1}. More precisely, for a given sequence $\lambda = \{\lambda_j\}$ with $\sum_{j=1}^\infty \lambda_j < \infty$, define

$$n_\lambda(\varepsilon) = \left\{ n \mid \sum_{j=n+1}^\infty \lambda_j \le \varepsilon^2 \sum_{j=1}^\infty \lambda_j \right\}.$$

We say that *the sequence λ admits weak tractability* iff

$$\lim_{\varepsilon^{-1} \to \infty} \frac{\ln n_\lambda(\varepsilon)}{\varepsilon^{-1}} = 0.$$

For example, if $\lambda_j = \Theta\left(j^{-1} \ln^{-p} j\right)$ then λ admits weak tractability iff $p > 3$, whereas if $\lambda_j = \mathcal{O}\left(j^{-p_1} \ln^{-p_2} j\right)$ with $p_1 > 1$ then λ always admits weak tractability.

Lemma 6.9. *Consider the linear weighted tensor product problem $S_\gamma = \{S_{d,\gamma}\}$ with $\lambda_2 > 0$ and with the weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ satisfying (6.20). We study the problem S_γ for the normalized error criterion in the average case setting and for the class Λ^{all}.*

- *We have*

$$n(\varepsilon, d) \geq (1 - \varepsilon^2) \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|}$$

  with $\alpha_1 = \sum_{j=2}^{\infty} \lambda_j / \lambda_1$.

- *If $\gamma_{1,\{1\}} > 0$ then*

$$n(\varepsilon, 1) = n_\lambda(\alpha \varepsilon) \quad \text{for all } \varepsilon \in (0, 1)$$

  with $\alpha = \left[(1 + \gamma_{1,\{1\}} \alpha_1)/(\gamma_{1,\{1\}}(1 + \alpha_1))\right]^{1/2}$.

Hence, $\gamma_{1,\{1\}} > 0$ and weak tractability of S_γ imply that

- *the sequence $\lambda = \{\lambda_j\}$ admits weak tractability,*

- *and*

$$\lim_{d \to \infty} \frac{\ln \left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|} \right)}{d} = 0, \tag{6.21}$$

  which

  - *for product-weights $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ with $\gamma_{d,j+1} \leq \gamma_{d,j}$ and $\gamma_{d,j} \in [0, 1]$ means that*

$$\lim_{d \to \infty} \frac{\sum_{j=1}^{d} \gamma_{d,j}}{d} = 0,$$

  - *for order-dependent weights $\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$ with $\Gamma_{d,k} \in [0, 1]$ means that*

$$\lim_{d \to \infty} \frac{\ln \left(\sum_{k=0}^{d} \binom{d}{k} \Gamma_{d,k} \alpha_1^k \right)}{d} = 0.$$

Proof. For $d \geq 1$ and $\varepsilon \in (0, 1)$, we have

$$n(\varepsilon, d) = \min \left\{ n \mid \sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j} \right\}.$$

For $n = n(\varepsilon, d)$ we know that

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

It was shown in the proof of Theorem 6.8 that

$$\sum_{j=1}^{\infty} \lambda_{d,j} = \lambda_1^d \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|}.$$

Since $\lambda_{d,1} = \lambda_1^d$, we obtain the estimate of $n(\varepsilon, d)$, as claimed in the first point of the lemma.

For $d = 1$, we have $\sum_{j=1}^{\infty} \lambda_{1,j} = \lambda_1 + \gamma_{1,\{1\}} \sum_{j=2}^{\infty} \lambda_j$ and $\sum_{j=n+1}^{\infty} \lambda_{1,j} = \gamma_{1,\{1\}} \sum_{j=n+1}^{\infty} \lambda_j$ for $n \geq 1$. This easily yields $n(\varepsilon, 1) = n_\lambda(\alpha\varepsilon)$, as claimed in the second point of the lemma.

Weak tractability implies, in particular, that $\ln n(\varepsilon, 1) = o(\varepsilon^{-1})$ which implies that λ admits weak tractability. For $\varepsilon < 1$, we also have $\lim_d (\ln n(\varepsilon, d))/d = 0$, and then the lower bound estimate on $n(\varepsilon, d)$ proven above shows the third point of the lemma. $\qquad\square$

Observe that for $\gamma_{d,\mathfrak{u}} = a \in (0, 1]$ for all $\mathfrak{u} \neq \emptyset$, we have

$$\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|} = 1 + a\big[(1 + \alpha_1)^d - 1\big] \geq 1 - a + a\left(1 + \frac{\lambda_2}{\lambda_1}\right)^d.$$

Hence for $\lambda_2 > 0$, we do not have weak tractability. This agrees with the first point of Theorem 6.6 for $a = 1$. Obviously Lemma 6.9 covers cases not addressed by Theorem 6.6. For example, take product weights. Then

$$\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|} = \prod_{j=1}^{d} \left(1 + \alpha_1 \gamma_{d,j}\right).$$

Then for $\lambda_2 > 0$, we have $\alpha_1 > 0$ and weak tractability does not hold if

$$\limsup_{d \to \infty} \frac{\sum_{j=1}^{d} \gamma_{d,j}}{d} > 0.$$

In particular, weak tractability does not hold if $\gamma_{d,j} \geq c_d$ with c_d not tending to 0, or if $\gamma_{d,j} \geq c_1$ for $j \leq c_2 d$, with positive c_1 and c_2 independent of d.

We now show that the necessary condition (6.21) for weak tractability is, in general, *not* sufficient. Indeed, consider S_γ with $\lambda_1 = \lambda_2 = 1$ and $\lambda_j = 0$ for all $j \geq 3$, and with $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} = 2^{-d}$ for all non-empty $\mathfrak{u}$. Then $\alpha_1 = 1$ and

$$\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|} = 1 + 2^{-d}(2^d - 1) = 2 - 2^{-d}.$$

Hence, (6.21) holds. On the other hand, note that the eigenvalues $\{\lambda_{d,j}\}$ are now given by $\lambda_{d,1} = 1$ and $\lambda_{d,j} = 2^{-d}$ for $j = 2, 3, \ldots, 2^d$, and $\lambda_{d,j} = 0$ for $j > 2^d$. For $n \geq 1$, we have $\sum_{j=n+1}^{\infty} \lambda_{d,j} = 1 - n2^{-d}$. Then $\sum_{j=n+1}^{\infty} \lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty} \lambda_{d,j}$ implies that

$$n(\varepsilon, d) \geq 2^d (1 - 2\varepsilon^2) + \varepsilon^2.$$

Hence, for a fixed $\varepsilon \in (0, \sqrt{2}/2)$, we conclude that

$$\lim_{d \to \infty} \frac{\ln n(\varepsilon, d)}{d} = 1,$$

which means that S_γ is *not* weakly tractable.

We are ready to present sufficient conditions for weak tractability. We recall that p_λ is the sum-exponent of the normalized univariate eigenvalues, and in the average case setting we have $p_\lambda \leq 1$. In what follows we assume that $p_\lambda < 1$ leaving the case $p_\lambda = 1$ as an open problem.

Lemma 6.10. *Consider the linear weighted tensor product problem $S_\gamma = \{S_{d,\gamma}\}$ with $\lambda_2 > 0$ and with the weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ satisfying (6.20). We study the problem S_γ for the normalized error criterion in the average case setting and for the class Λ^{all}. We assume that $p_\lambda < 1$. Then*

- *If there exists $\tau \in (p_\lambda, 1)$ such that*

$$\lim_{d \to \infty} \frac{\ln \left[\left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^\tau \alpha_\tau^{|\mathfrak{u}|} \right)^{1/\tau} / \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \alpha_1^{|\mathfrak{u}|} \right]}{d} = 0 \qquad (6.22)$$

then S_γ is weakly tractable.

- *For product weights, $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$, if there exist $\tau \in (p_\lambda, 1)$ such that*

$$\lim_{d \to \infty} \frac{\sum_{j=1}^d \gamma_{d,j}^\tau}{d} = 0 \qquad (6.23)$$

then S_γ is weakly tractable.

- *For order-dependent weights, $\gamma_{d,\mathfrak{u}} = \Gamma_{d,|\mathfrak{u}|}$, if there exist $\tau \in (p_\lambda, 1)$ such that*

$$\lim_{d \to \infty} \frac{\ln \left[\left(\sum_{k=0}^d \binom{d}{k} \Gamma_{d,k}^\tau \alpha_\tau^k \right)^{1/\tau} / \sum_{k=0}^d \binom{d}{k} \Gamma_{d,k} \alpha_1^k \right]}{d} = 0 \qquad (6.24)$$

then S_γ is weakly tractable.

- *For finite-order weights, $\gamma_{d,\mathfrak{u}} = 0$ for $|\mathfrak{u}| > \omega^*$ with order ω^*, S_γ is always weakly tractable.*

Proof. We know that $n^{1/\tau}\lambda_{d,n} \leq \left(\sum_{j=1}^{\infty}\lambda_{d,j}^{\tau}\right)^{1/\tau}$. This yields

$$\sum_{j=n+1}^{\infty}\lambda_{d,j} \leq \frac{\tau}{1-\tau}\,\frac{1}{n^{(1-\tau)/\tau}}\left(\sum_{j=1}^{\infty}\lambda_{d,j}^{\tau}\right)^{1/\tau}.$$

Hence, $\sum_{j=n+1}^{\infty}\lambda_{d,j} \leq \varepsilon^2 \sum_{j=1}^{\infty}\lambda_{d,j}$ holds for

$$n(\varepsilon, d) \leq \left[\left(\frac{\tau}{1-\tau}\frac{\left(\sum_{j=1}^{\infty}\lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty}\lambda_{d,j}}\right)^{\tau/(1-\tau)} + 1\right]\varepsilon^{-2\tau/(1-\tau)}.$$

In the proof of Theorem 6.8 we obtain

$$\frac{\left(\sum_{j=1}^{\infty}\lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty}\lambda_{d,j}} = \frac{\left(\sum_{\mathfrak{u}\subseteq[d]}\gamma_{d,\mathfrak{u}}^{\tau}\alpha_{\tau}^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u}\subseteq[d]}\gamma_{d,\mathfrak{u}}\alpha_1^{|\mathfrak{u}|}}.$$

From this we easily conclude that (6.22) implies $\lim_d (\ln n(\varepsilon, d))/(\varepsilon^{-1} + d) = 0$, which means that S_γ is weakly tractable.

For product weights, we have

$$\frac{\left(\sum_{j=1}^{\infty}\lambda_{d,j}^{\tau}\right)^{1/\tau}}{\sum_{j=1}^{\infty}\lambda_{d,j}} = \frac{\prod_{j=1}^{d}\left(1 + \alpha_{\tau}\gamma_{d,j}^{\tau}\right)^{1/\tau}}{\prod_{j=1}^{d}\left(1 + \alpha_1\gamma_{d,j}\right)} \leq \exp\left(\frac{\alpha_{\tau}}{\tau}\sum_{j=1}^{d}\gamma_{d,j}^{\tau}\right).$$

Then (6.23) implies (6.22) and weak tractability.

For order-dependent weights the condition (6.24) is the same as (6.22). Finally, for finite-order weights we have even polynomial tractability, so obviously weak tractability also holds. This completes the proof.   □

As an illustration, we can return to the example of S_γ before Lemma 6.10 with $\lambda_1 = \lambda_2 = 1$, $\lambda_j = 0$ for $j \geq 3$, and with $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} = 2^{-d}$ for all non-empty $\mathfrak{u}$. Then we have $p_\lambda = 0$, and it is easy to check that the limit in (6.22) is $(1 - \tau)/\tau$, and indeed S_γ is not weakly tractable. Clearly, if we change $\gamma_{d,\mathfrak{u}} = 2^{-d}$ to $\gamma_{d,\mathfrak{u}} = c^d$ with $c \in (0, \frac{1}{2})$ then we can take $\tau = (\ln 2)/(\ln 1/c)$, and (6.22) holds. Hence S_γ becomes weakly tractable.

We stress that Lemmas 6.9 and 6.10 supply necessary and sufficient conditions for weak tractability which are, unfortunately, not the same. Furthermore, in Lemma 6.10 we assumed that $p_\lambda < 1$, although it is not clear whether this assumption is really needed. Hence, much more work is needed to understand weak tractability for linear weighted tensor product problems in the average case setting. This leads us to the next open problem.

Open Problem 30.

- Consider the linear weighted tensor product problem in the average case setting $S_\gamma = \{S_{d,\gamma}\}$ with $\lambda_2 > 0$ as defined in this section. Find necessary and sufficient conditions for weak tractability of S_γ for the normalized error criterion and for the class Λ^{all}.

- In particular, consider the case for which $p_\lambda = 1$ and verify when S_γ is weakly tractable.

Example: Approximation for Continuous Functions (Continued)

We now equip the space $F_d = C([0, 1]^d)$ with a zero-mean Gaussian measure μ_d whose covariance kernel is the *weighted* Korobov kernel,

$$\int_{C([0,1]^d)} f(x) f(y) \mu_d(\mathrm{d} f) = K_{d,\alpha,\gamma}(x, y) \quad \text{for all } x, y \in [0, 1]^d,$$

where $K_{d,\alpha,\gamma}$ is given by (A.3) in Appendix A with $\alpha > 1/2$, and it is the reproducing kernel of the Korobov space $H_{d,\alpha,\gamma}$.

We consider the weighted multivariate approximation problem $\mathrm{APP}_\gamma = \{\mathrm{APP}_d\}$ similarly as before, but this time only for the normalized error criterion. The eigenvalues of the covariance operator C_{ν_d} are now given by

$$\lambda_{d,\gamma,h} = \gamma_{d,\mathfrak{u}_h} \beta_1^{d-|\mathfrak{u}_h|} \beta_2^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} |h_j|^{-2\alpha} \quad \text{for all } h \in \mathbb{Z}^d,$$

where, as always, $\mathfrak{u}_h = \{j \mid h_j \neq 0\}$. For simplicity we assume that $\beta_1 \geq \beta_2$. Then

$$\alpha_\tau = 2 \left(\frac{\beta_2}{\beta_1} \right)^\tau \zeta(2\alpha\tau) \quad \text{and} \quad p_\lambda = \tfrac{1}{2\alpha} < 1.$$

We are now ready to apply Theorem 6.8 as well Lemmas 6.9 and 6.10 on polynomial and weak tractability for the normalized error criterion. In particular, we obtain

- Let $\gamma_{d,\mathfrak{u}} = d^{-s|\mathfrak{u}|}$. Then APP_γ is strongly polynomially tractable iff $s > 1$. For $s > 1$, the exponent of strong polynomial tractability is

$$\max \left(\frac{2}{s-1}, \frac{1}{\alpha - \frac{1}{2}} \right).$$

  Note that this exponent may be, however, arbitrarily large if s is close to 1 or if α is close to $1/2$.

- Let $\gamma_{d,\mathfrak{u}} = d^{-s|\mathfrak{u}|}$. Then APP is weakly tractable iff $s > 0$.

As we know, APP_γ is intractable for the unweighted case and the normalized error criterion, since the problem suffers the curse of dimensionality. Hence, the weights can break the curse and we may even have strong polynomial tractability.

Example: Approximation for Weighted Korobov Space (Continued)

We now consider the weighted Korobov space $H_{d,\alpha,\gamma}$ defined in Appendix A, and studied in Chapter 5 in the worst case setting. We equip the space $H_{d,\alpha,\gamma}$ with a zero-mean Gaussian measure whose covariance operator C_{μ_d} has the eigenpairs $C_{\mu_d} e_h = \alpha_{d,h} e_h$ with e_h defined in Appendix A for $h \in \mathbb{Z}^d$. To preserve the structure of the weighted space $H_{d,\alpha,\gamma}$ we take the sequence $\xi = \{\xi_{d,\mathfrak{u}}\}$ of weights such that the eigenvalues $\alpha_{d,h} = \alpha_{d,\xi,h}$ depend on the weights $\xi_{d,\mathfrak{u}}$ and are given by

$$\alpha_{d,\xi,h} = \xi_{d,\mathfrak{u}_h} \eta_1^{d-|\mathfrak{u}_h|} \eta_2^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} |h_j|^{-2\beta} \quad \text{for all } h \in \mathbb{Z}^d,$$

where η_1 and η_2 are positive, and $\beta > 1/2$.

We consider multivariate approximation APP_γ for the normalized error criterion. Then C_{ν_d} has the eigenvalues

$$\lambda_{d,\gamma,h} = \gamma_{d,\mathfrak{u}_h} \xi_{d,\mathfrak{u}_h} (\eta_1 \beta_1)^{d-|\mathfrak{u}_h|} (\eta_2 \beta_2)^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} |h_j|^{-2(\alpha+\beta)} \quad \text{for all } h \in \mathbb{Z}^d.$$

Then $p_\lambda = 1/(2(\alpha+\beta)) < 1/2$ no matter how close α and β are to $1/2$.

Again we may now apply Theorem 6.8 as well Lemmas 6.9 and 6.10 on polynomial and weak tractability for the normalized error criterion. In particular, consider

$$\gamma_{d,\mathfrak{u}} = d^{-s_1|\mathfrak{u}|} \quad \text{and} \quad \xi_{d,\mathfrak{u}} = d^{-s_2|\mathfrak{u}|}$$

for non-negative s_1 and s_2. Let $s = s_1 + s_2$. We obtain

- APP_γ is strongly polynomially tractable iff $s > 1$. For $s > 1$, the exponent of strong polynomial tractability is

$$\max\left(\frac{2}{s-1}, \frac{1}{\alpha+\beta-\frac{1}{2}}\right) \leq \max\left(\frac{2}{s-1}, 2\right).$$

  Note that this exponent may be arbitrarily large only if s is close to 1.

- APP_γ is weakly tractable iff $s > 0$.

So we obtained similar results as for the previous example, with the only difference that the exponent of strong polynomial tractability does depend weakly on the smoothness parameters α and β.

We now briefly compare tractability results obtained for the space $H_{d,\alpha,\gamma}$ in the worst and average case settings for the normalized error criterion. To guarantee that the average case setting over the whole space is roughly the same as the average case over the unit ball, we need to assume that the trace of C_{μ_d} is of order 1. That is,

$$\mathrm{trace}(C_{\mu_d}) = \sum_{\mathfrak{u} \subseteq [d]} \xi_{d,\mathfrak{u}} \eta_1^{d-|\mathfrak{u}|} \eta_2^{|\mathfrak{u}|} [2\zeta(2\beta)]^{|\mathfrak{u}|} = \mathcal{O}(1) \quad \text{as } d \to \infty. \tag{6.25}$$

There is one more assumption to have a fair comparison between the worst and average case settings for the normalized error criterion. Namely, we need to guarantee that the initial errors in the worst and average case settings are comparable. For simplicity, we take $\beta_1 \geq \beta_2$, $\eta_1 \geq \eta_2$ and assume that $\gamma_{d,\mathfrak{u}}, \xi_{d,\mathfrak{u}} \in [0, 1]$ with $\gamma_{d,\emptyset} = \xi_{d,\emptyset} = 1$. Denote by R_d the ratio of the squares of the worst and average case initial errors. Then we assume that

$$R_d := \frac{\beta_1^d}{\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \xi_{d,\mathfrak{u}} (\eta_1 \beta_1)^{d - |\mathfrak{u}|} (\eta_2 \beta_2)^{|\mathfrak{u}|} [2\zeta(2(\alpha + \beta))]^{|\mathfrak{u}|}} = \mathcal{O}(1) \quad \text{as } d \in \infty.$$
(6.26)

We have

- APP_γ is polynomially tractable in the worst case setting iff

$$\sup_d \left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{\tau^{\mathrm{wor}}} [\beta_2/\beta_1]^{\tau^{\mathrm{wor}}|\mathfrak{u}|} [2\zeta(2\tau^{\mathrm{wor}}\alpha)]^{|\mathfrak{u}|} \right)^{1/\tau^{\mathrm{wor}}} d^{-q_2^{\mathrm{wor}}} < \infty$$

  for some $q_2^{\mathrm{wor}} \geq 0$ and $\tau^{\mathrm{wor}} > (2\alpha)^{-1}$. If so then

$$n^{\mathrm{wor}}(\varepsilon, d) = \mathcal{O}\big(d^{q_2^{\mathrm{wor}} \tau^{\mathrm{wor}}} \varepsilon^{-2\tau^{\mathrm{wor}}}\big).$$

- APP_γ is polynomially tractable in the average case setting iff

$$\sup_d \frac{\left(\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}}^{\tau^{\mathrm{avg}}} [\eta_2 \beta_2/(\eta_1 \beta_1)]^{\tau^{\mathrm{avg}}|\mathfrak{u}|} [2\zeta(2\tau^{\mathrm{avg}}(\alpha + \beta))]^{|\mathfrak{u}|} \right)^{1/\tau^{\mathrm{avg}}}}{d^{-q_2^{\mathrm{avg}}} \sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}} [\eta_2 \beta_2/(\eta_1 \beta_1)]^{|\mathfrak{u}|} [2\zeta(2(\alpha + \beta))]^{|\mathfrak{u}|}} < \infty$$

  with $\psi_{d,\mathfrak{u}} = \gamma_{d,\mathfrak{u}} \xi_{d,\mathfrak{u}}$, for some $q_2^{\mathrm{avg}} \geq 0$ and $\tau^{\mathrm{avg}} \in (2(\alpha + \beta))^{-1}, 1)$. If so then

$$n^{\mathrm{avg}}(\varepsilon, d) = \mathcal{O}\big(d^{q_2^{\mathrm{avg}} \tau^{\mathrm{avg}}/(1 - \tau^{\mathrm{avg}})} \varepsilon^{-2\tau^{\mathrm{avg}}/(1 - \tau^{\mathrm{avg}})}\big).$$

Obviously, if q_2^{wor} or q_2^{avg} is 0 then we have strong polynomial tractability in the corresponding setting.

We now show that if (6.25) and (6.26) hold then

- polynomial tractability in the worst case setting implies polynomial tractability in the average case setting. Furthermore, the exponents of ε^{-1} and d in the average case setting are not larger than their counterparts in the worst case setting.

To prove this claim, take $\tau = \tau^{\mathrm{wor}}/(1 + \tau^{\mathrm{wor}})$, so that $\tau/(1 - \tau) = \tau^{\mathrm{wor}}$. We want to show that we can take $\tau^{\mathrm{avg}} = \tau$.

We now use Hölder's inequality with $p = 1 + \tau$ and $q = 1 + 1/\tau^{\mathrm{wor}}$, so that $\tau p = \tau^{\mathrm{wor}}$ and $\tau q = 1$, and obtain

$$\sum_{j=1}^{\infty} \frac{1}{j^{2\tau(\alpha + \beta)}} \leq \Big(\sum_{j=1}^{\infty} \frac{1}{j^{2\tau^{\mathrm{wor}}\alpha}} \Big)^{1/p} \Big(\sum_{j=1}^{\infty} \frac{1}{j^{2\beta}} \Big)^{1/q}.$$

This implies that

$$2\zeta(2\tau(\alpha + \beta)) \leq [2\zeta(2\tau^{\mathrm{wor}}\alpha)]^{1/p} [2\zeta(2\beta)]^{1/q}.$$

Then

$$\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}}^{\tau} [\eta_2\beta_2/(\eta_1\beta_1)]^{\tau|\mathfrak{u}|} [2\zeta(2\tau(\alpha + \beta))]^{|\mathfrak{u}|}$$

$$\leq \sum_{\mathfrak{u} \subseteq [d]} \left(\gamma_{d,\mathfrak{u}}^{\tau p} [\beta_2/\beta_1]^{\tau p|\mathfrak{u}|} [2\zeta(2\tau^{\mathrm{wor}}\alpha)]^{|\mathfrak{u}|}\right)^{1/p} \left(\xi_{d,\mathfrak{u}}^{\tau q} (\eta_2/\eta_1)^{\tau q|\mathfrak{u}|} [2\zeta(2\beta)]^{|\mathfrak{u}|}\right)^{1/q}.$$

Applying now Hölder's inequality for finite sums, and using again the fact that $\tau p = \tau^{\mathrm{wor}}$ and $\tau q = 1$, we obtain

$$\left(\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}}^{\tau} [\eta_2\beta_2/(\eta_1\beta_1)]^{\tau|\mathfrak{u}|} [2\zeta(2\tau(\alpha + \beta))]^{|\mathfrak{u}|}\right)^{1/\tau}$$

$$\leq \left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{\tau^{\mathrm{wor}}} [\beta_2/\beta_1]^{\tau^{\mathrm{wor}}|\mathfrak{u}|} [2\zeta(2\tau^{\mathrm{wor}}\alpha)]^{|\mathfrak{u}|}\right)^{1/\tau^{\mathrm{wor}}} \mathrm{trace}(C_{\mu_d})\eta_1^{-d}.$$

Note also that

$$\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}} [\eta_2\beta_2/(\eta_1\beta_1)]^{|\mathfrak{u}|} [2\zeta(2(\alpha + \beta))]^{|\mathfrak{u}|}$$

$$= \frac{\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}} (\eta_1\beta_1)^{d-|\mathfrak{u}|} (\eta_2\beta_2)^{|\mathfrak{u}|} [2\zeta(2(\alpha + \beta))]^{|\mathfrak{u}|}}{(\eta_1\beta_1)^d}$$

$$= \frac{1}{\eta_1^d R_d}.$$

From this we conclude that

$$\frac{\left(\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}}^{\tau} [\eta_2\beta_2/(\eta_1\beta_1)]^{\tau|\mathfrak{u}|} [2\zeta(2\tau(\alpha + \beta))]^{|\mathfrak{u}|}\right)^{1/\tau}}{\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}} [\eta_2\beta_2/(\eta_1\beta_1)]^{|\mathfrak{u}|} [2\zeta(2(\alpha + \beta))]^{|\mathfrak{u}|}}$$

$$= \eta_1^d R_d \left(\sum_{\mathfrak{u} \subseteq [d]} \psi_{d,\mathfrak{u}}^{\tau} [\eta_2\beta_2/(\eta_1\beta_1)]^{\tau|\mathfrak{u}|} [2\zeta(2\tau(\alpha + \beta))]^{|\mathfrak{u}|}\right)^{1/\tau}$$

$$\leq \left(\sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{\tau^{\mathrm{wor}}} [\beta_2/\beta_1]^{\tau^{\mathrm{wor}}|\mathfrak{u}|} [2\zeta(2\tau^{\mathrm{wor}}\alpha)]^{|\mathfrak{u}|}\right)^{1/\tau^{\mathrm{wor}}} R_d \, \mathrm{trace}(C_{\mu_d}).$$

The rest is easy. Suppose we have polynomial tractability in the worst case setting. Since both R_d and $\mathrm{trace}(C_{\mu_d})$ are of order 1, then the last expression is bounded by $\mathcal{O}(d^{q_2^{\mathrm{wor}}})$. Hence, we have polynomial tractability also in the average case setting and we can take $\tau^{\mathrm{avg}} = \tau$, so that the exponent of ε^{-1} is at most $2\tau/(1 - \tau) = \tau^{\mathrm{wor}}$, and the exponent of d is at most $q_2^{\mathrm{wor}}\tau^{\mathrm{wor}}$, as claimed.

We now compare tractability results for the normalized error criterion in the worst and average case settings for a few specific weights. For simplicity we additionally assume that $\beta_1 = 1$, so that the initial error in the worst case setting is 1 for all $d \in \mathbb{N}$.

- Let $\gamma_{d,\mathfrak{u}} = d^{-s_1|\mathfrak{u}|}$ and $\xi_{d,\mathfrak{u}} = d^{-s_2|\mathfrak{u}|}$ for non-negative s_1 and s_2. Then we have strong polynomial tractability in the worst case setting iff $s_1 > 0$ and then the exponent of strong tractability is

$$p^{\mathrm{wor}} = \max\left(\frac{2}{s_1}, \frac{1}{\alpha}\right).$$

Assume then that $s_1 > 0$.

The square of the initial error in the average case setting and R_d are now

$$\mathrm{trace}(C_{\mu_d}) = \left(\eta_1 + \frac{2\eta_2\zeta(2\beta)}{d^{s_2}}\right)^d$$
$$= \exp\left(d\left[a_1 d^{-s_2} + \ln\eta_1\right]\right)(1 + o(1))$$
$$R_d = \left(\eta_1 + \frac{2\eta_2\beta_2\zeta(2(\alpha+\beta))}{d^{s_1+s_2}}\right)^{-d}$$
$$= \exp\left(d\left[-a_2 d^{-s_1-s_2} + \ln\eta_1^{-1}\right]\right)(1 = o(1)).$$

Here

$$a_1 = \frac{2\eta_2\zeta(2\beta)}{\eta_1} \quad \text{and} \quad a_2 = \frac{2\eta_2\beta_2\zeta(2(\alpha+\beta))}{\eta_1}.$$

It can be checked that (6.25) and (6.26) hold iff

$$s_2 \geq 1 \quad \text{and} \quad \eta_1 = 1.$$

In particular, this means that we *cannot* take $\xi_{d,\mathfrak{u}} \equiv 1$ if we want to have a fair comparison with the worst case setting.

As we know we have strong polynomial tractability in the average case setting iff $s_1 + s_2 > 1$ which is satisfied if $s_1 > 0$ since we must take $s_2 \geq 1$. However, if we take $s_1 = 0$ then we have polynomial intractability in the worst case setting and strong polynomial tractability in the average case setting if $s_1 > 1$.

If $s_1 + s_2 > 1$ then the exponent of strong tractability in the average case setting is

$$p^{\mathrm{avg}} = \max\left(\frac{2}{s_1 + s_2 - 1}, \frac{1}{\alpha + \beta - \frac{1}{2}}\right).$$

Note that $p^{\mathrm{avg}} \leq p^{\mathrm{wor}}$ since $\beta > \frac{1}{2}$. If $s_2 = 1$ then $p^{\mathrm{avg}} = p^{\mathrm{wor}}$, whereas for $s_2 > 1$ we have $p^{\mathrm{avg}} < p^{\mathrm{wor}}$. Furthermore, the difference $p^{\mathrm{wor}} - p^{\mathrm{avg}}$ can be arbitrarily large. Indeed, for α close to 0 and $s_1 = 2\alpha$ we have $p^{\mathrm{wor}} = 1/\alpha$ which can be arbitrarily large since we deal with not very smooth functions. On the other hand, for $s_2 = 2$ and $\beta = 3/2$, we obtain $p^{\mathrm{avg}} \leq 2$.

- Let $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ be product weights with $\gamma_{d,j+1} \leq \gamma_{d,j}$ and $\gamma_{d,j} \in (0, 1]$. Similarly, let $\xi_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \xi_{d,j}$ be also product weights with $\xi_{d,j+1} \leq \xi_{d,j}$ and $\xi_{d,j} \in (0, 1]$. Note that the previous weights are also product weights, but now we consider the general case of product weights.

Let $\eta_1 = 1$. To guarantee that the trace of C_{μ_d} is of order 1, we must assume that $\sup_d \sum_{j=1}^d \xi_{d,j} < \infty$ which implies that $p_\xi \leq 1$. This also implies that $R_d = \mathcal{O}(1)$.

As before, let $\psi = \{\psi_{d,\mathfrak{u}}\}$ with $\psi_{d,\mathfrak{u}} = \gamma_{d,\mathfrak{u}} \xi_{d,\mathfrak{u}}$. It is easy to verify that

$$p_\psi = \frac{1}{\frac{1}{p_\gamma} + \frac{1}{p_\xi}},$$

so that $2p_\psi$ is the harmonic mean of p_γ and p_ξ. Then we obtain strong polynomial tractability in the worst case setting iff $p_\gamma < \infty$ and then the exponent of strong tractability is

$$p^{\mathrm{wor}} = \max\left(2p_\gamma, \alpha^{-1}\right).$$

Strong polynomial tractability in the average case setting holds iff $p_\psi < 1$. The last condition may hold even if $p_\gamma = \infty$, i.e., when we do not have strong polynomial tractability in the worst case setting. Indeed, in this case, we have $p_\psi = p_\xi$ and it is enough to assume that $p_\xi < 1$.

If strong polynomial tractability holds in the average case setting then the exponent of strong tractability is

$$p^{\mathrm{avg}} = \frac{\max\left(2p_\psi, (\alpha + \beta)^{-1}\right)}{1 - \frac{1}{2}\max\left(2p_\psi, (\alpha + \beta)^{-1}\right)}.$$

Again $p^{\mathrm{avg}} \leq p^{\mathrm{wor}}$ which can also be directly checked. As for the previous case of weights, these exponents can be equal. For instance, take $p_\gamma = p_\xi = 1$ and $\alpha = \beta = \frac{1}{2}$. Then $p^{\mathrm{wor}} = p^{\mathrm{avg}} = 2$. On the other hand, if α is close to 0 and $2p_\gamma \leq \alpha^{-1}$ then $p^{\mathrm{wor}} = \alpha^{-1}$ can be arbitrarily large, whereas for $p_\xi = 2$ and $\beta + 1$, we have $p^{\mathrm{avg}} \leq 2$.

Example: Schrödinger Equation (Continued)

We now equip the space H_d as before with a zero-mean Gaussian measure. The only change is that the eigenvalues $\alpha_{d,j} = \alpha_{d,\gamma,j}$ of C_{μ_d} depend now on a given weight sequence,

$$\alpha_{d,\gamma,j} = \gamma_{d,\mathfrak{u}(j)} \frac{\left(q_0 + \sum_{k=1}^d \pi^2 j_k^2\right)^2}{(j_1 + 1)^\alpha (j_2 + 1)^\alpha \cdots (j_d + 1)^\alpha} \quad \text{for all } j \in \mathbb{N}^d,$$

where $\mathfrak{u}(j) = \{k \in [d] \mid j_k \geq 2\}$ and $\alpha > 5$. Then the eigenvalues of C_{ν_d} are

$$\lambda_{d,\gamma,j} = \frac{\gamma_{d,\mathfrak{u}(j)}}{(j_1 + 1)^\alpha (j_2 + 1)^\alpha \cdots (j_d + 1)^\alpha} \quad \text{for all } j \in \mathbb{N}^d.$$

For simplicity, we only consider the specific product weights, $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} j^{-s}$ for $s \geq 0$. Then $p_\lambda = 1/\alpha$ and $p_\gamma = 1/s$.

Using the proofs of Theorem 6.8, Lemmas 6.9 and 6.10, we conclude that the linear Schrödinger problem for the normalized error criterion in the average case setting with the zero-mean Gaussian measure considered here is

- strongly polynomially tractable iff polynomially tractable iff $s > 1$,

- not polynomially tractable but weakly tractable iff $s \in (0, 1]$,

- intractable iff $s = 0$.

6.4 Notes and Remarks

NR 6:1. The results on weak tractability presented in this chapter are new. The results on polynomial tractability have been studied in several papers and we will give proper references later. Here we only want to mention that tractability in the average case setting was already studied in the first tractability paper [287] which dealt with linear multivariate problems for the absolute error criterion.

NR 6.1:1. Theorem 6.1 on polynomial tractability for the absolute error criterion is a variant of Theorem 5.1 of [287], where the condition (6.1) is replaced by the equivalent condition (6.4). As in the worst case setting, tractability in ε^{-1} and d was also studied in [287] in the average case setting. The idea of using summability of the singular values as a technical tool for tractability study is from [94], and was also used in [93]. The reader is also referred to [269] for more information.

NR 6.1:2. Theorem 6.2 on polynomial tractability for the normalized error criterion is from [93], whereas strong tractability was proved before in [94].

NR 6.2:1. The approach for linear and linear weighted tensor product problems in the average case setting is taken from [93]. Theorem 6.6 for the normalized case can be easily deduced from Theorem 1 and Corollary 1 in [93], where the general weighted case is studied. The absolute error criterion has not been studied before. In any case, Theorem 5.5 is quite simple to prove.

NR 6.3:1. Theorem 6.8 is basically from Theorem 1 and Corollary 1 in [93], and the case of finite-order weights corresponds to Corollary 2 in [93].

NR 6.3:2. Comparison of tractability results in the worst and average case settings for multivariate approximation defined for weighted Korobov spaces with product weights and for lattice rules algorithm can be found in [122].

Chapter 7
Randomized Setting

In the previous chapters we studied the worst case and average case settings for the class Λ^{all} of all continuous linear functionals. In this chapter we switch to the randomized setting for linear problems defined between Hilbert spaces. We already know from Section 4.3.3 that there is a close relationship between the worst case and randomized settings for the class Λ^{all}, and randomization does not practically help, see Theorem 4.42. Therefore for the class Λ^{all}, all tractability results obtained for the worst case setting are also valid in the randomized setting, see Section 7.1 for details.

The purpose of this very short chapter is only to summarize this negative result that randomization does not help. This chapter can be also viewed as an introduction to study randomization for the class Λ^{std} of function values, which will be done in Volume II.

7.1 Tractability of Linear Problems for Λ^{all}

As in Chapter 4, for $d = 1, 2, \ldots$, let $S_d : H_d \to G_d$ be a compact linear operator, where H_d is a Hilbert space of real functions defined on $D_d \subseteq \mathbb{R}^d$, and G_d is also a Hilbert space. We deal with such problems in the randomized setting, and restrict ourselves to measurable randomized algorithms. For the class Λ^{all}, see Remark 4.37 and Theorem 4.42, we know that for all $n \geq 1$ we have

$$\tfrac{1}{2} e^{\text{wor}}(4n - 1, d; \Lambda^{\text{all}}) \leq e^{\text{ran}}(n, d; \Lambda^{\text{all}}) \leq e^{\text{wor}}(n, d; \Lambda^{\text{all}}), \qquad (7.1)$$

where $e^{\text{wor}}(n, d)$ and $e^{\text{ran}}(n, d)$ denote the minimal worst case and randomized errors if we use n deterministic and randomized information operations from Λ^{all}, respectively, for approximating the operator S_d.

From these two inequalities it is easy to conclude the equivalence of tractabilities in the worst case and randomized settings. More precisely, let $n^{\text{wor}}(\varepsilon, d; \Lambda^{\text{all}})$ and $n^{\text{ran}}(\varepsilon, d; \Lambda^{\text{all}})$ denote the minimal number of information operations from Λ^{all} needed to approximate S_d to within ε in the worst case and randomized setting for the absolute or normalized error criterion. Then for $n^{\text{ran}}(\varepsilon, d; \Lambda^{\text{all}}) \geq 1$ we have

$$\tfrac{1}{4} \left(n^{\text{wor}}(2\varepsilon, d; \Lambda^{\text{all}}) + 1 \right) \leq n^{\text{ran}}(\varepsilon, d; \Lambda^{\text{all}}) \leq n^{\text{wor}}(\varepsilon, d; \Lambda^{\text{all}}). \qquad (7.2)$$

This yields the following corollary.

Corollary 7.1. *Consider the linear problem $S = \{S_d\}$, with compact linear operators S_d between Hilbert spaces, for the class Λ^{all} and for the absolute or normalized*

error criterion. Then weak tractability, polynomial tractability and strong polynomial tractability in the randomized setting are equivalent to weak, polynomial tractability and strong polynomial tractability in the worst case setting for the class Λ^{all} and for the absolute or normalized error criterion, respectively. Furthermore, the exponents of polynomial tractability and strong polynomial tractability are the same in the two settings.

Estimates (7.2) can be also used for (T, Ω)-tractability, see Section 4.4.3. If $T((2\varepsilon)^{-1}, d) = \mathcal{O}\left(T(\varepsilon^{-1}, d)\right)$ for all $(\varepsilon^{-1}, d) \in \Omega$ then we have the equivalence of (T, Ω)-tractability of S in the worst case and randomized settings. Similarly, if $T((2\varepsilon)^{-1}, 1) = \mathcal{O}\left(T(\varepsilon^{-1}, 1)\right)$ for all $(\varepsilon^{-1}, d) \in \Omega$ then we have the equivalence of strong (T, Ω)-tractability of S in the worst case and randomized settings.

Corollary 7.1 presents a negative result that, as long as we consider linear problems defined over Hilbert spaces, the randomized setting is practically the same as the worst case setting for the class Λ^{all}. The assumption about Hilbert space is essential since randomization may help for some linear problems defined over certain Banach spaces, see Heinrich [79] and Mathé [141]. The assumption about the class Λ^{all} is also essential.

As we shall see later, the randomized setting is especially important for the class Λ^{std}, where only function values can be computed. For many multivariate problems, randomization significantly helps for Λ^{std}, although it cannot help more then the class Λ^{all} in the worst case setting. However, for some multivariate problems the class Λ^{all} is so powerful that $e^{\mathrm{wor}}(n, d, \Lambda^{\mathrm{all}}) = 0$ for all $n \geq n_0$ with n_0 independent of d. For instance, for continuous linear functionals, we have $n_0 = 1$, and for linear multivariate problems with S_d of rank k, we have $n_0 = k$. This really makes such problems trivial for the class Λ^{all}. The only interesting case for such problems is to study the power of randomization for the class Λ^{std}. The most known case is, of course, Monte Carlo for multivariate integration. As we shall see, Monte Carlo may break intractability of the worst case setting for many classes of functions. In fact, we will see examples of commonly used spaces for which multivariate integration is intractable in the worst case setting and strongly polynomially tractable in the randomized setting.

Chapter 8
Generalized Tractability

In the previous chapters we studied polynomial and weak tractability of multivariate problems in the worst, average, and randomized settings for the class Λ^{all}. In this chapter we study *generalized* tractability that was already briefly defined in Chapter 4. We restrict ourselves only to the worst case setting and the class Λ^{all}. The class Λ^{std} will be studied in Volume II for continuous linear functionals in the worst case setting. Generalized tractability in other settings has not yet been studied. This chapter is based on [69], [70], and all results reported here were originally proved in these two papers.

Generalized tractability is defined in terms of a function T and a set Ω. We assume that

- T, called a *tractability function*, is defined on $[1, \infty) \times [1, \infty)$ and is non-decreasing in both variables ε^{-1} and d and grows slower than exponentially to infinity,

- Ω, called a *tractability domain set*, is a subset of $[1, \infty) \times \mathbb{N}$ for which at least one of the variables can go to infinity.

As always, let $n(\varepsilon, d)$ denote the information complexity in a given setting for the absolute or normalized error criterion. To obtain *generalized tractability*, we need to verify that $n(\varepsilon, d)$ can be bounded by a multiple of a power of $T(\varepsilon^{-1}, d)$ for all $(\varepsilon^{-1}, d) \in \Omega$. To obtain *strong generalized tractability*, we need to verify that $n(\varepsilon, d)$ can be bounded by a multiple of a power of $T(\varepsilon^{-1}, 1)$ for all $(\varepsilon^{-1}, d) \in \Omega$. The smallest powers of $T(\varepsilon^{-1}, d)$ or $T(\varepsilon^{-1}, 1)$ are called the exponents of (generalized) tractability.

We will mainly study linear (unweighted) tensor product problems for the class Λ^{all}. We present necessary and sufficient conditions on T such that generalized tractability holds for such multivariate problems. In particular, we exhibit a number of examples for which polynomial tractability does *not* hold but generalized tractability *does*.

We add in passing that for linear tensor product problems, we needed to introduce weights to obtain polynomial tractability under suitable assumptions on the decay of weights. As we shall see, generalized tractability often holds even for unweighted spaces. Hence, there is a trade-off. If we want to guarantee polynomial tractability we must, in general, work with weighted spaces and impose suitable conditions on weights. On the other hand, if we want to work with unweighted spaces then we must relax the notion of polynomial tractability and switch to generalized tractability for functions T which are non-exponential but tend to infinity faster than polynomials.

Generalized tractability may differ from polynomial tractability in two ways. The first is the domain of (ε, d). For polynomial tractability, ε and d are independent, and $(\varepsilon^{-1}, d) \in [1, \infty) \times \mathbb{N}$. For some applications, as in mathematical finance, d is huge but we are only interested in a rough approximation, so that ε is not too small.

There may be also problems for which d is relatively small and we are interested in a very accurate approximation which corresponds to a very small ε. For generalized tractability, we assume that $(\varepsilon^{-1}, d) \in \Omega$, where

$$[1, \infty) \times \{1, 2, \ldots, d^*\} \cup [1, \varepsilon_0^{-1}) \times \mathbb{N} \subseteq \Omega \subseteq [1, \infty) \times \mathbb{N} \qquad (8.1)$$

for some non-negative integer d^* and some $\varepsilon_0 \in (0, 1]$ such that

$$d^* + (1 - \varepsilon_0) > 0.$$

The importance of the case $d^* = 0$ will be explained later.

The essence of (8.1) is that for all such Ω, we know that at least one of the parameters (ε^{-1}, d) may go to infinity but not necessarily both of them. Hence, for generalized tractability we assume that $(\varepsilon^{-1}, d) \in \Omega$ and we may choose an arbitrary Ω satisfying (8.1) for some d^* and ε_0.

The second way in which generalized tractability may differ from polynomial tractability is how we measure the lack of exponential dependence. For polynomial tractability, we want to bound the information complexity $n(\varepsilon, d)$ by a polynomial in ε^{-1} and d, whereas for generalized tractability we want to bound $n(\varepsilon, d)$ by a multiple of a power of $T(\varepsilon^{-1}, d)$ that can go faster to infinity than polynomially.

We are mainly interested in how the choice of Ω and T affects the class of tractable problems. The first promising result was obtained in [293] for

$$\Omega = [1, \infty) \times \mathbb{N} \quad \text{and} \quad T(x, y) = f_1(x) f_2(y) \text{ with } f_i(t) = \exp(\ln^{1+\alpha_i} t).$$

Here α_1 and α_2 are non-negative. Note that for $\alpha_1 = \alpha_2 = 0$ we have $T(x, y) = xy$, which corresponds to polynomial tractability. Namely, it was proved that (unweighted) linear tensor product problems with polynomially decaying eigenvalues are tractable iff $\alpha_1 \alpha_2 \geq 1$. Hence, these problems are *not* polynomially tractable, but are tractable if, for example, $\alpha_1 = \alpha_2 = 1$. This shows that we can regain tractability of the (unweighted) linear tensor product problems if we agree to consider functions T going to infinity like $\exp(\ln^2 x + \ln^2 y)$.

In this chapter, we study linear tensor product problems for the class Λ^{all}. Let $\lambda = \{\lambda_j\}$ be the sequence of the singular values for the univariate case. We assume that $\lambda_1 = 1$ so that the absolute and normalized error criteria coincide. We first choose a "smallest" set,

$$\Omega = \Omega^{\text{res}} = [1, \infty) \times \{1, 2, \ldots, d^*\} \cup [1, \varepsilon_0^{-1}) \times \mathbb{N},$$

which is called the *restricted tractability domain*.

We provide necessary and sufficient conditions on the tractability function T such that generalized tractability holds for Ω^{res}. These conditions depend on the parameters d^* and ε_0, as well as on the sequence λ. In particular, the following results hold. Assume that $d^* \geq 1$ and $\varepsilon_0 < 1$. If the largest eigenvalue has multiplicity at least two, i.e., if $\lambda_2 = 1$, then generalized tractability does *not* hold, no matter how we choose the tractability function T.

Assume then that $\lambda_2 < 1$ and that we have a polynomial rate of convergence of the singular eigenvalues, i.e., $\lambda_j = \Theta(j^{-\beta})$ for a positive β. This case is typical and corresponds to many classical Sobolev or Korobov tensor product spaces of smooth functions whose smoothness is measured by the parameter β.

Assume first that $\varepsilon_0^2 < \lambda_2$. Then generalized strong tractability does *not* hold, no matter how we choose T. Generalized tractability holds iff

$$\liminf_{x\to\infty} \frac{\ln T(x,1)}{\ln x} \in (0,\infty] \quad \text{and} \quad \liminf_{d\to\infty} \inf_{\varepsilon\in[\varepsilon_0,\sqrt{\lambda_2})} \frac{\ln T(\varepsilon^{-1},d)}{\alpha(\varepsilon)\ln d} \in (0,\infty],$$

where $\alpha(\varepsilon) = \lceil 2\ln(1/\varepsilon)/\ln(1/\lambda_2)\rceil - 1$. In particular, if we take $T(x,y) = xy$ then polynomial tractability *holds* with the exponent

$$t^{\text{tra}} = \max\left\{\frac{2}{\beta}, \alpha(\varepsilon_0)\right\}.$$

Note that t^{tra} goes to $\max\{2/\beta, 1\}$ as $\lambda_2 - \varepsilon_0^2$ tends to 0, and t^{tra} goes to infinity as ε_0 tends to 0.

Assume now that $\lambda_2 \leq \varepsilon_0^2$. Then generalized strong tractability holds iff

$$\liminf_{x\to\infty} \frac{\ln T(x,1)}{\ln x} \in (0,\infty].$$

For $T(x,y) = xy$, we obtain strong polynomial tractability with the exponent $t^{\text{str}} = 2/\beta$.

Then we consider the unrestricted tractability domain

$$\Omega^{\text{unr}} = [1,\infty) \times \mathbb{N}.$$

We consider three cases of linear tensor product problems depending on the behaviour of their singular eigenvalues for the univariate case. These cases are:

- only finitely many singular values are positive,

- singular values decay exponentially fast,

- singular values decay polynomially fast.

As we know from Chapter 5, weak tractability holds for these three cases, and even for all linear tensor product problems for which the singular values decay slightly faster that logarithmically. In this chapter, we present necessary and sufficient conditions on the function T such that generalized tractability holds. These conditions are obtained in terms of the singular values for the univariate case and limiting properties of T. The tractability conditions tell us how fast T must go to infinity. As already indicated, T must go to infinity faster than polynomially. We show that generalized tractability is obtained for such functions as $T(x,y) = x^{1+\ln y}$.

We also study tractability functions T of product form, $T(x, y) = f_1(x) f_2(x)$. Assume that

$$a_i = \liminf_{x \to \infty} (\ln \ln f_i(x))/(\ln \ln x) < \infty \quad \text{for } i = 1, 2.$$

Then we obtain generalized tractability iff

$$a_i > 1 \quad \text{and} \quad (a_1 - 1)(a_2 - 1) \geq 1,$$

and if $(a_1 - 1)(a_2 - 1) = 1$ then we need to assume one more condition (8.28). If

$$(a_1 - 1)(a_2 - 1) > 1$$

then the exponent of tractability is 0, and if

$$(a_1 - 1)(a_2 - 1) = 1$$

then the exponent of tractability is positive. If T is of product form, the tractability conditions as well as the exponent of tractability depend only on the second singular eigenvalue and they do *not* depend on the rate of their decay.

Finally, we compare the results obtained for the unrestricted and restricted domains. In general, the tractability results are quite different. We may have generalized tractability for the restricted domain and no generalized tractability for the unrestricted domain which is the case, for instance, for polynomial tractability $T(x, y) = xy$. We may also have generalized tractability for both domains with different or with the same exponents of tractability.

8.1 Motivation of Generalized Tractability

The essence of tractability is to guarantee that the information complexity $n(\varepsilon, d)$ does *not* depend exponentially on ε^{-1} and d. There are various ways to measure the lack of exponential dependence. First of all, we must agree how the parameters ε and d vary. In most previous work on tractability, and so far also in this book, it was assumed that ε and d are independent, and $\varepsilon \in (0, 1]$, $d \in \mathbb{N}$. In particular, it was assumed that both ε^{-1} and d may go to infinity. For some applications, as in finance, we are interested in huge d and relatively small ε^{-1}. For instance, d may be in the hundreds or thousands, however, we may have $\varepsilon \geq 0.01$. The reason is that since financial models are relatively weak, depending for instance on future re-financing rates, there is no merit in a more accurate solution. In such cases, the assumption that both ε^{-1} and d may go to infinity is too demanding.

That is why we assume that (ε^{-1}, d) belongs to Ω, where the domain Ω is, in general, a proper subset of $[1, \infty) \times \mathbb{N}$. Obviously, the domain Ω should be big enough to properly model the essence of a given multivariate problem.

As before, we use the notation $[n] := \{1, 2, \ldots, n\}$ for any integer n. In particular, $[n] = \emptyset$ if $n \leq 0$. We assume that

$$[1, \infty) \times [d^*] \cup [1, \varepsilon_0^{-1}) \times \mathbb{N} \subseteq \Omega \tag{8.2}$$

for some $d^* \in \mathbb{N}_0$ and some $\varepsilon_0 \in (0, 1]$ such that $d^* + (1 - \varepsilon_0) > 0$. Condition (8.2) is the only restriction we impose on Ω. The constraint $d^* + (1 - \varepsilon_0) > 0$ excludes the case $d^* = 0$ and $\varepsilon_0 = 1$ corresponding to no restriction on Ω.

Polynomial tractability for multivariate problems is defined by demanding that $n(\varepsilon, d)$ is bounded by a polynomial in ε^{-1} and d. Obviously there are different ways of guaranteeing that $n(\varepsilon, d)$ does not depend exponentially on ε^{-1} and d.

For instance, in theoretical computer science, tractability for discrete problems is usually understood by demanding that the cost bound of an algorithm is a polynomial in $k = \lceil \log_2(1 + \varepsilon^{-1}) \rceil$. That is, we want to compute k correct bits of the solution in time polylog in ε^{-1}.

We note in passing that if one adopts this definition of tractability then most multivariate problems become intractable since even for the univariate case, $d = 1$, we typically have that $n(\varepsilon, 1)$ is a polynomial in ε^{-1}.

One may also take a point of view opposite to the one presented above, and consider a problem to be tractable when $n(\varepsilon, d)$ can be bounded by a function of ε^{-1} and d that grows faster than polynomials. This has been partially done in [293] by demanding that $n(\varepsilon, d)$ is bounded by a multiple of powers of $f_1(\varepsilon^{-1})$ and $f_2(d)$ with functions f_i such as $f_i(x) = \exp(\ln^{1+\alpha_i}(x))$ with $\alpha_i > 0$. Indeed, such functions grow faster than any polynomial as x tends to infinity, but slower than any exponential function a^x with $a > 1$. It was shown in [293] that the class of tractable multivariate problems is larger for such functions f_i than the tractability class studied before.

The approach of [293] is not fully general. Its notion of tractability decouples the parameters ε^{-1} and d since functions f_i depend only on one of these parameters. For some multivariate problems, such as tensor product problems, this restriction is essential. It is therefore better not to insist on independence of ε^{-1} and d, and study tractability without assuming this property.

Hence, we study tractability defined by a function T of two variables, using a multiple of a power of $T(\varepsilon^{-1}, d)$ in the definition of generalized tractability. Obviously, we need to assume that T satisfies several natural properties. First of all, the problem of computing an ε-approximation usually becomes harder as ε decreases. Furthermore, with a proper definition of the operators S_d, the problem should become harder when d increases. That is why we assume that the function T is non-decreasing in both its arguments. Moreover, to rule out the exponential behavior of T, we assume that $T(x, y)/a^{x+y}$ tends to 0 as $x + y$ tends to infinity for any $a > 1$. This is equivalent to assuming that $\ln T(x, y)/(x + y)$ tends to 0 as $x + y$ approaches infinity. As we shall see in a moment, it will be convenient to define the domain of T as the set $[1, \infty) \times [1, \infty)$. In particular, this domain allows us to say that T is non-decreasing, and will be useful for the concept of generalized strong tractability. This discussion motivates the following definitions.

A function $T : [1, \infty) \times [1, \infty) \to [1, \infty)$ is a *tractability function* if T is non-decreasing in x and y and

$$\lim_{(x,y)\in[1,\infty)\times[1,\infty),\ x+y\to\infty} \frac{\ln T(x, y)}{x + y} = 0. \tag{8.3}$$

Consider a multivariate problem $S = \{S_d\}$ in some setting for a class $\Lambda = \{\Lambda_d\}$ of information operations, and for the absolute or normalized error criterion. Here by Λ_d we denote the class of information operations used for approximation of S_d. Let $n(\varepsilon, S_d, \Lambda_d)$ denote its information complexity.

The multivariate problem S is (T, Ω)-*tractable in the class* Λ if there exist non-negative numbers C and t such that

$$n(\varepsilon, S_d, \Lambda_d) \leq C\, T(\varepsilon^{-1}, d)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega. \tag{8.4}$$

The *exponent* t^{tra} *of* (T, Ω)-*tractability* in the class Λ is defined as the infimum of all non-negative t for which there exists a $C = C(t)$ such that (8.4) holds.

Let Ω be an arbitrary domain Ω satisfying (8.2) with $\varepsilon_0 < 1$. Then it is easy to see that if

$$n(\varepsilon_0, S_d, \Lambda_d) \geq \kappa^d \quad \text{for almost all } d \in \mathbb{N} \text{ with } \kappa > 1, \tag{8.5}$$

then S is *not* (T, Ω)-tractable in the class Λ for an arbitrary tractability function T. Indeed, suppose on the contrary that S is (T, Ω)-tractable in the class Λ. Then

$$\frac{\ln C + t \ln T(\varepsilon_0^{-1}, d)}{\varepsilon_0^{-1} + d} \geq \frac{d \ln \kappa}{\varepsilon_0^{-1} + d},$$

implying that

$$\liminf_{d\to\infty} \frac{\ln T(\varepsilon_0^{-1}, d)}{\varepsilon_0^{-1} + d} \geq \frac{\ln \kappa}{t} > 0,$$

which contradicts (8.3).

Similarly, let Ω be an arbitrary domain Ω satisfying (8.2) with $d^* \geq 1$. If there exist $d \in [d^*]$ and $\kappa > 1$ such that

$$n(\varepsilon, S_d, \Lambda_d) \geq \kappa^{1/\varepsilon} \quad \text{for sufficiently small } \varepsilon, \tag{8.6}$$

then S is *not* (T, Ω)-tractable in the class Λ for an arbitrary tractability function T. As before, this follows from the fact that

$$\liminf_{\varepsilon\to0} \frac{\ln T(\varepsilon^{-1}, d)}{\varepsilon_0^{-1} + d} \geq \frac{\ln \kappa}{t} > 0,$$

which contradicts (8.3).

For some multivariate problems, it has been shown that $n(\varepsilon, S_d, \Lambda_d)$ is bounded by a multiple of some power of ε^{-1} that does not depend on d. This property is called *strong tractability*. In our case, we can define generalized strong tractability by

insisting that the bound in (8.4) is independent of d. Formally, we replace $T(\varepsilon^{-1}, d)$ by $T(\varepsilon^{-1}, 1)$. We stress that (ε^{-1}, d) from Ω does not necessarily imply that $(\varepsilon^{-1}, 1)$ is in Ω. Nevertheless, due to the more general domain of T, the value $T(\varepsilon^{-1}, 1)$ is well defined, and since T is monotonic, we have $T(\varepsilon^{-1}, 1) \le T(\varepsilon^{-1}, d)$.

The multivariate problem S is *strongly (T, Ω)-tractable* in the class Λ if there exist non-negative numbers C and t such that

$$n(\varepsilon, S_d, \Lambda_d) \le C\, T(\varepsilon^{-1}, 1)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega. \tag{8.7}$$

The *exponent t^{str} of strong (T, Ω)-tractability* in the class Λ is defined as the infimum of all non-negative t for which there exists a $C = C(t)$ such that (8.7) holds.

Clearly, strong (T, Ω)-tractability in the class Λ implies (T, Ω)-tractability in the class Λ. Furthermore, $t^{\mathrm{tra}} \le t^{\mathrm{str}}$. For some multivariate problems the exponents t^{tra} and t^{str} are the same, and for some they are different. We shall see such examples also in this chapter.

When it will cause no confusion, we simplify our notation and terminology as follows. If Ω and Λ are clear from the context, we say that S is T-tractable or strongly T-tractable. If T is also clear from the context, we say that S is tractable or strongly tractable. Finally, we talk about generalized tractability or generalized strong tractability if we consider various T, Ω and Λ.

We note in passing what happens if two tractability functions T_1 and T_2 are such that $T_1 = T_2^\alpha$ for some positive α. It is clear that the concepts of T_i-tractability are then essentially the same, with the obvious changes of their exponents. Therefore we can obtain substantially different tractability results for T_1 and T_2 only if they are *not* polynomially related.

We now introduce a couple of specific cases of generalized tractability depending on the domain Ω and the form of the function T. We begin with two examples of Ω which seem especially interesting.

- **Restricted tractability domain.** Let

$$\Omega^{\mathrm{res}} = [1, \infty) \times [d^*] \cup [1, \varepsilon_0^{-1}) \times \mathbb{N}$$

  for some $d^* \in \mathbb{N}_0$ and $\varepsilon_0 \in (0, 1]$ with $d^* + (1 - \varepsilon_0) > 0$. This corresponds to a smallest set Ω used for tractability study. This case is called a *restricted tractability domain* independently of the function T.

  We may consider the special sub-cases where $d^* = 0$ or $\varepsilon_0 = 1$. If $d^* = 0$ then $\varepsilon_0 < 1$ and we have $\Omega^{\mathrm{res}} = [1, \varepsilon_0^{-1}) \times \mathbb{N}$. Hence, we now want to compute an ε-approximation for only $\varepsilon \in (\varepsilon_0, 1]$ and for all d. We call this sub-case *restricted tractability in ε*.

  If $\varepsilon_0 = 1$ then $d^* \ge 1$ and we have $\Omega^{\mathrm{res}} = [1, \infty) \times [d^*]$. Hence, we now want to compute an ε-approximation for all $\varepsilon \in (0, 1]$ but only for $d \le d^*$. We call this sub-case *restricted tractability in d*.

- **Unrestricted tractability domain.** Let

$$\Omega^{\mathrm{unr}} = [1, \infty) \times \mathbb{N}.$$

  This corresponds to the largest set Ω used for tractability study. This case is called the *unrestricted tractability domain* independently of the function T. This domain has been used so far in the book.

We now present several examples of generalized tractability in terms of specific functions T that we think are of a particular interest.

- **Polynomial tractability.** Let

$$T(x, y) = xy.$$

  In this case $(T, \Omega^{\mathrm{unr}})$-tractability coincides with tractability previously studied in most of tractability papers as well as so far in the book. For this function T, independently of Ω, tractability means that $n(\varepsilon, S_d, \Lambda_d)$ is bounded by a polynomial in ε^{-1} and d, explaining the name.

- **Separable tractability.** Let

$$T(x, y) = f_1(x) f_2(y)$$

  with non-decreasing functions $f_1, f_2 \colon [1, \infty) \to [1, \infty)$. To guarantee (8.3) we assume that

$$\lim_{x \to \infty} \frac{\ln f_i(x)}{x} = 0 \quad \text{for } i = 1, 2.$$

  Now $(T, \Omega^{\mathrm{unr}})$-tractability coincides with the notion of (f_1, f_2)-tractability studied in [293]. For this T, independently of Ω, the roles of ε^{-1} and d are separated, explaining the name. Observe that polynomial tractability is a special case of separable tractability for $f_1(x) = f_2(x) = x$.

  For separable tractability, we can modify the condition (8.4) by taking possibly different exponents of ε^{-1} and d. That is, the problem S is (T, Ω)-tractable in the class Λ if there are non-negative numbers C, p and q such that

$$n(\varepsilon, S_d, \Lambda_d) \leq C f_1(\varepsilon^{-1})^p f_2(d)^q \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega. \qquad (8.8)$$

  The exponents p and q are called the ε-exponent and the d-exponent. We stress that, in general, they do not need to be uniquely defined. Note that we obtain (8.4) from (8.8) by taking $t = \max\{p, q\}$. Similarly, the notion of strong (T, Ω)-tractability in the class Λ is obtained if $q = 0$ in the bound above, and the exponent t^{str} is the infimum of p satisfying the bound above with $q = 0$. Again for $f_1(x) = f_2(x) = x$ these notions coincide with the notions of polynomial tractability.

- **Separable restricted tractability.** Let

$$T(x, y) := \begin{cases} f_1(x) & \text{if } (x, y) \in [1, \infty) \times [1, d^*], \\ f_2(y) & \text{if } (x, y) \in [1, \varepsilon_0^{-1}) \times \mathbb{N} \setminus [1, d^*], \\ \max\{f_1(x), f_2(y)\} & \text{otherwise,} \end{cases}$$

where f_1, f_2 are as above with $f_2(d^*) \geq f_1(\varepsilon_0^{-1})$.

It is easy to check that T is indeed a generalized tractability function. Suppose that the function T is considered over the restricted tractability domain Ω^{res}. Then (T, Ω^{res})-tractability corresponds to the smallest set Ω and we have a separate dependence on ε and d, explaining the name. As already discussed, such generalized tractability seems especially relevant for the case when for huge d we are only interested in a rough approximation to the solution.

- **Non-separable symmetric tractability.** Let

$$T(x, y) = \exp\left(f(x) f(y)\right) \tag{8.9}$$

with a non-decreasing function $f : [1, \infty) \to \mathbb{R}_+$. To guarantee (8.3) we need to assume that $\lim_{x+y \to \infty} f(x) f(y)/(x + y) = 0$. This holds, for example, if $f(x) = x^\alpha$ with $\alpha \in (0, 1/2)$ or if $f(x) = \ln^{1+\alpha}(x + 1)$ with a positive α. The tractability function corresponding to $f(x) = \ln^{1+\alpha}(x + 1)$ will be useful in the study of linear tensor product problems.

It is easy to see that this tractability function is *not* separable if f is not a constant function. Indeed, assume on the contrary that $T(x, y) = f_1(x) f_2(y)$ for some functions f_1 and f_2. For $x = 1$, we get $f_2(y) = f_1(1)^{-1} \exp(f(1) f(y))$, and similarly by taking $y = 1$, we obtain $f_1(x) = f_2(1)^{-1} \exp(f(1) f(x))$. Hence,

$$\exp(f(x) f(y)) = [f_1(1) f_2(1)]^{-1} \exp(f(1)(f(x) + f(y))).$$

Now $f_1(1) f_2(1) = \exp(f^2(1))$. Taking $x = y$, we obtain

$$f^2(x) = 2f(1) f(x) - f^2(1),$$

which leads to the conclusion that $f(x) = f(1)$ for all x. This contradicts the requirement that f is not a constant function. Thus, T is not separable. Since the roles of ε^{-1} and d are the same, this motivates the name of this generalized tractability.

We finish this subsection by an example of a function T that is *not* a tractability function. Consider $T(x, y) = \exp(y^{1-1/x})$. This function is bounded in x for fixed y and increases sub-exponentially in y for fixed x. Nevertheless,

$$\limsup_{x+y \to \infty} \frac{\ln T(x, y)}{x + y} \geq \lim_{x=y \to \infty} \frac{x^{1-1/x}}{2x} = \frac{1}{2},$$

proving that T is not a tractability function. This example shows that the notion of tractability functions does not admit functions that increase asymptotically as fast as an exponential function in some direction.

8.2 Linear Tensor Product Problems

In this section we consider multivariate problems defined as linear tensor product problems and study generalized tractability in the worst case setting mostly for the class Λ^{all}. As we know from Chapter 5, the information complexity $n(\varepsilon, d) := n(\varepsilon, S_d, \Lambda_d)$ depends entirely on the sequence of the singular values $\lambda = \{\lambda_j\}$ for the univariate case, and $\Lambda_d = \Lambda^{\text{all}}$ is the class of all continuous linear functionals used for approximation of S_d. We normalize the problem by assuming that $\lambda_1 = 1$, so that $1 \geq \lambda_2 \geq \cdots$ and $\lim_k \lambda_k = 0$. Hence, the absolute and normalized error criteria coincide. We have

$$n(\varepsilon, d) = |\{(i_1, \ldots, i_d) \in \mathbb{N}^d \mid \lambda_{i_1} \cdots \lambda_{i_d} > \varepsilon^2\}|, \tag{8.10}$$

with the convention that the cardinality of the empty set is 0. Since $n(\varepsilon/\sqrt{\lambda_j}, d-1) \leq n(\varepsilon, d-1)$ we obtain for all $d \geq 2$,

$$n(\varepsilon, 1) \leq n(\varepsilon, d-1) \leq n(\varepsilon, d) \leq n(\varepsilon, 1)^d. \tag{8.11}$$

We show a simple lemma relating generalized tractability to the sequence $\{\lambda_i\}$.

Lemma 8.1. *Let T be an arbitrary tractability function, Ω be a domain satisfying (8.2) with $\varepsilon_0 < 1$, $S = \{S_d\}$ be a linear tensor product problem, and $\Lambda = \{\Lambda_d\}$ be an arbitrary class of information operations.*

- *Let $\lambda_2 = 1$. Then S is* not *(T, Ω)-tractable in the class Λ.*

- *Let $\varepsilon_0^2 < \lambda_2 < 1$. Then S is* not *strongly (T, Ω)-tractable in the class Λ.*

- *Let $\lambda_2 = 0$. Then S is strongly (T, Ω)-tractable in the class Λ^{all} since $n(\varepsilon, d) = 1$ for all $(\varepsilon^{-1}, d) \in \Omega$ with $\varepsilon < 1$, and $t^{\text{str}} = 0$.*

Proof. Since $\Lambda_d \subseteq \Lambda^{\text{all}}$ we have $n(\varepsilon, S_d, \Lambda_d) \geq n(\varepsilon, d)$. If $\lambda_2 = 1$, then we can take $i_1, i_2, \ldots, i_d \in \{1, 2\}$ to conclude from (8.10) that

$$n(\varepsilon_0, d) \geq 2^d \quad \text{for all } d.$$

Hence (8.5) holds with $\kappa = 2$, and S is not (T, Ω)-tractable in the class Λ.

If $\varepsilon_0^2 < \lambda_2 < 1$, then we take $d-1$ values of $i_j = 1$ and one value of $i_j = 2$. Since we have at least d products of eigenvalues λ_{i_j} equal to λ_2, we get

$$n(\varepsilon_0, d) \geq d \quad \text{for all} d.$$

This contradicts strong (T, Ω)-tractability in the class Λ, since $n(\varepsilon_0, d)$ cannot be bounded by $C T(\varepsilon_0^{-1}, 1)^t$ for all d.

Finally, if $\lambda_2 = 0$ then S_1, as well as S_d, is equivalent to a bounded linear functional, which can be computed exactly using only one information evaluation. This completes the proof. $\qquad \square$

In what follows we will need a simple bound for $n(\varepsilon, d)$, which was proved in [288, Remark 3.1]. For the sake of completeness, we restate the short proof of this bound.

Lemma 8.2.

- *For $\varepsilon \in (0, 1)$ and $\lambda_2 \in (0, 1)$, let*

$$\alpha(\varepsilon) := \lceil 2 \ln(1/\varepsilon) / \ln(1/\lambda_2) \rceil - 1,$$

  $\beta(\varepsilon) := n(\varepsilon, 1)$, *and* $a := \min\{\alpha(\varepsilon), d\}$. *Then*

$$\binom{d}{a} \leq n(\varepsilon, d) \leq \binom{d}{a} \beta(\varepsilon)^a. \tag{8.12}$$

- *If $\lambda_2 \leq \varepsilon_0^2 < 1$ then*

$$n(\varepsilon, d) = 1 \quad \text{for all } \varepsilon \in [\varepsilon_0, 1) \text{ and for all } d \in \mathbb{N}. \tag{8.13}$$

Proof. Let us consider a product $\lambda_{i_1} \cdots \lambda_{i_d}$ such that $\lambda_{i_1} \cdots \lambda_{i_d} > \varepsilon^2$. Let k denote the number of indices i_j, where $j \in [d]$, with $i_j \geq 2$. Then necessarily $\lambda_2^k > \varepsilon^2$, which implies $k \leq \alpha(\varepsilon)$. Consequently we have at most a indices that are not one. From (8.10), it follows that $\beta(\varepsilon) = |\{j \mid \lambda_j > \varepsilon^2\}|$, which implies that $i_j \leq \beta(\varepsilon)$ for all $j \in [d]$. This leads to (8.12). For $\lambda_2 \leq \varepsilon_0^2$, we may assume without loss of generality that $\lambda_2 > 0$, and we have $\alpha(\varepsilon) = a = 0$ for all $\varepsilon \in [\varepsilon_0, 1)$ and for all d. Then (8.12) implies (8.13). $\qquad\square$

8.3 Restricted Tractability Domain

In this section, we study generalized tractability for the linear tensor product problem S and a restricted tractability domain

$$\Omega^{\text{res}} = [1, \infty) \times [d^*] \cup [1, \varepsilon_0^{-1}) \times \mathbb{N}$$

for $d^* \in \mathbb{N}_0$ and $\varepsilon_0 \in (0, 1]$ with $d^* + (1 - \varepsilon_0) > 0$.

We first treat the two sub-cases of restricted tractability in ε and in d. We will see that in the first case, when $d^* = 0$, the second largest eigenvalue λ_2 is the only eigenvalue that effects tractability, while in the second case, when $\varepsilon_0 = 1$, the convergence rate of the sequence λ is the important criterion for tractability. Then we consider the case of the restricted tractability domain with $d^* \geq 1$ and $\varepsilon_0 < 1$.

8.3.1 Restricted Tractability in ε

We now provide necessary and sufficient conditions for restricted tractability in ε, which we then illustrate for several tractability functions. In this subsection $\varepsilon_0 < 1$, and from Lemma 3.1 we see that we can restrict our attention to the case when $\lambda_2 < 1$.

Theorem 8.3. *Let $\varepsilon_0 < 1$ and $d^* = 0$, so that*

$$\Omega^{\mathrm{res}} = [1, \varepsilon_0^{-1}) \times \mathbb{N}.$$

Let S be a linear tensor product problem with $\lambda_2 < \lambda_1 = 1$.

- *S is strongly $(T, \Omega^{\mathrm{res}})$-tractable in the class Λ^{all} iff $\lambda_2 \leq \varepsilon_0^2$. If this holds, then $n(\varepsilon, d) = 1$ for all $(\varepsilon, d) \in [\varepsilon_0, 1) \times \mathbb{N}$, and the exponent of strong restricted tractability is $t^{\mathrm{str}} = 0$.*

- *Let $\lambda_2 > \varepsilon_0^2$. Then S is $(T, \Omega^{\mathrm{res}})$-tractable in the class Λ^{all} iff*

$$B := \liminf_{d \to \infty} \inf_{\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln d} \in (0, \infty],$$

where, as in Lemma 3.2, $\alpha(\varepsilon) = \lceil 2 \ln(1/\varepsilon) / \ln(1/\lambda_2) \rceil - 1$. If this holds then the exponent of restricted tractability is $t^{\mathrm{tra}} = 1/B$.

Proof. The first part of the lemma follows directly from Lemma 8.1 and 8.2. Before we verify the second part, we present an estimate of $n(\varepsilon, d)$. Let $\varepsilon \in [\varepsilon_0, 1)$. For $d \geq \alpha(\varepsilon)$, we get from (8.12) of Lemma 3.2,

$$n(\varepsilon, d) \leq \frac{\beta(\varepsilon)^{\alpha(\varepsilon)}}{\alpha(\varepsilon)!} d(d-1) \cdots (d - \alpha(\varepsilon) + 1) \leq C_1 d^{\alpha(\varepsilon)}, \qquad (8.14)$$

where C_1 depends only on ε_0 and S_1.

Let now $B \in (0, \infty]$. We want to show the existence of some positive C and t such that

$$n(\varepsilon, d) \leq C T(\varepsilon^{-1}, d)^t \quad \text{for all } (\varepsilon, d) \in [\varepsilon_0, 1) \times \mathbb{N}. \qquad (8.15)$$

Let $\{B_n\}$ be a sequence in $(0, B)$ that converges to B. Then we find for each $n \in \mathbb{N}$ a number $d_n \in \mathbb{N}$ such that

$$\inf_{\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln d} \geq B_n \quad \text{for all } d \geq d_n.$$

Due to (8.14), to prove (8.15) it is sufficient to show that $C_1 d^{\alpha(\varepsilon)} \leq C T(\varepsilon^{-1}, d)^t$, which is equivalent to

$$\frac{\ln(C_1/C)}{t \ln d} + \frac{\alpha(\varepsilon)}{t} \leq \frac{\ln T(\varepsilon^{-1}, d)}{\ln d}.$$

If $C \geq C_1$ and $1/t = B_n$, then for all $d \geq d_n$ and all $\varepsilon \in [\varepsilon_0, 1]$, we have

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

To make the last estimate hold for every $(\varepsilon, d) \in [\varepsilon_0, 1] \times \mathbb{N}$, we only have to increase the number C if necessary. Letting n tend to infinity, we see that $t^{\mathrm{tra}} \leq 1/B$.

Now let (8.15) hold for some positive C and t. To prove that $B \in (0, \infty]$, we apply (8.12) of Lemma 8.2 for $\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})$ and $d \geq \alpha(\varepsilon_0)$. Then

$$n(\varepsilon, d) \geq \binom{d}{\alpha(\varepsilon)} \geq \left(\frac{d}{\alpha(\varepsilon)}\right)^{\alpha(\varepsilon)} \geq C_2 d^{\alpha(\varepsilon)}$$

with[1] $C_2 = \alpha(\varepsilon_0)^{-\alpha(\varepsilon_0)}$. Thus for $\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})$ we have

$$C_2 d^{\alpha(\varepsilon)} \leq C T(\varepsilon^{-1}, d)^t \quad \text{for all } d \geq \alpha(\varepsilon_0),$$

which is equivalent to

$$\frac{\ln T(\varepsilon^{-1}, d)}{\ln d} \geq \frac{\alpha(\varepsilon)}{t} + \frac{\ln(C_2/C)}{t \ln d}.$$

The condition $\varepsilon^2 < \lambda_2$ implies $\alpha(\varepsilon) \geq 1$, and we get

$$\liminf_{d \to \infty} \inf_{\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln d} \geq \frac{1}{t}.$$

This proves that $B > 0$ and $t^{\text{tra}} \geq 1/B$, and completes the proof. $\quad\square$

We illustrate Theorem 8.3 for a number of tractability functions T, assuming that $\lambda_2 \in (\varepsilon_0^2, 1)$. In this case we do *not* have strong tractability. However, tractability depends on the particular function T.

- Polynomial tractability, $T(x, y) = xy$. Then (T, Ω^{res})-tractability in the class Λ^{all} holds with the exponent $t^{\text{tra}} = 1/B$ with

$$B = \frac{1}{\alpha(\varepsilon_0)} = \frac{1}{\lceil 2\ln(1/\varepsilon_0)/\ln(1/\lambda_2)\rceil - 1}.$$

- Separable restricted tractability, $T(x, y) = f_2(y)$ for $x, y \in [1, \varepsilon_0^{-1}] \times \mathbb{N}$, and with a non-decreasing function $f_2 : [1, \infty) \to [1, \infty)$ such that

$$\lim_{y \to \infty} \frac{\ln f_2(y)}{y} = 0.$$

Then (T, Ω^{res})-tractability in the class Λ^{all} holds iff

$$B_1 := \liminf_{d \to \infty} \frac{\ln f_2(d)}{\ln d} \in (0, \infty];$$

in this case we get $t^{\text{tra}} = 1/B$, where

$$B = \frac{B_1}{\alpha(\varepsilon_0)} = \frac{B_1}{\lceil 2\ln(1/\varepsilon_0)/\ln(1/\lambda_2)\rceil - 1}.$$

---

[1]Here we use the inequality $\binom{d}{k} \geq (d/k)^k$ for $d \geq k$, which can be easily checked by induction on d.

Note that $B_1 > 0$ iff $f_2(d)$ is at least of order d^β for some positive β. Hence, if we take $f_2(d) = \lceil \ln(d+1) \rceil$ then we do not have tractability. On the other hand, if $f_2(d) = d^\beta$ for a positive β then $B_1 = \beta$. For $f_2(d) = \exp(\ln^{1+\beta}(d))$ with $\beta > 0$, we obtain $B_1 = \infty$ and $t^{\mathrm{tra}} = 0$. This means that in this case for an arbitrarily small positive t we have

$$n(\varepsilon, d) = \mathcal{O}\left(T(\varepsilon^{-1}, d)^t\right) \quad \text{for all } \varepsilon \in [\varepsilon_0, 1], \ d \in \mathbb{N}.$$

- Non-separable symmetric tractability, $T(x, y) = \exp(f(x)f(y))$ with f as in (8.9). Then $(T, \Omega^{\mathrm{res}})$-tractability in the class Λ^{all} holds iff

$$B_2 = \liminf_{d \to \infty} \frac{f(d)}{\ln d} \in (0, \infty],$$

and the exponent $t^{\mathrm{tra}} = 1/B$, with

$$B = B_2 \inf_{\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})} \frac{f(x)}{\alpha(x)}.$$

Note that $B_2 > 0$ iff $f(d)$ is at least of order $\ln(d)$. For example, if $f(x) = \beta \ln d$ for a positive β then $B_2 = \beta$, whereas $f(d) = d^\alpha$ with $\alpha > 0$ yields $B_2 = \infty$ and $t^{\mathrm{tra}} = 0$.

8.3.2 Restricted Tractability in d

We now assume that $d^* \geq 1$ and $\varepsilon_0 = 1$ so that

$$\Omega^{\mathrm{res}} = [1, \infty) \times [d^*].$$

We provide necessary and sufficient conditions for restricted tractability in d in terms of the sequence of eigenvalues $\lambda = \{\lambda_j\}$ of the compact operator $W_1 = S_1^* S_1$. Assume first that W_1 has a finite number of positive eigenvalues λ_j. Then

$$\lim_{\varepsilon \to 0} n(\varepsilon, 1) < \infty$$

and (8.11) yields that

$$\lim_{\varepsilon \to 0} n(\varepsilon, d) \leq \left(\lim_{\varepsilon \to 0} n(\varepsilon, 1)\right)^d < \infty$$

for all d. In our case, we have $d \leq d^*$. Hence, the problem is strongly $(T, \Omega^{\mathrm{res}})$-tractable with $t^{\mathrm{str}} = 0$ for all tractability functions T, since

$$n(\varepsilon, d) \leq C := \left(\lim_{\varepsilon \to 0} n(\varepsilon, 1)\right)^{d^*} \quad \text{for all } (\varepsilon, d) \in \Omega^{\mathrm{res}}.$$

Assume then that W_1 has infinitely many positive eigenvalues λ_j which is equivalent to assuming that $\lim_{\varepsilon \to 0} n(\varepsilon, 1) = \infty$. In this case we have the following theorem.

Theorem 8.4. *Let*

$$\Omega^{\text{res}} = [1, \infty) \times [d^*] \quad \textit{with } d^* \geq 1, \textit{ and } \lim_{\varepsilon \to 0} n(\varepsilon, 1) = \infty.$$

Then the following three statements are equivalent:

(i)

$$A := \liminf_{\varepsilon \to 0} \frac{\ln T(\varepsilon^{-1}, 1)}{\ln n(\varepsilon, 1)} \in (0, \infty],$$

(ii) *S is (T, Ω^{res})-tractable in the class Λ^{all},*

(iii) *S is strongly (T, Ω^{res})-tractable in the class Λ^{all}.*

If (i) *holds then the exponent of strong (T, Ω^{res})-tractability and the exponent of (T, Ω^{res})-tractability satisfy*

$$\frac{1}{A} \leq t^{\text{tra}} \leq t^{\text{str}} \leq \frac{d^*}{A}.$$

Proof. It is enough to show that (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) $\Rightarrow$ (iii).

(iii) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (i). For $d = 1$ we now know that

$$C T(\varepsilon^{-1}, 1)^t \geq n(\varepsilon, 1)$$

for some positive C and t with $t \geq t^{\text{tra}}$. Taking logarithms we obtain

$$\frac{\ln T(\varepsilon^{-1}, 1)}{\ln n(\varepsilon, 1)} \geq \frac{1}{t} + \frac{\ln C^{-1}}{t \ln n(\varepsilon, 1)}.$$

Since $n(\varepsilon, 1)$ goes to infinity, we conclude that $A \geq 1/t > 0$, as claimed. Furthermore, $t \geq 1/A$ and since t can be arbitrarily close to t^{tra}, we have $t^{\text{tra}} \geq 1/A$.

(i) $\Rightarrow$ (iii). We now know that for any $\delta \in (0, A)$ there exists a positive ε_δ such that

$$n(\varepsilon, 1) \leq T(\varepsilon^{-1}, 1)^{1/(A-\delta)} \quad \text{for all } \varepsilon \in (0, \varepsilon_\delta].$$

Hence, there is a number $C_\delta \geq 1$ such that

$$n(\varepsilon, 1) \leq C_\delta T(\varepsilon^{-1}, 1)^{1/(A-\delta)} \quad \text{for all } \varepsilon \in (0, 1].$$

From (8.11) we obtain that

$$n(\varepsilon, d) \leq C_\delta^{d^*} T(\varepsilon^{-1}, 1)^{d^*/(A-\delta)} \quad \text{for all } \varepsilon \in (0, 1] \text{ and } d \in [d^*].$$

This proves strong tractability with the exponent at most $d^*/(A - \delta)$. Since δ can be arbitrarily small, $t^{\text{tra}} \leq t^{\text{str}} \leq d^*/A$, which completes the proof. $\qquad \square$

Theorem 8.4 states that (T, Ω^{res})-tractability is equivalent to $A > 0$, where A depends only on the behavior of the eigenvalues for $d = 1$. The condition $A > 0$ means that $\ln T(\varepsilon^{-1}, 1)$ goes to infinity at least as fast as $\ln n(\varepsilon, 1)$. Note that for a finite positive A and for $d^* > 1$, we do not have sharp bounds on the exponents. We shall see later that both bounds in Theorem 8.4 may be attained for some specific multivariate problems and tractability functions T. It may also happen that $A = \infty$. In this case $t^{\text{tra}} = t^{\text{str}} = 0$, which means that for all $d \in [d^*]$, and all positive t we have

$$n(\varepsilon, d) = o(T(\varepsilon^{-1}, d)^t) \quad \text{as } \varepsilon \to 0.$$

To verify the condition $A > 0$ and find better bounds on the exponents of tractability, we study different rates of convergence of the sequence $\lambda = \{\lambda_j\}$. We consider exponential, polynomial and logarithmic rates of λ. That is, we assume

- exponential rate: λ_j is of order $\exp(-\beta j)$ for some positive β, or a little more generally, λ_j is of order $\exp(-\beta j^\alpha)$ for some positive α and β;

- polynomial rate: λ_j is of order $j^{-\beta} = \exp(-\beta \ln j)$, or a little more generally, λ_j is of order $\exp(-\beta (\ln j)^\alpha)$ for some positive α and β;

- logarithmic rate: λ_j is of order $(\ln j)^{-\beta} = \exp(-\beta \ln \ln j)$ for some positive β.

Note that for $\alpha < 1$, we have sub-exponential or sub-polynomial behavior of the eigenvalues, whereas for $\alpha > 1$, we have super-exponential or super-polynomial decay of the eigenvalues. For the sake of simplicity we omit the prefixes sub and super and talk only about exponential or polynomial rates.

As we shall see, tractability will depend on some limits. We will denote these limits using subscripts indicating the rate of convergence of λ. Hence, the subscript e indicates an exponential rate, the subscript p a polynomial rate, and the subscript l a logarithmic one.

Exponential Rate

Theorem 8.5. *Let* $\Omega^{\text{res}} = [1, \infty) \times [d^*]$ *with* $d^* \geq 1$. *Let S be a linear tensor product problem with* $\lambda_1 = 1$ *and with exponentially decaying eigenvalues* λ_j, *so that*

$$K_1 \exp\left(-\beta_1 j^{\alpha_1}\right) \leq \lambda_j \leq K_2 \exp\left(-\beta_2 j^{\alpha_2}\right) \quad \text{for all } j \in \mathbb{N}$$

for some positive numbers α_1, α_2, β_1, β_2, K_1 *and* K_2.

Then S is (T, Ω^{res})-*tractable* (*as well as strongly* (T, Ω^{res})-*tractable due to Theorem 8.4*) *in the class* Λ^{all} *iff*

$$A_e := \liminf_{x \to \infty} \frac{\ln T(x, 1)}{\ln \ln x} \in (0, \infty].$$

If $A_e > 0$ then the exponent of $(T, \Omega^{\mathrm{res}})$-tractability satisfies

$$\frac{1}{\alpha_1} \max_{d \in [d^*]} \frac{d}{A_{e,d}} \le t^{\mathrm{tra}} \le \frac{1}{\alpha_2} \max_{d \in [d^*]} \frac{d}{A_{e,d}},$$

where

$$A_{e,d} = \liminf_{x \to \infty} \frac{\ln T(x, d)}{\ln \ln x},$$

(clearly, $A_{e,d} \ge A_{e,1} = A_e > 0$), and the exponent of strong $(T, \Omega^{\mathrm{res}})$-tractability satisfies

$$\frac{d^*}{\alpha_1 A_e} \le t^{\mathrm{str}} \le \frac{d^*}{\alpha_2 A_e}.$$

Proof. We have

$$n(\varepsilon, 1) = \min\{j \mid \lambda_{j+1} \le \varepsilon^2\}.$$

Using the estimates of λ_j we obtain

$$\min\{j \mid g_1(j) \le \varepsilon^2\} \le n(\varepsilon, 1) \le \min\{j \mid g_2(j) \le \varepsilon^2\},$$

where $g_i(j) = K_i \exp\left(-\beta_i (j + 1)^{\alpha_i}\right)$. This yields

$$\left(\frac{1}{\beta_1} \ln\left(K_1 \varepsilon^{-2}\right)\right)^{1/\alpha_1} - 1 \le n(\varepsilon, 1) \le \left(\frac{1}{\beta_2} \ln\left(K_2 \varepsilon^{-2}\right)\right)^{1/\alpha_2}.$$

For small ε this leads to

$$\frac{\ln \ln \varepsilon^{-1}}{\alpha_1} (1 + o(1)) \le \ln n(\varepsilon, 1) \le \frac{\ln \ln \varepsilon^{-1}}{\alpha_2} (1 + o(1)).$$

Therefore A from (i) of Theorem 8.4 satisfies $\alpha_2 A_e \le A \le \alpha_1 A_e$. Hence, $A > 0$ iff $A_e > 0$, and (i) of Theorem 8.4 yields the first part of Theorem 8.5.

We now find bounds on the exponents assuming that $A_e > 0$. First we estimate the information complexity $n(\varepsilon, d)$. With $x := \ln((K_2^d \varepsilon^{-2})^{1/\beta_2})$ we use (8.10) to obtain

$$n(\varepsilon, d) \le m_e(x, d) := \left|\{(i_1, \ldots, i_d) \mid \sum_{j=1}^d i_j^{\alpha_2} < x\}\right|.$$

We now prove that

$$\kappa_{d,x}\left(\left(\frac{x}{d}\right)^{1/\alpha_2} - 1\right)^d \le m_e(x, d) \le x^{d/\alpha_2}, \tag{8.16}$$

where $\kappa_{d,x} = 1$ for $x \ge d$, and $\kappa_{d,x} = 0$ for $x < d$. We prove (8.16) by induction on d. Let $\alpha = \alpha_2$. For $d = 1$ we have $m(x, 1) = |\{i \mid i < x^{1/\alpha}\}|$, i.e., $x^{1/\alpha} - 1 \le m(x, 1) < x^{1/\alpha}$. For $d > 1$, we have

$$m_e(x, d) = \sum_{k < x^{1/\alpha}} m_e(x - k^\alpha, d - 1).$$

From our induction hypothesis we get

$$m_e(x, d) \leq \sum_{k < x^{1/\alpha}} (x - k^\alpha)^{(d-1)/\alpha} \leq \sum_{k < x^{1/\alpha}} x^{(d-1)/\alpha} \leq x^{d/\alpha}.$$

To prove a lower bound, we can assume that $x > d$, so that

$$m_e(x, d) \geq \sum_{k:\, k^\alpha + d - 1 \leq x} \left(\left(\frac{x - \xi^\alpha}{d-1} \right)^{1/\alpha} - 1 \right)^{d-1}$$

$$\geq \int_1^{(x+1-d)^{1/\alpha}} \left(\left(\frac{x - \xi^\alpha}{d-1} \right)^{1/\alpha} - 1 \right)^{d-1} d\xi.$$

Since $x + 1 - d \geq x/d$ and $(x - \xi^\alpha)/(d-1) \geq x/d$ for $\xi \in [1, (x/d)^{1/\alpha}]$, we have

$$m_e(x, d) \geq \int_1^{(x/d)^{1/\alpha}} \left(\left(\frac{x}{d} \right)^{1/\alpha} - 1 \right)^{d-1} d\xi = \left(\left(\frac{x}{d} \right)^{1/\alpha} - 1 \right)^d,$$

as claimed.

Consequently, we have

$$n(\varepsilon, d) \leq \left(\frac{\ln \left(K_2^d \varepsilon^{-2} \right)}{\beta_2} \right)^{d/\alpha_2} \tag{8.17}$$

for all $\varepsilon \in (0, 1]$ and $d \in [d^*]$. Take $C^* := \sup\{(1/\beta_2)^{d/\alpha_2} \mid d \in [d^*]\}$. It is easy to see that $K_2 \geq 1$. We want to show the existence of some positive C and t such that

$$n(\varepsilon, d) \leq C^* \ln \left(K_2^{d^*} \varepsilon^{-2} \right)^{d^*/\alpha_2} \leq C T(\varepsilon^{-1}, 1)^t \tag{8.18}$$

for all $\varepsilon \in (0, 1]$. The right-hand inequality is equivalent to

$$\frac{\ln(C^*/C)}{t \ln \ln \varepsilon^{-1}} + \frac{d^*}{\alpha_2 t} \frac{\ln \ln \left(K_2^{d^*} \varepsilon^{-2} \right)}{\ln \ln \varepsilon^{-1}} \leq \frac{\ln T(\varepsilon^{-1}, 1)}{\ln \ln \varepsilon^{-1}}. \tag{8.19}$$

Let $\{A_n\}$ be a sequence in $(0, A_e)$ converging to A_e. Hence for every n there exists a positive ε_n such that

$$\frac{\ln T(\varepsilon^{-1}, 1)}{\ln \ln \varepsilon^{-1}} \geq A_n \quad \text{for all } \varepsilon \in (0, \varepsilon_n].$$

Therefore, decreasing ε_n if necessary, we obtain (8.19) for all $\varepsilon \in (0, \varepsilon_n]$ as long as we choose $C \geq C^*$ and $t > d^*/(\alpha_2 A_n)$. To establish (8.18) for all $\varepsilon \in (\varepsilon_n, 1]$, we can keep the same t and, if necessary, increase C. Hence, we have strong tractability with the exponent $t^{\text{str}} \leq d^*/(\alpha_2 A_n)$, and with n tending to infinity, we conclude that $t^{\text{str}} \leq d^*/(\alpha_2 A_e)$.

We know that the problem is also tractable. To obtain an upper bound on the exponent of tractability, we use (8.17) and we find positive C and t for which

$$n(\varepsilon, d) \leq \left(\frac{\ln(K_2^{d^*} \varepsilon^{-2})}{\beta_2} \right)^{d/\alpha_2} \leq C\,T(\varepsilon^{-1}, d)^t \quad \text{for all } d \in [d^*].$$

Proceeding as before, we conclude that $t^{\mathrm{tra}} \leq \max_{d \in [d^*]} d/(\alpha_2 A_{e,d})$.

To obtain lower bounds on the exponents, we use the estimate

$$n(\varepsilon, d) \geq \tilde{m}_e(z, d) := \left| \left\{ (i_1, \ldots, i_d) \mid \sum_{j=1}^d i_j^{\alpha_1} < z \right\} \right|,$$

where $z = z(\varepsilon, d) := \ln((K_1^d \varepsilon^{-2})^{1/\beta_1})$. For sufficiently small ε, we can use the left-hand side of (8.16) with α_2 replaced by α_1 which yields

$$n(\varepsilon, d) \geq c\,z^{d/\alpha_1} = c\left(\ln\left((K_1^d \varepsilon^{-2})^{1/\beta_1} \right) \right)^{d/\alpha_1}$$

for all $d \in [d^*]$, where c is independent of ε and d. Thus, for all $t > t^{\mathrm{str}}$ there exists a $C > 0$ such that for small ε we have the inequality

$$C\,T(\varepsilon^{-1}, 1)^t \geq c\left(\ln\left((K_1^{d^*} \varepsilon^{-2})^{1/\beta_1} \right) \right)^{d^*/\alpha_1},$$

which is equivalent to

$$\frac{\ln T(\varepsilon^{-1}, 1)}{\ln \ln \varepsilon^{-1}} \geq \frac{\ln(c/C)}{t \ln \ln \varepsilon^{-1}} + \frac{d^*}{\alpha_1 t} \frac{\ln\left(\beta_1^{-1} \ln\left(K_1^{d^*} \varepsilon^{-2} \right) \right)}{\ln \ln \varepsilon^{-1}}.$$

This implies $A_e \geq d^*/(\alpha_1 t)$, and $t^{\mathrm{str}} \geq d^*/(\alpha_1 A_e)$.

For tractability, we know that there are positive C and t such that

$$C\,T(\varepsilon^{-1}, d)^t \geq c\left(\ln\left((K_1^d \varepsilon^{-2})^{1/\beta_1} \right) \right)^{d/\alpha_1} \quad \text{for all } d \in [d^*].$$

Proceeding as before, we conclude that $t^{\mathrm{tra}} \geq \max_{d \in [d^*]} d/(\alpha_1 A_{e,d})$. This concludes the proof. ☐

For exponentially decaying eigenvalues, Theorem 8.5 states that strong tractability (and tractability) are equivalent to the condition $A_e > 0$. If we know the precise order of convergence of λ, i.e., when $\alpha_1 = \alpha_2 = \alpha > 0$, then we know the exponents of tractability and strong tractability,

$$t^{\mathrm{tra}} = \frac{1}{\alpha} \max_{d \in [d^*]} \frac{d}{A_{e,d}},$$
$$t^{\mathrm{str}} = \frac{1}{\alpha} \frac{d^*}{A_e}.$$

As we shall see it may happen that $t^{\mathrm{str}} > t^{\mathrm{tra}}$.

We now illustrate Theorem 8.5 for a number of tractability functions T.

- Polynomial tractability, $T(x, y) = xy$. Then $A_{e,d} = A_e = \infty$, and we have strong tractability with $t^{\mathrm{tra}} = t^{\mathrm{str}} = 0$.

- Separable restricted tractability, $T(x, y) = f_1(x)$ for $(x, y) \in \Omega^{\mathrm{res}}$ and a non-decreasing function

$$f_1 : [1, \infty) \to [1, \infty) \quad \text{with} \quad \lim_{x \to \infty} \frac{\ln f_1(x)}{x} = 0.$$

Then strong $(T, \Omega^{\mathrm{res}})$-tractability holds iff

$$A_{e,d} = A_e = \liminf_{x \to \infty} \frac{\ln f_1(x)}{\ln \ln x} \in (0, \infty].$$

Note that $A_e > 0$ iff $f_1(x)$ is at least of order $(\ln x)^\beta$ for some positive β. If we take $f(x) = \lceil \ln(x + 1) \rceil$ then we have strong tractability with $A_e = 1$. For $\alpha_1 = \alpha_2 = \alpha > 0$, the exponents are $t^{\mathrm{str}} = t^{\mathrm{tra}} = d^*/\alpha$.

- Non-separable symmetric tractability, $T(x, y) = \exp(f(x)f(y))$ with f as in (8.9). Then $(T, \Omega^{\mathrm{res}})$-tractability holds iff

$$A_{e,d} = f(d) \liminf_{x \to \infty} \frac{f(x)}{\ln \ln x} \in (0, \infty].$$

Hence, $A_e = A_{e,1} > 0$ iff $f(x)$ is at least of order $\beta \ln \ln x$ for some positive β. For example, if we take $f(x) = \ln^{1+\alpha}(x + 1)$ for $\alpha > -1$, then $A_{e,d} = \infty$ and $t^{\mathrm{str}} = t^{\mathrm{tra}} = 0$. For $f(x) = \beta \ln \ln(x + c)$ with $c > \exp(1) - 1$ and a positive β, we have $f(1) > 0$ and

$$A_{e,d} = f(d)\beta = \beta^2 \ln \ln(d + c).$$

For $\alpha_1 = \alpha_2 = \alpha > 0$, we now have

$$t^{\mathrm{str}} = \frac{d^*}{\alpha \beta^2 \ln \ln(1 + c)}.$$

Assume for simplicity that $d^* = 2$ and take c close to $\exp(1) - 1$. Then the maximum of the function $d / \ln \ln(d + c)$ is attained for $d = 1$, and we have

$$t^{\mathrm{tra}} = \frac{1}{\alpha \beta^2 \ln \ln(1 + c)} = \frac{t^{\mathrm{str}}}{d^*}.$$

Polynomial Rate

Theorem 8.6. *Let $\Omega^{\mathrm{res}} = [1, \infty) \times [d^*]$ with $d^* \geq 1$. Let S be a linear tensor product problem with $\lambda_1 = 1$ and with polynomially decaying eigenvalues λ_j, so that*

$$K_1 \exp\left(-\beta_1(\ln j)^\alpha\right) \leq \lambda_j \leq K_2 \exp\left(-\beta_2(\ln j)^\alpha\right) \quad \text{for all } j \in \mathbb{N}$$

for some positive numbers α, β_1, β_2, K_1 and K_2.

Then S is (T, Ω^{res})-*tractable* (*as well strongly* (T, Ω)-*tractable due to Theorem* 8.4) *in the class* Λ^{all} *iff*

$$A_p := \liminf_{x \to \infty} \frac{\ln T(x, 1)}{(\ln x)^{1/\alpha}} \in (0, \infty].$$

If $\alpha \in (0, 1]$ *and* $A_p > 0$ *then the exponents of* (T, Ω^{res})-*tractability satisfy*

$$\left(\frac{2}{\beta_1}\right)^{1/\alpha} A_p^{-1} \leq t^{\text{tra}} \leq t^{\text{str}} \leq \left(\frac{2}{\beta_2}\right)^{1/\alpha} A_p^{-1}.$$

If $\alpha \in (1, \infty)$ *and* $A_p > 0$ *then the exponent of* (T, Ω^{res})-*tractability satisfies*

$$\left(\frac{2}{\beta_1}\right)^{1/\alpha} \max_{d \in [d^*]} \frac{d^{1-1/\alpha}}{A_{p,d}} \leq t^{\text{tra}} \leq \left(\frac{2}{\beta_2}\right)^{1/\alpha} \max_{d \in [d^*]} \frac{d^{1-1/\alpha}}{A_{p,d}},$$

where

$$A_{p,d} = \liminf_{x \to \infty} \frac{\ln T(x, d)}{(\ln x)^{1/\alpha}},$$

(*clearly,* $A_{p,d} \geq A_{p,1} = A_p > 0$), *and the exponent of strong* (T, Ω^{res})-*tractability satisfies*

$$(d^*)^{1-1/\alpha} \left(\frac{2}{\beta_1}\right)^{1/\alpha} A_p^{-1} \leq t^{\text{str}} \leq (d^*)^{1-1/\alpha} \left(\frac{2}{\beta_2}\right)^{1/\alpha} A_p^{-1}.$$

Proof. We now have

$$\min\{j \mid g_1(j) \leq \varepsilon^2\} \leq n(\varepsilon, 1) \leq \min\{j \mid g_2(j) \leq \varepsilon^2\}$$

with $g_i(j) = K_i \exp\left(-\beta_i (\ln(j + 1))^\alpha\right)$. This yields

$$\exp\left((\beta_1^{-1} \ln\left(K_1 \varepsilon^{-2}\right))^{1/\alpha}\right) - 1 \leq n(\varepsilon, 1) \leq \exp\left((\beta_2^{-1} \ln\left(K_2 \varepsilon^{-2}\right))^{1/\alpha}\right).$$

For small ε this leads to

$$\left(\frac{2 \ln \varepsilon^{-1}}{\beta_1}\right)^{1/\alpha} (1 + o(1)) \leq \ln n(\varepsilon, 1) \leq \left(\frac{2 \ln \varepsilon^{-1}}{\beta_2}\right)^{1/\alpha} (1 + o(1)).$$

Hence, A from (i) of Theorem 8.4 satisfies $(\beta_2/2)^{1/\alpha} A_p \leq A \leq (\beta_1/2)^{1/\alpha} A_p$. Hence, $A > 0$ iff $A_p > 0$, and (i) of Theorem 8.4 yields the first part of Theorem 8.6, and the bound $t^{\text{tra}} \geq (2/\beta_1)^{1/\alpha} A_p^{-1}$.

We now find bounds on the exponents assuming that $A_p > 0$. First we estimate the information complexity $n(\varepsilon, d)$. With $x = x(\varepsilon, d) := \ln((K_2^d \varepsilon^{-2})^{1/\beta_2})$, we have

$$n(\varepsilon, d) \leq m_p(x, d) := \left|\left\{(i_1, \ldots, i_d) \mid \sum_{j=1}^d (\ln i_j)^\alpha < x\right\}\right|.$$

We now prove the following estimates on $m_p(x, d)$. Let $s > 1$. If $\alpha \in (0, 1]$ then there exists a positive number $C(s, d)$ such that

$$\exp\left(x^{1/\alpha}\right) - 1 \leq m_p(x, d) \leq C(s, d) \exp\left(s x^{1/\alpha}\right). \tag{8.20}$$

If $\alpha \in [1, \infty)$ then there exists a positive number $C(s, d)$ such that

$$\left(\exp \left(\left(\frac{x}{d} \right)^{1/\alpha} \right) - 1 \right)^d \leq m_p(x, d) \leq C(s, d) \exp \left(s d^{1-1/\alpha} x^{1/\alpha} \right). \qquad (8.21)$$

For $d = 1$ we have $m_p(x, 1) = |\{j \mid j < \exp(x^{1/\alpha})\}|$ and $\exp \left(x^{1/\alpha} \right) - 1 \leq m_p(x, 1) < \exp \left(x^{1/\alpha} \right)$.

We start with $\alpha \in (0, 1]$. The lower bound is already proved since $m_p(x, d) \geq m_p(x, 1)$. To obtain an upper bound on $m_p(x, d)$, we modify an argument from the proof of Theorem 3.1(ii) in [288]. Let $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$ denote, as always, the Riemann zeta function. We show by induction on d that

$$m_p(x, d) \leq \zeta(s)^{d-1} \exp \left(s x^{1/\alpha} \right).$$

Clearly, this holds for $d = 1$. Assume that our claim holds for d. Then

$$m_p(x, d + 1) = \sum_{k < \exp(x^{1/\alpha})} m_p \left(x - (\ln k)^\alpha, d \right)$$

$$\leq \zeta(s)^{d-1} \sum_{k < \exp(x^{1/\alpha})} \exp \left(s \left(x - (\ln k)^\alpha \right)^{1/\alpha} \right).$$

Since $(a - b)^{1/\alpha} \leq a^{1/\alpha} - b^{1/\alpha}$ for all $a \geq b \geq 0$ and $\alpha \in (0, 1]$, we obtain

$$m_p(x, d + 1) \leq \zeta(s)^{d-1} \sum_{k < \exp(x^{1/\alpha})} \exp \left(s x^{1/\alpha} \right) \exp(-s \ln k)$$

$$= \zeta(s)^{d-1} \exp \left(s x^{1/\alpha} \right) \sum_{k < \exp(x^{1/\alpha})} k^{-s} \leq \zeta(s)^d \exp \left(s x^{1/\alpha} \right).$$

Let now $\alpha \in [1, \infty)$. Again we proceed by induction on d. The estimate (8.21) clearly holds for $d = 1$. Assume that our claim holds for d. Again we have

$$m_p(x, d + 1) = \sum_{k < \exp(x^{1/\alpha})} m_p \left(x - (\ln k)^\alpha, d \right).$$

To get a lower bound on $m_p(x, d + 1)$, we obtain

$$m_p(x, d + 1) \geq \int_1^{\exp(x^{1/\alpha})} \left(\exp \left(\left(\frac{x - (\ln \xi)^\alpha}{d} \right)^{1/\alpha} \right) - 1 \right)^d d\xi$$

$$\geq \int_1^{\exp((x/(d+1))^{1/\alpha})} \left(\exp \left(\left(\frac{x}{d + 1} \right)^{1/\alpha} \right) - 1 \right)^d d\xi$$

$$\geq \left(\exp \left(\left(\frac{x}{d + 1} \right)^{1/\alpha} \right) - 1 \right)^{d+1}.$$

We now obtain an upper bound on $m_p(x, d + 1)$. Let $r = (1 + s)/2$. Since $r > 1$, we can use the upper bound on $m_p(x, d)$ and obtain

$$m_p(x, d + 1) \leq C(r, d)\left\{ \exp\left(rd^{1-1/\alpha}x^{1/\alpha}\right) \right.$$

$$\left. + \int_1^{\exp(x^{1/\alpha})} \exp\left(rd^{1-1/\alpha}\left(x - (\ln \xi)^\alpha\right)^{1/\alpha}\right) d\xi \right\}.$$

The substitution $z = \ln \xi$ leads to

$$\int_1^{\exp(x^{1/\alpha})} \exp\left(rd^{1-1/\alpha}\left(x - (\ln \xi)^\alpha\right)^{1/\alpha}\right) d\xi \leq \int_0^{x^{1/\alpha}} \exp(rh(z)) \, dz,$$

where $h(z) = d^{1-1/\alpha}(x - z^\alpha)^{1/\alpha} + z$. Since

$$h'(z) = 1 - d^{1-1/\alpha}\left(\frac{x}{z^\alpha} - 1\right)^{1/\alpha - 1},$$

the function h takes its maximum at $z = (x/(d + 1))^{1/\alpha}$, and we get

$$m_p(x, d + 1) \leq C(r, d)\left\{ \exp\left(rd^{1-1/\alpha}x^{1/\alpha}\right) + x^{1/\alpha} \exp\left(r(d + 1)^{1-1/\alpha}x^{1/\alpha}\right) \right\}$$

$$\leq C(r, d)(1 + x^{1/\alpha}) \exp\left(r(d + 1)^{1-1/\alpha}x^{1/\alpha}\right).$$

Since

$$a := \sup_{x > 0}(1 + x^{1/\alpha}) \exp\left(- (s - r)(d + 1)^{1-1/\alpha}x^{1/\alpha}\right)$$

$$= \sup_{x > 0}(1 + x^{1/\alpha}) \exp\left(- (s - 1)(d + 1)^{1-1/\alpha}x^{1/\alpha}/2\right) < \infty$$

we take $C(s, d + 1) = aC(r, d)$ and conclude that

$$m_p(x, d + 1) \leq C(s, d + 1) \exp\left(s(d + 1)^{1-1/\alpha}x^{1/\alpha}\right),$$

as claimed.

Let $\gamma := \max\{0, 1 - 1/\alpha\}$. Then (8.20) and (8.21) yield that for every $s > 1$ there exists a positive C_s such that

$$n(\varepsilon, d) \leq C_s \exp\left(sd^\gamma \left(\ln \varepsilon^{-2/\beta_2}\right)^{1/\alpha}\right) \tag{8.22}$$

for all $\varepsilon \in (0, 1]$ and $d \in [d^*]$. Knowing that $A_p > 0$, we want to show that

$$C_s \exp\left(s(d^*)^\gamma (\ln \varepsilon^{-2/\beta_2})^{1/\alpha}\right) \leq CT(\varepsilon^{-1}, 1)^t \tag{8.23}$$

for some positive C and t. Let $\{A_n\}$ be a sequence in $(0, A_p)$ converging to A_p. Then for every n there exists a positive ε_n such that

$$\frac{\ln T(\varepsilon^{-1}, 1)}{(\ln \varepsilon^{-1})^{1/\alpha}} \geq A_n \quad \text{for all } \varepsilon \in (0, \varepsilon_n].$$

Observe that (8.23) is equivalent to

$$\frac{s(d^*)^\gamma}{t}\left(\frac{2}{\beta_2}\right)^{1/\alpha} + \frac{\ln(C_s/C)}{t(\ln\varepsilon^{-1})^{1/\alpha}} \leq \frac{\ln T(\varepsilon^{-1},1)}{(\ln\varepsilon^{-1})^{1/\alpha}}.$$

This holds for all $\varepsilon \in (0, \varepsilon_n]$ if $t \geq s(d^*)^\gamma(2/\beta_2)^{1/\alpha}A_n^{-1}$ and $C \geq C_s$. For $\varepsilon > \varepsilon_n$ we can keep the same t and, if necessary, increase C. Hence (8.23) holds with $t = s(d^*)^\gamma(2/\beta_2)^{1/\alpha}A_n^{-1}$. Thus, S is strongly (T, Ω^{res})-tractable. Taking s arbitrarily close to 1 and letting n tend to infinity, we conclude that $t^{\text{str}} \leq (d^*)^\gamma(2/\beta_2)^{1/\alpha}A_p^{-1}$.

We now show that in the case $\alpha \in (1, \infty)$ the exponent of strong tractability satisfies $t^{\text{str}} \geq (d^*)^{1-1/\alpha}(2/\beta_1)^{1/\alpha}A_p^{-1}$. Here we use the estimate

$$n(\varepsilon, d) \geq m_p(z, d),$$

where $z = z(\varepsilon, d) := \ln((K_1^d\varepsilon^{-2})^{1/\beta_1})$. For small ε, the left-hand side of (8.21) implies that there is a positive $c(d)$ such that

$$n(\varepsilon, d) \geq c(d)\exp\left(d^{1-1/\alpha}\left(\frac{1}{\beta_1}\ln\left(K_1^d\varepsilon^{-2}\right)\right)^{1/\alpha}\right). \tag{8.24}$$

Thus for all $t > t^{\text{str}}$ there exists a $C > 0$ such that for small ε, we have

$$CT(\varepsilon^{-1}, 1)^t \geq c(d^*)\exp\left((d^*)^{1-1/\alpha}\left(\frac{1}{\beta_1}\ln\left(K_1^d\varepsilon^{-2}\right)\right)^{1/\alpha}\right),$$

which is equivalent to

$$\frac{\ln T(\varepsilon^{-1}, 1)}{(\ln\varepsilon^{-1})^{1/\alpha}} \geq \frac{\ln(c(d^*)/C)}{t(\ln\varepsilon^{-1})^{1/\alpha}} + \frac{(d^*)^{1-1/\alpha}}{t}\left(\frac{1}{\beta_1}\frac{\ln\left(K_1^d\varepsilon^{-2}\right)}{\ln\varepsilon^{-1}}\right)^{1/\alpha}.$$

Taking the limit inferior as $\varepsilon \to 0$, we obtain $A_p \geq (d^*)^{1-1/\alpha}(2/\beta_1)^{1/\alpha}t^{-1}$, and $t^{\text{str}} \geq (d^*)^{1-1/\alpha}(2/\beta_1)^{1/\alpha}A_p^{-1}$.

We finally find estimates of the exponent of tractability for $\alpha \in (1, \infty)$. We proceed similarly as before and assume that

$$CT(\varepsilon^{-1}, d)^t \geq n(\varepsilon, d) \quad \text{for all } d \in [d^*].$$

By (8.24), this implies that

$$\frac{t\ln T(\varepsilon^{-1}, d)}{(\ln\varepsilon^{-1})^{1/\alpha}} \geq d^{1-1/\alpha}\left(\frac{2}{\beta_1}\right)^{1/\alpha}(1 + o(1))$$

for small ε. This yields that $t^{\text{tra}} \geq (2/\beta_1)^{1/\alpha}\max_{d\in[d^*]} d^{1-1/\alpha}/A_{p,d}$.

To get an upper bound on t^{tra}, we use (8.22), and conclude that it is enough to find positive C and t such that

$$C_s\exp\left(sd^{1-1/\alpha}\left(\ln\varepsilon^{-2/\beta_2}\right)^{1/\alpha}\right) \leq CT(\varepsilon^{-1}, d)^t \quad \text{for all } d \in [d^*].$$

This holds for $t \geq s \max_{d \in [d*]} d^{1-1/\alpha} (2/\beta_2)^{1/\alpha} A_{p,d}^{-1}$. Since s can be arbitrarily close to one, we get that $t^{\text{tra}} \leq \max_{d \in [d*]} d^{1-1/\alpha} (2/\beta_2)^{1/\alpha} A_{p,d}^{-1}$, which completes the proof. $\qquad \square$

For polynomially decaying eigenvalues, Theorem 8.6 states that strong tractability (and tractability) are equivalent to the condition $A_p > 0$. If we know the precise order of convergence of λ, so that $\beta_1 = \beta_2 = \beta > 0$, then we know the exponents of tractability. For $\alpha \in (0, 1]$ we have

$$t^{\text{tra}} = t^{\text{str}} = \left(\frac{2}{\beta}\right)^{1/\alpha} A_p^{-1},$$

whereas for $\alpha \in (1, \infty)$ we have

$$t^{\text{tra}} = \left(\frac{2}{\beta}\right)^{1/\alpha} \max_{d \in [d*]} \frac{d^{1-1/\alpha}}{A_{p,d}},$$

$$t^{\text{str}} = (d*)^{1-1/\alpha} \left(\frac{2}{\beta}\right)^{1/\alpha} A_p^{-1}.$$

As before, it may happen that $t^{\text{str}} > t^{\text{tra}}$.

We now illustrate Theorem 8.6 for a number of tractability functions T.

- Polynomial tractability, $T(x, y) = xy$. Then $A_{p,d} = A_p$ and its value depends on α. We have $A_p = 0$ for $\alpha < 1$, and $A_p = 1$ for $\alpha = 1$, and $A_p = \infty$ for $\alpha > 1$. Hence, we have strong tractability (and tractability) iff $\alpha \geq 1$. For $\alpha > 1$, we have $t^{\text{tra}} = t^{\text{str}} = 0$, whereas for $\alpha = 1$ and $\beta_1 = \beta_2 = \beta > 0$, we have $t^{\text{tra}} = t^{\text{str}} = 2/\beta$.

- Separable restricted tractability, $T(x, y) = f_1(x)$ with f_1 as for exponential decaying eigenvalues. Then strong (T, Ω^{res})-tractability holds iff

$$A_{p,d} = A_p = \liminf_{x \to \infty} \frac{\ln f_1(x)}{(\ln x)^{1/\alpha}} \in (0, \infty].$$

Note that $A_p > 0$ iff $f_1(x)$ is at least of order $\exp\left(\eta(\ln x)^{1/\alpha}\right)$ for some positive η. If we take $f_1(x) = \exp\left(\eta(\ln x)^{1/\alpha}\right)$ then we have strong tractability with $A_p = \eta$. For $\beta_1 = \beta_2 = \beta > 0$, the exponents are $t^{\text{str}} = t^{\text{tra}} = (d*)^{(1-1/\alpha)} + (2/\beta)^{1/\alpha} \eta^{-1}$.

- Non-separable symmetric tractability, $T(x, y) = \exp(f(x)f(y))$ with f as in (8.9). Then (T, Ω^{res})-tractability holds iff

$$A_{p,d} = f(d) \liminf_{x \to \infty} \frac{f(x)}{(\ln x)^{1/\alpha}} \in (0, \infty].$$

Hence, $A_p = A_{p,1} > 0$ iff $f(x)$ is at least of order $\eta(\ln x)^{1/\alpha}$ for some positive η. For example, if we take $f(x) = \eta(\ln(x + c))^{1/\alpha}$ with a positive c, then $A_{p,d} =$

$f(d)\eta$. For a given $\alpha \in [1, \infty)$, $\beta_1 = \beta_2 = \beta > 0$, and sufficiently small c, the maximum of the function $d^{1-1/\alpha}/A_{p,d}$ is attained for $d = 1$, and we have

$$t^{\text{tra}} = \frac{2^{1/\alpha}}{\eta^2(\beta \ln(1+c))^{1/\alpha}} = \frac{t^{\text{str}}}{(d^*)^{1-1/\alpha}}.$$

Logarithmic Rate

Theorem 8.7. *Let $\Omega^{\text{res}} = [1, \infty) \times [d^*]$ with $d^* \geq 1$. Let S be a linear tensor product problem with $\lambda_1 = 1$ and with logarithmically decaying eigenvalues λ_j, so that*

$$K_1 \exp(-\beta \ln(\ln(j) + 1)) \leq \lambda_j \leq K_2 \exp(-\beta \ln(\ln(j) + 1)) \quad \text{for all } j \in \mathbb{N}$$

for some positive numbers β, K_1 and K_2.

Let $\beta \leq 2$. Then S is not (T, Ω^{res})-tractable in the class Λ^{all}.

Let $\beta > 2$. Then S is (T, Ω^{res})-tractable (as well strongly (T, Ω^{res})-tractable due to Theorem 8.4) in the class Λ^{all} iff

$$A_l := \liminf_{x \to \infty} \frac{\ln T(x, 1)}{x^{2/\beta}} \in (0, \infty].$$

If $\beta > 2$ and $A_l > 0$ then the exponent of (T, Ω^{res})-tractability satisfies

$$\max_{d \in [d^*]} \frac{K_1^{d/\beta}}{A_{l,d}} \leq t^{\text{tra}} \leq \max_{d \in [d^*]} \frac{K_2^{d/\beta}}{A_{l,d}},$$

where

$$A_{l,d} := \liminf_{x \to \infty} \frac{\ln T(x, d)}{x^{2/\beta}},$$

(clearly, $A_{l,d} \geq A_{l,1} = A_l > 0$), and the exponent of strong (T, Ω^{res})-tractability satisfies

$$\frac{K_1^{1/\beta}}{A_l} \leq t^{\text{str}} \leq \frac{K_2^{d^*/\beta}}{A_l}.$$

(Note that the numbers K_1 and K_2 must satisfy $K_1 \leq 1 \leq K_2$. Thus, if $K_1 = K_2 = 1$, we have also $K_1^{1/\beta} = K_2^{d^*/\beta}$, and the last inequality becomes an equality.)

Proof. We now have

$$\min\{j \mid g_1(j) \leq \varepsilon^2\} \leq n(\varepsilon, 1) \leq \min\{j \mid g_2(j) \leq \varepsilon^2\}$$

with $g_i(j) = K_i \exp(-\beta \ln(\ln(j + 1) + 1))$. This yields

$$\exp(K_1^{1/\beta} \varepsilon^{-2/\beta} - 1) - 1 \leq n(\varepsilon, 1) \leq \exp(K_2^{1/\beta} \varepsilon^{-2/\beta} - 1).$$

For small ε this leads to

$$K_1^{1/\beta} \varepsilon^{-2/\beta} (1 + o(1)) \leq \ln n(\varepsilon, 1) \leq K_2^{1/\beta} \varepsilon^{-2/\beta} (1 + o(1)).$$

Assume first that $\beta \leq 2$. Then

$$\liminf_{\varepsilon \to 0} \frac{\ln T(\varepsilon^{-1}, 1)}{\ln n(\varepsilon, 1)} \leq K_1^{-1/\beta} \liminf_{\varepsilon \to 0} \frac{\ln T(\varepsilon^{-1}, 1)}{\varepsilon^{-1} + 1} \frac{\varepsilon^{-1} + 1}{\varepsilon^{-2/\beta}} = 0$$

due to (8.3). Therefore A from (i) of Theorem 8.4 is 0, and we do not have tractability, as claimed.

Assume then that $\beta > 2$. Then $K_2^{-1/\beta} A_l \leq A \leq K_1^{-1/\beta} A_l$. Hence, $A > 0$ iff $A_l > 0$, and (i) of Theorem 8.4 yields the first part of Theorem 8.6, and that $t^{\mathrm{str}} \geq K_1^{1/\beta} A_l^{-1}$.

We now find bounds on the exponents, assuming that $A_l > 0$. First we estimate the information complexity $n(\varepsilon, d)$. With $x = x(\varepsilon, d) := \ln((K_2^d / \varepsilon^2)^{1/\beta})$ we get

$$n(\varepsilon, d) \leq m_l(x, d) := \left| \left\{ (i_1, \ldots, i_d) \mid \sum_{j=1}^d \ln(\ln(i_j) + 1) < x \right\} \right|.$$

We prove that for every $s > 1$, there exists a positive number $C(s, d)$ such that

$$\exp(\exp(x) - 1) - 1 \leq m_l(x, d) \leq C(s, d) \exp\big(s(\exp(x) - 1)\big). \tag{8.25}$$

Let $\eta := \exp(x)$. Clearly we have $m_l(x, 1) = |\{j \mid j < \exp(\eta - 1)\}|$, which implies that

$$\exp(\eta - 1) - 1 \leq m_l(x, 1) \leq \exp(\eta - 1).$$

Let now $d \geq 1$ and assume that (8.25) holds for d. Then

$$m_l(x, d + 1) = \sum_{k < \exp(\eta - 1)} m_l\big(x - \ln(\ln(k) + 1), d\big).$$

Thus, we get the trivial lower bound estimate

$$m_l(x, d + 1) \geq m_l(x, d) \geq \exp(\eta - 1) - 1.$$

We now obtain an upper bound on $m_l(x, d + 1)$. Let $r = (1 + s)/2$. Then

$$m_l(x, d + 1) \leq C(r, d) \bigg\{ \exp(r(\eta - 1)) $$
$$+ \int_1^{\exp(\eta - 1)} \exp\big(r\big(\exp\big(x - \ln(\ln(\xi) + 1)\big) - 1\big)\big) \, d\xi \bigg\}.$$

The last integral is of the form

$$\int_1^{\exp(\eta - 1)} \exp\Big(r\Big(\frac{\eta}{\ln(\xi) + 1} - 1\Big)\Big) d\xi = \int_1^\eta \exp(rh(z)) \, dz,$$

where $z = \ln(\xi) + 1$, and $h \colon [1, \eta] \to \mathbb{R}$ with $h(z) = \eta/z + z/r - (1 + 1/r)$. It is easy to check that h takes its maximum $\eta - 1$ at the point $z = 1$. So we have

$$\int_1^\eta \exp(rh(z)) \, dz \leq (\eta - 1) \exp(r(\eta - 1)).$$

This implies that

$$
\begin{aligned}
m_l(x, d+1) &\leq C(r,d)\eta \exp(r(\eta-1)) \\
&= C(r,d)\eta \exp\big((r-s)(\eta-1)\big)\exp\big(s(\eta-1)\big) \\
&\leq C(r,d)\big(\sup_{\xi\geq 1}\xi \exp\big(-(s-1)(\xi-1)/2\big)\big)\exp\big(s(\eta-1)\big) \\
&\leq C(s,d+1)\exp(s(\eta-1))
\end{aligned}
$$

for suitably large $C(s, d+1)$, as claimed.

Due to (8.25), we conclude that

$$
n(\varepsilon, d) \leq C(s,d)\exp\big(s(K_2^{d/\beta}\varepsilon^{-2/\beta}-1)\big).
$$

For $A_l \in (0, \infty]$, $\varepsilon \in (0,1]$ and $d \in [d^*]$, we want to show that

$$
C(s,d)\exp\big(s(K_2^{d/\beta}\varepsilon^{-2/\beta}-1)\big) \leq CT(\varepsilon^{-1},1)^t \tag{8.26}
$$

for some positive C and t. Therefore let $\{A_n\}$ be a sequence in $(0, A_l)$ converging to A_l. Thus for every n there exists a positive ε_n such that

$$
\frac{\ln T(\varepsilon^{-1},1)}{\varepsilon^{-2/\beta}} \geq A_n \quad \text{for all } \varepsilon \in (0, \varepsilon_n].
$$

Then (8.26) is equivalent to

$$
\frac{sK_2^{d/\beta}}{t} + \frac{\ln(C(s,d)/C)-s}{t\varepsilon^{-2/\beta}} \leq \frac{\ln T(\varepsilon^{-1},1)}{\varepsilon^{-2/\beta}}.
$$

This holds for all $\varepsilon \in (0, \varepsilon_n]$ if $t \geq sK_2^{d/\beta}A_n^{-1}$ and $C \geq C(s,d)$. For $\varepsilon \in (\varepsilon_n, 1]$ we can keep the same t and, if necessary, increase C. Letting s tend to 1 and n tend to infinity, we conclude $t^{\mathrm{str}} \leq K_2^{d^*/\beta}A_l^{-1}$.

We can similarly show bounds on t^{tra}, since $(\ln T(\varepsilon^{-1}, d))/\varepsilon^{-2/\beta}$ is arbitrarily close to $A_{l,d}$ for small ε. This leads to $t^{\mathrm{tra}} \leq \max_{d\in[d^*]}K_2^{d/\beta}/A_{l,d}$. To get a lower bound on t^{tra}, we use the left-hand side inequality in (8.25) to conclude that

$$
n(\varepsilon, d) \geq \exp\big(K_1^{d/\beta}\varepsilon^{-2/\beta}-1\big)-1.
$$

This yields that $t^{\mathrm{tra}} \geq \max_{d\in[d^*]}K_1^{d/\beta}/A_{l,d}$, and completes the proof. $\qquad\square$

For logarithmically decaying eigenvalues, Theorem 8.7 states that for $\beta \leq 2$, we do not have tractability. This means that the eigenvalues λ_j converge to 0 too slowly, no matter how we choose the tractability function T. For $\beta > 2$, strong tractability (and tractability) are equivalent to the condition $A_l > 0$. In this case, and for $K_1 = K_2 = 1$, we know the exponents of tractability satisfy

$$
t^{\mathrm{tra}} = t^{\mathrm{str}} = A_l^{-1}.
$$

We now illustrate Theorem 8.7 for a number of tractability functions T.

- Polynomial tractability, $T(x, y) = xy$. Then for $\beta > 2$, we have $A_{l,d} = A_l = 0$. Hence, strong tractability (and tractability) does *not* hold.

- Separable restricted tractability, $T(x, y) = f_1(x)$ with f_1 as for exponential decaying eigenvalues. Let $\beta > 2$. Then strong $(T, \Omega^{\mathrm{res}})$-tractability holds iff

$$A_{l,d} = A_l = \liminf_{x \to \infty} \frac{\ln f_1(x)}{x^{2/\beta}} \in (0, \infty].$$

  Note that $A_l > 0$ iff $\ln f_1(x)$ is at least of order x^α with $\alpha \in [2/\beta, 1)$. If we take $f_1(x) = \exp(x^\alpha)$ then we have strong tractability with $A_l = 0$ for $\alpha \in (2/\beta, 1)$, and then $t^{\mathrm{str}} = t^{\mathrm{tra}} = 0$, whereas $A_l = 1$ for $\alpha = 2/\beta$ and $t^{\mathrm{str}} = t^{\mathrm{tra}} = 1$ for $K_2 = K_1 = 1$.

- Non-separable symmetric tractability, $T(x, y) = \exp(f(x)f(y))$ with f as in (8.9). For $\beta > 2$, $(T, \Omega^{\mathrm{res}})$-tractability holds iff

$$A_{l,d} = f(d) \liminf_{x \to \infty} \frac{f(x)}{x^{2/\beta}} \in (0, \infty].$$

  Hence, $A_l = A_{l,1} > 0$ iff $f(x)$ is at least of order $x^{2/\beta}$. For example, if we take $f(x) = x^{2/\beta}$ then $A_{l,d} = f(d)$. For $K_2 \leq \exp(1/d^*)$, the maximum of the function $K_2^{d/\beta}/f(d)$ is attained for $d = 1$, and $t^{\mathrm{tra}} \leq K_2^{1/\beta}$ and $t^{\mathrm{str}} \leq \exp(1)^{1/\beta}$.

8.3.3 Restricted Tractability with $d^* \geq 1$ and $\varepsilon_0 < 1$

Based on the results for restricted tractability in ε and d, it is easy to study restricted tractability with $d^* \geq 1$ and $\varepsilon \in (0, 1)$. In this subsection we let

$$\Omega^{\mathrm{res}} = \Omega^{\mathrm{res}}(\varepsilon_0, d^*) = [1, \infty) \times [d^*] \cup [1, \varepsilon_0^{-1}) \times \mathbb{N}$$

for $d^* \in \mathbb{N}_0$ and $\varepsilon_0 \in (0, 1]$.

Hence, restricted tractability in ε corresponds to $\Omega^{\mathrm{res}}(\varepsilon_0, 0) = [1, \varepsilon_0^{-1}) \times \mathbb{N}$ with $\varepsilon_0 \in (0, 1)$, and restricted tractability in d corresponds to $\Omega(1, d^*) = [1, \infty) \times [d^*]$ with $d^* \geq 1$.

Since $\Omega^{\mathrm{res}}(\varepsilon_0, d^*) = \Omega^{\mathrm{res}}(\varepsilon_0, 0) \cup \Omega^{\mathrm{res}}(1, d^*)$, it is obvious that strong tractability and tractability for $d^* \geq 1$ and $\varepsilon_0 \in (0, 1)$ are equivalent to restricted strong tractability and tractability in ε and d, respectively. We summarize this simple fact in the following lemma.

Lemma 8.8. *Let $d^* \geq 1$ and $\varepsilon_0 \in (0, 1)$. Let S be a linear tensor product problem with $\lambda_1 = 1$. Then*

- *S is strongly $(T, \Omega^{\mathrm{res}}(\varepsilon_0, d^*))$-tractable in the class Λ^{all} iff S is strongly $(T, \Omega^{\mathrm{res}}(\varepsilon_0, 0))$- and strongly $(T, \Omega^{\mathrm{res}}(1, d^*))$-tractable in the class Λ^{all}.*

- S is $(T, \Omega^{\mathrm{res}}(\varepsilon_0, d^*))$-tractable in the class Λ^{all} iff S is $(T, \Omega^{\mathrm{res}}(\varepsilon_0, 0))$- and $(T, \Omega^{\mathrm{res}}(1, d^*))$-tractable in the class Λ^{all}.

- The exponents of strong tractability and tractability for $\Omega^{\mathrm{res}}(\varepsilon_0, d^*)$ are the respective maxima of the exponents for $\Omega^{\mathrm{res}}(\varepsilon_0, 0)$ and $\Omega^{\mathrm{res}}(1, d^*)$.

Proof. It is obviously enough to show that (strong) tractability for $\Omega^{\mathrm{res}}(\varepsilon_0, 0)$ and $\Omega^{\mathrm{res}}(1, d^*)$ imply (strong) tractability for $\Omega^{\mathrm{res}}(\varepsilon_0, d^*)$. Let us consider only tractability since the reasoning for strong tractability is the same. We have

$$n(\varepsilon, d) \le C_1 T(\varepsilon^{-1}, d)^{t_1} \quad \text{for all } (\varepsilon, d) \in \Omega^{\mathrm{res}}(\varepsilon_0, 0),$$
$$n(\varepsilon, d) \le C_2 T(\varepsilon^{-1}, d)^{t_2} \quad \text{for all } (\varepsilon, d) \in \Omega^{\mathrm{res}}(1, d^*),$$

for some positive C_1, C_2, t_1 and t_2. Furthermore, we can take t_i arbitrarily close to the exponents of tractability t_i^{tra} for $i = 1, 2$.

For $(\varepsilon, d) \in \Omega^{\mathrm{res}}(\varepsilon_0, d^*)$ we have

$$(\varepsilon, d) \in \Omega^{\mathrm{res}}(\varepsilon_0, 0) \quad \text{if } d > d^*, \text{ and}$$
$$(\varepsilon, d) \in \Omega^{\mathrm{res}}(1, d^*) \quad \text{if } d \le d^*.$$

Since $T(\varepsilon^{-1}, d) \ge 1$, we then have

$$n(\varepsilon, d) \le \max(C_1, C_2) T(\varepsilon^{-1}, d)^{\max(t_1, t_2)} \quad \text{for all } (\varepsilon, d) \in \Omega^{\mathrm{res}}(\varepsilon_0, d^*).$$

This implies tractability with the exponent $t^{\mathrm{tra}} \le \max(t_1, t_2)$. If we take t_i tending to t_i^{tra}, then $t^{\mathrm{tra}} \le \max(t_1^{\mathrm{tra}}, t_2^{\mathrm{tra}})$. The last bound is sharp since for $(\varepsilon, d) \in \Omega^{\mathrm{res}}(\varepsilon_0, 0)$ we must have $t^{\mathrm{tra}} \ge t_1^{\mathrm{tra}}$, and for $(\varepsilon, d) \in \Omega^{\mathrm{res}}(1, d^*)$ we must have $t^{\mathrm{tra}} \ge t_2^{\mathrm{tra}}$. This completes the proof. $\square$

We now combine the results of the previous subsections and present two theorems on the tractability of S for $\Omega^{\mathrm{res}}(\varepsilon_0, d^*)$. In these theorems, strong tractability of S means that S is strongly $(T, \Omega^{\mathrm{res}}(\varepsilon_0, d^*))$-tractable in the class Λ^{all}, and tractability of S means that S is $(T, \Omega^{\mathrm{res}}(\varepsilon_0, d^*))$-tractable in the class Λ^{all}.

Theorem 8.9. *Let $d^* \ge 1$ and $\varepsilon_0 \in (0, 1)$. Let S be a linear tensor product problem with $\lambda_1 = 1$.*

- *Let $\lambda_2 = 1$. Then S is not tractable.*

- *Let $\varepsilon_0^2 < \lambda_2 < 1$. Then S is not strongly tractable, and S is tractable iff*

$$A = \liminf_{\varepsilon \to 0} \frac{\ln T(\varepsilon^{-1}, 1)}{\ln n(\varepsilon, 1)} \in (0, \infty],$$

$$B = \liminf_{d \to \infty} \inf_{\varepsilon \in [\varepsilon_0, \sqrt{\lambda_2})} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln d} \in (0, \infty],$$

where $\alpha(\varepsilon) = \lceil 2 \ln(1/\varepsilon) / \ln(1/\lambda_2) \rceil - 1$.

If $A > 0$ and $B > 0$ then

$$\max\left(A^{-1}, B^{-1}\right) \le t^{\mathrm{tra}} \le \max\left(d^* A^{-1}, B^{-1}\right).$$

- *Let $0 < \lambda_2 \le \varepsilon_0^2$.*

  Let $\lim_{\varepsilon \to 0} n(\varepsilon, 1) < \infty$. Then S is strongly tractable and $t^{\mathrm{str}} = 0$.

  Let $\lim_{\varepsilon \to 0} n(\varepsilon, 1) = \infty$. Then S is strongly tractable iff S is tractable iff

$$A = \liminf_{\varepsilon \to 0} \; \frac{\ln T(\varepsilon^{-1}, 1)}{\ln n(\varepsilon, 1)} \; \in \; (0, \infty].$$

  If $A > 0$ then

$$A^{-1} \le t^{\mathrm{tra}} \le t^{\mathrm{str}} \le d^* A^{-1}.$$

- *Let $\lambda_2 = 0$. Then $n(\varepsilon, d) = 1$ for all $(\varepsilon, d) \in \Omega(\varepsilon_0, d^*)$, and S is strongly tractable with $t^{\mathrm{str}} = 0$.*

Proof. For $\lambda_2 = 1$, it is enough to apply the first part of Lemma 8.1.

Let $\varepsilon_0^2 < \lambda_2 < 1$. The lack of strong tractability follows from the second part of Lemma 8.1. Tractability in ε holds iff $B \in (0, \infty]$ due to the second part of Theorem 8.3. Let $\lim_{\varepsilon \to 0} n(\varepsilon, 1) < \infty$. Then tractability in d holds and, in this case, $A \in (0, \infty]$, due to the reasoning before Theorem 8.4. Let $\lim_{\varepsilon \to 0} n(\varepsilon, 1) = \infty$. Then tractability in d holds iff $A \in (0, \infty]$ due to Theorem 8.4. Hence, Lemma 8.8 implies that S is tractable iff both $A, B \in (0, \infty]$. The bounds on t^{tra} now follow from Theorems 8.3 and 8.4 along with Lemma 8.8.

For $0 < \lambda_2 \le \varepsilon_0^2$ and $\lim_{\varepsilon \to 0} n(\varepsilon, 1) < \infty$, we conclude that S is strongly tractable due to the first part of Theorem 8.3, the reasoning before Theorem 8.4 and Lemma 8.8. In this case, $t^{\mathrm{str}} = 0$.

For $0 < \lambda_2 \le \varepsilon_0^2$ and $\lim_{\varepsilon \to 0} n(\varepsilon, 1) = \infty$, strong tractability in ε holds with $t^{\mathrm{str}} = 0$ due to the first part of Theorem 8.3, and strong tractability in d is equivalent to tractability in d and equivalent to $A \in (0, \infty]$ due to Theorem 8.4. This and Lemma 8.8 yield that S is strongly tractable iff S is tractable iff $A \in (0, \infty]$. The bounds on t^{tra} and t^{str} follow from Theorem 8.4.

For $\lambda_2 = 0$, the problem is trivial due to the last part of Lemma 8.1. $\qquad\square$

We now summarize tractability conditions for $\Omega(\varepsilon_0, d^*)$, assuming the specific rates of convergence of the eigenvalues $\lambda = \{\lambda_j\}$ as discussed in Theorems 8.5, 8.6 and 8.7.

Theorem 8.10. *Let $d^* \ge 1$ and $\varepsilon_0 \in (0, 1)$. Let S be a linear tensor product problem with $\lambda_2 < \lambda_1 = 1$.*

- *Let $\lambda_j = \Theta\big(\exp(-\beta j^\alpha)\big)$ converge to 0 with an exponential rate for some positive α and β.*

  - *Let $\varepsilon_0^2 < \lambda_2$. Then S is not strongly tractable, and S is tractable iff $A_e = A_{e,1} \in (0, \infty]$ and $B \in (0, \infty]$ with*

$$A_{e,d} = \liminf_{x \to \infty} \frac{\ln T(x, d)}{\ln \ln x} \in (0, \infty],$$

and B as in Theorem 8.9. *If $A_e > 0$ and $B > 0$ then*

$$t^{\text{tra}} = \max\left(\frac{1}{\alpha} \max_{d \in [d^*]} \frac{d}{A_{e,d}} , \frac{1}{B}\right).$$

– *Let $\lambda_2 \leq \varepsilon_0^2$. Then S is strongly tractable iff $A_e \in (0, \infty]$. If $A_e > 0$ then*

$$t^{\text{str}} = \frac{d^*}{\alpha A_e} \quad \text{and} \quad t^{\text{tra}} = \frac{1}{\alpha} \max_{d \in [d^*]} \frac{d}{A_{e,d}}.$$

- *Let $\lambda_j = \Theta\big(\exp(-\beta(\ln j)^\alpha)\big)$ converge to 0 with a polynomial rate for some positive α and β.*

  – *Let $\varepsilon_0^2 < \lambda_2$. Then S is not strongly tractable, and S is tractable iff $A_p = A_{p,1} \in (0, \infty]$ and $B \in (0, \infty]$ with*

  $$A_{p,d} = \liminf_{x \to \infty} \frac{\ln T(x, d)}{(\ln x)^{1/\alpha}} \in (0, \infty],$$

  and B as in Theorem 8.9. *If $A_p > 0$ and $B > 0$ then*

  $$t^{\text{tra}} = \max\left(\left(\frac{2}{\beta}\right)^{1/\alpha} \max_{d \in [d^*]} \frac{d^{(1-1/\alpha)+}}{A_{p,d}}, \frac{1}{B}\right).$$

  – *Let $\lambda_2 \leq \varepsilon_0^2$. Then S is strongly tractable iff $A_p \in (0, \infty]$. If $A_p > 0$ then*

  $$t^{\text{str}} = \left(\frac{2}{\beta}\right)^{1/\alpha} \frac{(d^*)^{(1-1/\alpha)+}}{A_p} \quad \text{and} \quad t^{\text{tra}} = \left(\frac{2}{\beta}\right)^{1/\alpha} \max_{d \in [d^*]} \frac{d^{(1-1/\alpha)+}}{A_{p,d}}.$$

- *Let $\lambda_j = \exp\big(-\beta\big(\ln(\ln(j) + 1)\big)\big)$ converge to 0 with a logarithmic rate for some positive β. For $\beta \leq 2$, S is not tractable. For $\beta > 2$, we have the following.*

  – *Let $\varepsilon_0^2 < \lambda_2$. Then S is not strongly tractable, and S is tractable iff $A_l \in (0, \infty]$ and $B \in (0, \infty]$ with*

  $$A_l = \liminf_{x \to \infty} \frac{\ln T(x, 1)}{x^{2/\beta}} \in (0, \infty]$$

  and B as in Theorem 8.9. *If $A_l > 0$ and $B > 0$ then*

  $$t^{\text{tra}} = \max\left(\frac{1}{A_l}, \frac{1}{B}\right).$$

  – *Let $\lambda_2 \leq \varepsilon_0^2$. Then S is strongly tractable iff $A_l \in (0, \infty]$. If $A_l > 0$ then*

  $$t^{\text{str}} = t^{\text{tra}} = \frac{1}{A_l}.$$

Proof. For the exponential rate and $\varepsilon_0^2 < \lambda_2$, the lack of strong tractability follows from Theorem 8.9, whereas tractability is equivalent to A_e, $B \in (0, \infty]$ due to Theorems 8.5 and 8.3. The formula for t^{tra} also follows from these two theorems and Lemma 8.8.

For the exponential rate and $\lambda_2 \le \varepsilon_0^2$, strong tractability in ε trivially holds, and strong tractability in d holds iff $A_e > 0$ due to Theorem 8.5. The formulas for t^{str} and t^{tra} are also from Theorem 8.5.

For the polynomial and logarithmic rates, we proceed in the same way and use Theorem 8.6 for the polynomial case, and Theorem 8.7 for the logarithmic case, instead of Theorem 8.5. $\qquad\square$

We illustrate Theorems 8.9 and 8.10 for a number of tractability functions T.

- Polynomial tractability, $T(x, y) = xy$. Then $A_{e,d} = A_{p,d} = \infty$ for $\alpha > 1$, whereas $A_{p,d} = 1$ if $\alpha = 1$, and $A_{p,d} = 0$ for $\alpha < 1$. Finally, $A_{l,d} = 0$ for $\beta > 2$. Hence, for logarithmically and polynomially decaying eigenvalues with $\alpha < 1$, S is not tractable.

  Let $\varepsilon_0^2 < \lambda_2$. We have $B = 1/\alpha(\varepsilon_0)$. Then for exponentially and polynomially decaying eigenvalues with $\alpha > 1$, S is not strongly tractable but is tractable with the exponent
  $$t^{\mathrm{tra}} = \alpha(\varepsilon_0) = \lceil 2 \ln(1/\varepsilon_0)/\ln(1/\lambda_2) \rceil - 1.$$

  For polynomially decaying eigenvalues with $\alpha = 1$, S is not strongly tractable but is tractable with the exponent
  $$t^{\mathrm{tra}} = \max\left(\frac{2}{\beta}, \alpha(\varepsilon_0)\right).$$

  Let $\lambda_2 \le \varepsilon_0^2$. Then for exponentially and polynomially decaying eigenvalues with $\alpha > 1$, S is strongly tractable with $t^{\mathrm{str}} = 0$. For polynomially decaying eigenvalues with $\alpha = 1$, S is strongly tractable with $t^{\mathrm{str}} = 2/\beta$.

- Separable restricted tractability, $T(x, y) = f_1(x)$ for $(x, y) \in \Omega(1, d^*)$, and $T(x, y) = f_2(y)$ for $(x, y) \in \Omega(\varepsilon_0, 0)$ with non-decreasing f_1 and f_2 such that $\lim_{t \to \infty}(\ln f_i(t))/t = 0$.

  For simplicity, let us take $f_i(t) = \exp\left((\ln t)^{\alpha_i}\right)$ for some positive α_i. Then $A_{e,d} = \infty$, whereas $A_{p,d} = \infty$ if $\alpha_1 > 1/\alpha$, and $A_{p,d} = 1$ if $\alpha_1 = 1/\alpha$, and $A_{p,d} = 0$ if $\alpha_1 < 1/\alpha$. Finally, $A_{l,d} = 0$. Hence, for polynomially decaying eigenvalues with $\alpha_1 < 1/\alpha$, and for logarithmically decaying eigenvalues S is not tractable.

  Let $\varepsilon_0^2 < \lambda_2$. If $\alpha_2 < 1$, then $B = 0$ and S is not tractable. Let $\alpha_2 \ge 1$. Then for exponentially and polynomially decaying eigenvalues with $\alpha_1 > 1/\alpha$, S is not strongly tractable but S is tractable. The exponent of tractability is $t^{\mathrm{tra}} = \alpha(\varepsilon_0)$ if $\alpha_2 = 1$ and $t^{\mathrm{tra}} = 0$ if $\alpha_2 > 1$. For polynomially decaying eigenvalues with

$\alpha_1 = 1/\alpha$, S is not strongly tractable but is tractable with exponent

$$t^{\mathrm{tra}} = \max\left\{\left(\frac{2}{\beta}\right)^{1/\alpha} (d^*)^{(1-1/\alpha)_+}, \alpha(\varepsilon_0)\right\} \quad \text{if } \alpha_2 = 1$$

and

$$t^{\mathrm{tra}} = \left(\frac{2}{\beta}\right)^{1/\alpha} (d^*)^{(1-1/\alpha)_+} \quad \text{if } \alpha_2 > 1.$$

Let $\lambda_2 \le \varepsilon_0^2$. Then for exponentially and polynomially decaying eigenvalues with $\alpha_1 > 1/\alpha$, S is strongly tractable and $t^{\mathrm{str}} = 0$. For polynomially decaying eigenvalues with $\alpha_1 = 1/\alpha$, S is strongly tractable with

$$t^{\mathrm{tra}} = t^{\mathrm{str}} = \left(\frac{2}{\beta}\right)^{1/\alpha} (d^*)^{(1-1/\alpha)_+}.$$

- Non-separable symmetric tractability, $T(x, y) = \exp(f(x)f(y))$ with f as in (8.9). For simplicity, let us take $f(x) = (\ln(x+1))^{\eta}$ for some positive η. Then $A_{e,d} = \infty$, whereas $A_{p,d} = \infty$ for $\eta > 1/\alpha$, and $A_{p,d} = f(d)$ for $\eta = 1/\alpha$, and $A_{p,d} = 0$ for $\eta < 1/\alpha$. Finally, $A_{l,d} = 0$. Hence, S is not tractable for logarithmically and polynomially decaying eigenvalues with $\eta < 1/\alpha$.

Let $\varepsilon_0^2 < \lambda_2$. If $\eta < 1$, then $B = 0$ and S is not tractable. Let $\eta \ge 1$. Then for exponentially and polynomially decaying eigenvalues with $\eta > 1/\alpha$, S is not strongly tractable but S is tractable. In the case $\eta > 1$ we have $t^{\mathrm{tra}} = 0$. For polynomially decaying eigenvalues with $\eta = 1/\alpha$, S is not strongly tractable but tractable. If we have $\eta > 1$, then $\alpha \in (0, 1)$ and

$$t^{\mathrm{tra}} = \left(\frac{2}{\beta \ln 2}\right)^{1/\alpha}.$$

Let $\lambda_2 \le \varepsilon_0^2$. Then for exponentially and polynomially decaying eigenvalues with $\eta > 1/\alpha$, S is strongly tractable and $t^{\mathrm{str}} = 0$. For polynomially decaying eigenvalues with $\eta = 1/\alpha$, S is strongly tractable with

$$t^{\mathrm{str}} = (d^*)^{(1-1/\alpha)_+} \left(\frac{2}{\beta \ln 2}\right)^{1/\alpha}, \quad t^{\mathrm{tra}} = \left(\frac{2}{\beta}\right)^{1/\alpha} \max_{d \in [d^*]} \frac{d^{(1-1/\alpha)_+}}{(\ln(d+1))^{1/\alpha}}.$$

8.4 Unrestricted Tractability Domain

In this section, we study generalized tractability for the unrestricted tractability domain

$$\Omega^{\mathrm{unr}} = [1, \infty) \times \mathbb{N}.$$

We consider linear tensor product problems for the class Λ^{all} in the worst case setting, and the sequence of the singular values $\lambda = \{\lambda_j\}$ for the univariate case. As before, we assume that $1 = \lambda_1 \geq \lambda_2 \geq \cdots$ with $\lim_j \lambda_j = 0$. We analyze three cases of the singular values in the three subsequent subsections.

8.4.1 Finitely Many Eigenvalues

In this section we consider the case when we have only finitely many positive eigenvalues λ_j. First we consider the case where we have $k \geq 2$ eigenvalues different from 0 and $k - 1$ of them are equal. We need an auxiliary lemma which will be helpful in the course of the proof of the first theorem.

Lemma 8.11. *Let $d, k \in \mathbb{N}$ and let α be an integer satisfying $0 \leq \alpha \leq \frac{k-1}{k}(d + 1)$. Then*

$$\max_{0 \leq \nu \leq \alpha} \binom{d}{\nu}(k - 1)^\nu = \binom{d}{\alpha}(k - 1)^\alpha. \tag{8.27}$$

Proof. For $0 \leq \nu \leq \alpha$, the inequality

$$\binom{d}{\nu - 1}(k - 1)^{\nu-1} \leq \binom{d}{\nu}(k - 1)^\nu$$

holds iff $\nu \leq (d - \nu + 1)(k - 1)$, and the last inequality holds iff $\nu \leq \frac{k-1}{k}(d + 1)$. This shows that the function

$$\nu \mapsto \binom{d}{\nu}(k - 1)^\nu$$

is non-decreasing on $[0, \alpha] \cap \mathbb{N}$. $\square$

Theorem 8.12. *Let T be a tractability function. Let*

$$\lambda_1 = 1, \quad 0 < \lambda_2 = \cdots = \lambda_k < 1, \quad and \quad \lambda_l = 0 \ for \ l > k \geq 2.$$

Then the linear tensor product problem $S = \{S_d\}$ is $(T, \Omega^{\mathrm{unr}})$-tractable in the class Λ^{all} iff

$$B_k := \liminf_{d \to \infty} \ \inf_{1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k}d} \frac{\ln T(\varepsilon^{-1}, d)}{m_k(\varepsilon, d)} \in (0, \infty], \tag{8.28}$$

where $m_k(\varepsilon, d) := \alpha(\varepsilon) \ln \left(\frac{d}{\alpha(\varepsilon)}(k - 1)\right) + (d - \alpha(\varepsilon)) \ln \left(\frac{d}{d-\alpha(\varepsilon)}\right)$.
   If $B_k > 0$, then the exponent t^{tra} of tractability is given by

$$t^{\mathrm{tra}} = B_k^{-1}. \tag{8.29}$$

Proof. For the eigenvalues specified in Theorem 8.12, it is easy to check that (8.10) yields

$$n(\varepsilon, d) = \sum_{\nu=0}^{\min\{\alpha(\varepsilon), d\}} \binom{d}{\nu} (k-1)^{\nu}. \tag{8.30}$$

Let us first assume that S is $(T, \Omega^{\mathrm{unr}})$-tractable, i.e., that there exist $C, t > 0$ such that $n(\varepsilon, d) \leq C \, T(\varepsilon^{-1}, d)^t$. Let $1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k} d$. From (8.27) and (8.30) we get the estimate

$$\binom{d}{\alpha(\varepsilon)} (k-1)^{\alpha(\varepsilon)} \leq n(\varepsilon, d) \leq (\alpha(\varepsilon) + 1) \binom{d}{\alpha(\varepsilon)} (k-1)^{\alpha(\varepsilon)}. \tag{8.31}$$

Using Stirling's formula for factorials $m! = m^{m+1/2} e^{-m} \sqrt{2\pi}(1 + o(1))$, we obtain

$$
\begin{aligned}
\ln & \left(\binom{d}{\alpha(\varepsilon)} (k-1)^{\alpha(\varepsilon)} \right) \\
&= \ln(d!) - \ln(\alpha(\varepsilon)!) - \ln((d - \alpha(\varepsilon))!) + \alpha(\varepsilon) \ln(k-1) \\
&= \left(d + \frac{1}{2} \right) \ln(d) - \left(\alpha(\varepsilon) + \frac{1}{2} \right) \ln(\alpha(\varepsilon)) - \left(d - \alpha(\varepsilon) + \frac{1}{2} \right) \ln(d - \alpha(\varepsilon)) \\
&\quad - \ln(\sqrt{2\pi}) + \ln(\mathcal{O}(1)) + \alpha(\varepsilon) \ln(k-1) \\
&= m_k(\varepsilon, d) + \frac{1}{2} \ln \left(\frac{d}{\alpha(\varepsilon)(d - \alpha(\varepsilon))} \right) + \mathcal{O}(1).
\end{aligned}
$$

Thus

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_k(\varepsilon, d)} \geq \frac{1}{t} + \frac{\ln \left(\frac{d}{\alpha(\varepsilon)(d - \alpha(\varepsilon))} \right)}{2t m_k(\varepsilon, d)} - \frac{\ln(C)}{t m_k(\varepsilon, d)} + \frac{\mathcal{O}(1)}{t m_k(\varepsilon, d)}. \tag{8.32}$$

Let $\{(\varepsilon_\nu^{-1}, d_\nu)\}$ be a sequence in Ω^{unr} such that $1 \leq \alpha(\varepsilon_\nu) \leq (k-1)d/k$, and $\lim_{\nu \to \infty}(\alpha(\varepsilon_\nu)/d_\nu)$ exists (and obviously is at most $(k-1)/k$) with $\lim_{\nu \to \infty} d_\nu = \infty$.

If $\lim_{\nu \to \infty}(\alpha(\varepsilon_\nu)/d_\nu) > 0$ then $m_k(\varepsilon_\nu, d_\nu) = \Theta(d_\nu)$ and the right hand side of (8.32) tends to $1/t$ for $\nu \to \infty$.

If $\lim_{\nu \to \infty}(\alpha(\varepsilon_\nu)/d_\nu) = 0$ then

$$m_k(\varepsilon_\nu, d_\nu) = \Theta\left(\alpha(\varepsilon_\nu) \ln \left(\frac{d_\nu}{\alpha(\varepsilon_\nu)}(k-1) \right) \right),$$

since

$$(d_\nu - \alpha(\varepsilon_\nu)) \ln \left(\frac{d_\nu}{d_\nu - \alpha(\varepsilon_\nu)} \right) = \Theta(\alpha(\varepsilon_\nu)).$$

Furthermore,

$$\left| \ln \left(\frac{d_\nu}{\alpha(\varepsilon_\nu)(d_\nu - \alpha(\varepsilon_\nu))} \right) \right| = \Theta(\ln(\alpha(\varepsilon_\nu))).$$

Hence, again, the right hand side of (8.32) tends to $1/t$. Since an arbitrary sequence $\{(\varepsilon_\nu^{-1}, d_\nu)\}$ with $\lim_{\nu\to\infty} d_\nu = \infty$ and $1 \le \alpha(\varepsilon_\nu) \le \frac{k-1}{k}d$ has a sub-sequence $\{(\varepsilon_\mu^{-1}, d_\mu)\}$ for which $\{\alpha(\varepsilon_\mu)/d_\mu\}$ converges, we conclude that

$$B_k \ge \frac{1}{t} > 0 \quad \text{and} \quad t^{\text{tra}} \ge B_k^{-1}. \tag{8.33}$$

Assume now $B_k > 0$. We want to show that for all $t > B_k^{-1}$ there exists a $C = C(t) > 0$ such that $n(\varepsilon, d) \le C T(\varepsilon^{-1}, d)^t$ for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$. From (8.30) we see that this inequality is trivial if $\alpha(\varepsilon) = 0$, and, since $T(\varepsilon^{-1}, d)$ is non-decreasing in ε^{-1}, that the case $\alpha(\varepsilon) > d$ is settled if we have the inequality for $\alpha(\varepsilon) = d$. Thus it remains to consider the following two cases:

Case 1: $1 \le \alpha(\varepsilon) \le \frac{k-1}{k}d$. We now show that for all $t > B_k^{-1}$ there exists a $C = C(t) > 0$ such that for all $d \in \mathbb{N}$,

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_k(\varepsilon, d)} \ge \frac{1}{t} + \frac{\ln(1 + \alpha(\varepsilon))}{t m_k(\varepsilon, d)} + \frac{\ln\left(\frac{d}{\alpha(\varepsilon)(d - \alpha(\varepsilon))}\right)}{2 t m_k(\varepsilon, d)}$$
$$- \frac{\ln(C)}{t m_k(\varepsilon, d)} + \frac{\mathcal{O}(1)}{t m_k(\varepsilon, d)}. \tag{8.34}$$

Due to (8.31) and the formula for $\ln\left(\binom{d}{\alpha(\varepsilon)}(k-1)^{\alpha(\varepsilon)}\right)$, we conclude that $n(\varepsilon, d) \le C T(\varepsilon^{-1}, d)^t$ for all $d \in \mathbb{N}$, $1 \le \alpha(\varepsilon) \le \frac{k-1}{k}d$.

To prove (8.34), observe that $\ln\left(\frac{d}{\alpha(\varepsilon)(d - \alpha(\varepsilon))}\right) \le \ln(k)$. Obviously,

$$m_k(\varepsilon, d) \ge \alpha(\varepsilon) \ln\left(\frac{d}{\alpha(\varepsilon)}(k-1)\right).$$

For a given $l \in \mathbb{N}$ let $x_l \in \mathbb{R}$ be so large that for all $x \ge x_l$ we have

$$\frac{\ln(1 + x)}{x \ln(k)} \le \frac{1}{l}.$$

Let now $d_l \ge (1 + x_l)^{l+1}$. Then we get for $d \ge d_l$ and $1 \le \alpha(\varepsilon) \le \frac{k-1}{k}d$,

$$\frac{\ln(1 + \alpha(\varepsilon))}{\alpha(\varepsilon) \ln(\frac{d}{\alpha(\varepsilon)}(k-1))} \le \frac{1}{l}.$$

Furthermore, let $\{\tilde{B}_l\}$ be a sequence in $(0, B_k)$ that converges to B_k. For each l we find a d_l' such that for all $d \ge d_l'$ and all $1 \le \alpha(\varepsilon) \le \frac{k-1}{k}d$

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_k(\varepsilon, d)} \ge \tilde{B}_l.$$

Choose $t_l := (1 + \frac{1}{l})\tilde{B}_l^{-1}$. For all $d \ge \max\{d_l, d_l'\}$ and all $1 \le \alpha(\varepsilon) \le \frac{k-1}{k}d$ we have

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_k(\varepsilon, d)} \ge \frac{1}{t_l}\left(1 + \frac{\ln(1 + \alpha(\varepsilon))}{m_k(\varepsilon, d)}\right).$$

It is now easy to see that (8.34) holds for $t = t_l$ and all $d \in \mathbb{N}$ and all $1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k}d$ if we just choose $C = C(t_l)$ suitably large. Observe that t_l converges to B_k^{-1} as l tends to infinity.

Case 2: $\frac{k-1}{k}d < \alpha(\varepsilon) \leq d$. Let $\delta \in (0, B_k^{-1})$ and $t \geq (B_k - \delta)^{-1}$. There exists a d_δ such that for all $d \geq d_\delta$ and all ε_* with $1 \leq \alpha(\varepsilon_*) \leq \frac{k-1}{k}d$, we have

$$\frac{\ln T(\varepsilon_*^{-1}, d)}{m_k(\varepsilon_*, d)} \geq B_k - \delta.$$

For $d \geq d_\delta$ and $\alpha(\varepsilon) \geq \frac{k-1}{k}d$, choose $\varepsilon_* \in [\varepsilon, 1)$ such that $\alpha(\varepsilon_*) = \lfloor \frac{k-1}{k}d \rfloor = d - \lceil \frac{d}{k} \rceil$. Then

$$m_k(\varepsilon_*, d) \geq d \ln(k) - \left\lceil \frac{d}{k} \right\rceil \ln\left(1 + \frac{k}{d}\right),$$

and

$$t \ln T(\varepsilon^{-1}, d) \geq (B_k - \delta)^{-1} \ln T(\varepsilon_*^{-1}, d) \geq m_k(\varepsilon_*, d).$$

We find a number C not depending on d such that $\ln(C) \geq \lceil \frac{d}{k} \rceil \ln(1 + \frac{k}{d})$. From (8.30) we know that $n(\varepsilon, d) \leq k^d$, and this yields

$$t \ln T(\varepsilon^{-1}, d) \geq m_k(\varepsilon_*, d) \geq \ln n(\varepsilon, d) - \ln(C),$$

implying $C T(\varepsilon^{-1}, d)^t \geq n(\varepsilon, d)$. Choosing $C = C(t)$ sufficiently large the last inequality extends to all d and all ε with $\frac{k-1}{k}d \leq \alpha(\varepsilon) \leq d$.

The statement of the theorem follows from Cases 1 and 2. □

We illustrate Theorem 8.12 by two tractability functions.

- Let $T(x, y) = xy$ which corresponds to polynomial tractability. Then it is easy to check that $B_k = 0$ for all $k \geq 2$. This means that we do not have polynomial tractability for any linear tensor product problem with at least two positive eigenvalues for $d = 1$. This result has been known before.

- Let $T(x, y) = x^{1+\ln y}$. Then it can be checked that

$$B_k = \tfrac{1}{2}\ln(\lambda_2^{-1}) \text{ for all } k \geq 2, \quad \text{and} \quad t^{\text{tra}} = \frac{2}{\ln(\lambda_2^{-1})}.$$

Hence, the exponent of tractability only depends on the second largest eigenvalue and is independent of its multiplicity. Note that the exponent of tractability goes to infinity as λ_2 approaches 1.

We now consider the general case of finitely many positive eigenvalues.

Corollary 8.13. *Let T be a tractability function. Let $k \geq 2$ and $\lambda_1 = 1$, $\lambda_2 \in (0, 1)$, and $\lambda_l = 0$ for $l > k$. Then the linear tensor product problem $S = \{S_d\}$ is (T, Ω^{unr})-tractable in the class Λ^{all} iff for some (and thus for all) $j \in \{2, 3, \ldots, k\}$*

$$B_j = \liminf_{d \to \infty} \inf_{1 \leq \alpha(\varepsilon) \leq \frac{j-1}{j}d} \frac{\ln T(\varepsilon^{-1}, d)}{m_j(\varepsilon, d)} \in (0, \infty], \tag{8.35}$$

where $m_j(\varepsilon, d) = \alpha(\varepsilon) \ln \left(\frac{d}{\alpha(\varepsilon)} (j-1) \right) + (d - \alpha(\varepsilon)) \ln \left(\frac{d}{d - \alpha(\varepsilon)} \right)$. In this case the exponent t^{tra} of tractability satisfies

$$B_2^{-1} \leq t^{\text{tra}} \leq B_k^{-1}. \tag{8.36}$$

Proof. Obviously we have $B_2 \geq B_3 \geq \cdots \geq B_k$. We need to show that $B_2 > 0$ implies that $B_k > 0$. We first show that $B_2 > 0$ implies

$$\liminf_{d \to \infty} \inf_{1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k} d} \frac{\ln T(\varepsilon^{-1}, d)}{m_2(\varepsilon, d)} > 0. \tag{8.37}$$

Let ε_* satisfy $\alpha(\varepsilon_*) = \lfloor \frac{d}{2} \rfloor$. We have $m_2(\varepsilon_*, d) \geq d \ln(2) - \lceil \frac{d}{2} \rceil \ln(1 + \frac{2}{d})$. Thus for $\frac{d}{2} < \alpha(\varepsilon) \leq \frac{k-1}{k} d$ we get

$$m_2(\varepsilon, d) \leq d \ln(2) + \frac{d}{2} \ln(k) \leq C m_2(\varepsilon_*, d)$$

for d and C sufficiently large. Since T is non-decreasing with respect to the first variable, it is easy to see that $B_2 > 0$ implies (8.37).

Now we prove that

$$\frac{m_k(\varepsilon, d)}{m_2(\varepsilon, d)} = 1 + \frac{\alpha(\varepsilon) \ln(k-1)}{\alpha(\varepsilon) \ln \left(\frac{d}{\alpha(\varepsilon)} \right) + (d - \alpha(\varepsilon)) \ln \left(\frac{d}{d - \alpha(\varepsilon)} \right)} \tag{8.38}$$

is bounded uniformly for all $d \in \mathbb{N}$ and all ε with $1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k} d$. This follows easily from

$$m_2(\varepsilon, d) \geq \alpha(\varepsilon) \ln \left(\frac{d}{\alpha(\varepsilon)} \right) \geq \alpha(\varepsilon) \ln \left(\frac{k}{k-1} \right).$$

Thus $B_2 > 0$ implies

$$B_k \geq \left(\liminf_{d \to \infty} \inf_{1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k} d} \frac{\ln T(\varepsilon^{-1}, d)}{m_2(\varepsilon, d)} \right) \left(\inf_{d \in \mathbb{N}; 1 \leq \alpha(\varepsilon) \leq \frac{k-1}{k} d} \frac{m_2(\varepsilon, d)}{m_k(\varepsilon, d)} \right) > 0.$$

Since the linear tensor product problem S' having only the two non-zero eigenvalues $\lambda_1' = \lambda_1$ and $\lambda_2' = \lambda_2$ is at most as difficult as S and the problem S'' having eigenvalues $\lambda_1'' = \lambda_1, \lambda_2'' = \cdots = \lambda_k'' = \lambda_2$ and $\lambda_l'' = 0$ for $l > k$ is at least as difficult as S, the corollary follows from Theorem 8.12. $\square$

Remark 8.14. Theorem 8.12 shows that in the case $\lambda_3 = \cdots = \lambda_k = 0$ we have $t^{\text{tra}} = B_2^{-1}$, while in the case $\lambda_2 = \lambda_3 = \cdots = \lambda_k$ we have $t^{\text{tra}} = B_k^{-1}$.

If we consider a fixed tractability function T, a sequence $\{S^{(n)}\}$ of tensor product problems whose eigenvalues $\{\lambda_i^{(n)}\}$ satisfy $\lambda_1^{(n)} = \lambda_1 = 1, \lambda_2^{(n)} = \lambda_2 \in (0, 1)$, $\lambda_3^{(n)}, \ldots, \lambda_k^{(n)} > 0$, and $\lim_{n \to \infty} \lambda_3^{(n)} = 0$, then we do not necessarily have that the

corresponding exponents of tractability t_n^{tra} converge to B_2^{-1} as the following counterexample shows. Let

$$T(\varepsilon^{-1}, d) = \sum_{\nu=0}^{\min\{\alpha(\varepsilon), d\}} \binom{d}{\nu}.$$

Then it is not hard to see that T is indeed a tractability function and that $B_2 = 1$ (we showed that implicitly in the proof of Theorem 8.12). According to Corollary 8.13 each problem $S^{(n)}$ is (T, Ω^{unr})-tractable. For $d \in \mathbb{N}$ we obviously have $\sup_{\varepsilon \in (0,1)} T(\varepsilon^{-1}, d) = 2^d$. If we choose $\varepsilon = \varepsilon_d^{(n)} = \frac{1}{2}(\lambda_k^{(n)})^{d/2}$, we get $n(\varepsilon, S_d^{(n)}) = k^d$. This implies $t_n^{\text{tra}} \geq \ln(k)/\ln(2)$ for all n. This shows that the sequence $\{t_n^{\text{tra}}\}$ does not converge to $B_2^{-1} = 1$.

Example 8.15. Let the conditions of Corollary 8.13 hold. We consider the special tractability function $T(x, y) = \exp(f_1(x) f_2(y))$, where $f_i : [1, \infty) \to (0, \infty)$, $i = 1, 2$, are non-decreasing functions. Let

$$a_i := \liminf_{x \to \infty} \frac{f_i(x)}{\ln x} \quad \text{for } i = 1, 2.$$

Let us assume that S is (T, Ω^{unr})-tractable. According to Corollary 8.13 we have $B_2 > 0$, and from $m_2(\varepsilon, d) \geq \alpha(\varepsilon)(\ln(d) - \ln(\alpha(\varepsilon)))$ for all ε satisfying $1 \leq \alpha(\varepsilon) \leq d/2$ we get

$$0 < B_2 \leq \liminf_{d \to \infty} \frac{f_1(\varepsilon^{-1}) f_2(d)}{\alpha(\varepsilon)(\ln(d) - \ln(\alpha(\varepsilon)))}$$

$$\leq \frac{f_1(\varepsilon^{-1})}{\alpha(\varepsilon)} \left(\liminf_{d \to \infty} \frac{f_2(d)}{\ln(d)} \right) \left(\limsup_{d \to \infty} \frac{\ln(d)}{\ln(d) - \ln(\alpha(\varepsilon))} \right) = \frac{f_1(\varepsilon^{-1})}{\alpha(\varepsilon)} a_2.$$

Thus $a_2 > 0$, and

$$0 < \frac{B_2}{a_2} \leq \liminf_{\varepsilon \to 0} \frac{f_1(\varepsilon^{-1})}{\alpha(\varepsilon)} = \liminf_{\varepsilon \to 0} \left(\frac{\ln(\varepsilon^{-1})}{\alpha(\varepsilon)} \frac{f_1(\varepsilon^{-1})}{\ln(\varepsilon^{-1})} \right) = \frac{\ln(\lambda_2^{-1})}{2} a_1.$$

Hence, $a_1 > 0$ and $a_2 > 0$ are necessary conditions for the problem S to be (T, Ω^{unr})-tractable, and the exponent of tractability is bounded from below by

$$t^{\text{tra}} \geq B_2^{-1} \geq \frac{2}{a_1 a_2 \ln(\lambda_2^{-1})}.$$

In Corollary 8.24 we will show in particular that the conditions $a_1 > 0$, $a_2 > 0$ are also sufficient for (T, Ω^{unr})-tractability.

Remark 8.16. Under the conditions of Corollary 8.13 we can state a slightly simpler criterion to characterize (T, Ω^{unr})-tractability. The linear tensor product problem $S = \{S_d\}$ is (T, Ω^{unr})-tractable in the class Λ^{all} iff

$$B := \liminf_{d \to \infty} \inf_{1 \leq \alpha(\varepsilon) \leq d/2} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln(d/\alpha(\varepsilon))} \in (0, \infty]. \tag{8.39}$$

The necessity and sufficiency of $B > 0$ follows from (8.35) and the (easy to check) inequalities

$$\frac{1}{2}m_2(\varepsilon, d) \leq \alpha(\varepsilon) \ln\left(\frac{d}{\alpha(\varepsilon)}\right) \leq m_2(\varepsilon, d)$$

for all ε satisfying $1 \leq \alpha(\varepsilon) \leq \frac{d}{2}$ and large d. A drawback of (8.39) is that the quantity B is not related to the exact exponent of tractability as B_k in Theorem 8.12.

Example 8.17. The tractability criteria (8.35) and (8.39) depend on the second largest eigenvalue λ_2 via $\alpha(\varepsilon)$. In fact, for a given tractability function T, a linear tensor product problem $S = \{S_d\}$ with only two positive eigenvalues for $S_1^* S_1$ may be $(T, \Omega^{\mathrm{unr}})$-tractable, but if we increase the value of λ_2 this may not necessarily be the case any more. Choose, e.g.,

$$T(x, y) := \begin{cases} 1 & \text{if } x \in [1, \lambda_2^{-1/2}], \\ e^{\ln(x)(1+\ln(y))} & \text{otherwise.} \end{cases}$$

From criterion (8.39) it easily follows that S is $(T, \Omega^{\mathrm{unr}})$-tractable. But if we consider the problem $\widetilde{S}$ where we only increase the second eigenvalue to $\tilde{\lambda}_2 > \lambda_2$, we see that for $\tilde{\lambda}_2^{-1/2} < \varepsilon^{-1} \leq \lambda_2^{-1/2}$ we have

$$n(\varepsilon, \widetilde{S}_d) \geq \sum_{\nu=0}^{\min\{\tilde{\alpha}(\varepsilon), d\}} \binom{d}{\nu} \geq \binom{d}{1} = d, \quad \text{where } \tilde{\alpha}(\varepsilon) := \left\lceil \frac{2\ln(\varepsilon^{-1})}{\ln(\tilde{\lambda}_2^{-1})} \right\rceil - 1 \geq 1.$$

Thus the problem $\widetilde{S}_d$ is obviously not $(T, \Omega^{\mathrm{unr}})$-tractable since $C T(\varepsilon^{-1}, d)^t = C$ cannot be larger than d for $d > C$.

The counterexample above motivates us to state a sufficient condition on T ensuring $(T, \Omega^{\mathrm{unr}})$-tractability of all linear tensor product problems S with finitely many eigenvalues regardless of the specific value of λ_2.

Corollary 8.18. *Let T be a tractability function. If*

$$\widetilde{B} := \liminf_{d \to \infty} \inf_{1 < \varepsilon^{-1} \leq e^d} \frac{\ln T(\varepsilon^{-1}, d)}{\ln(\varepsilon^{-1})(1 + \ln(d/\ln(\varepsilon^{-1})))} \in (0, \infty] \qquad (8.40)$$

then arbitrary linear tensor product problem S with finitely many eigenvalues is $(T, \Omega^{\mathrm{unr}})$-tractable. However, the exponent of tractability goes to infinity as λ_2 approaches 1.

Proof. The proof of the corollary is easy. For values of $\varepsilon \in [e^{-d}, 1)$ satisfying $\alpha(\varepsilon) \in [1, d/2]$ one can simply show that $\alpha(\varepsilon) \ln(d/\alpha(\varepsilon)) \leq C \ln(\varepsilon^{-1})(1 + \ln(d/\ln(\varepsilon^{-1})))$, where the number C depends only on λ_2. If we substitute the upper bound on $\alpha(\varepsilon)$ in the definition of B in (8.39) by the minimum of $d/2$ and $\lceil 2d/\ln(\lambda_2^{-1}) \rceil - 1$, we therefore see that this modified quantity is strictly positive. From that we can deduce similarly as in Case 2 in the proof of Theorem 8.12 that $B > 0$, and due to Remark 8.16, the problem S is $(T, \Omega^{\mathrm{unr}})$-tractable. Obviously, $n(\varepsilon, d) \geq 2^d$ for $\varepsilon^2 < \lambda_2^d$. Hence, the exponent of tractability must go to infinity as λ_2 goes to one. $\qquad\square$

Remark 8.19. Condition (8.40) in the corollary above is sufficient for $(T, \Omega^{\mathrm{unr}})$-tractability for all linear tensor product problems S with finitely many eigenvalues, but not necessary as the example $T(\varepsilon^{-1}, d) = \exp(\ln(\varepsilon^{-1})(1 + \ln(d)))$ shows, see Corollary 8.24.

8.4.2 Exponential Decay of Eigenvalues

We begin to study linear tensor problems with infinitely many positive eigenvalues. As we shall see, tractability results depend on the behavior of the eigenvalues for $d = 1$. In this section we assume that they are exponentially decaying whereas in the next section that they are polynomially decaying.

Theorem 8.20. *Let T be a tractability function. Let S be a linear tensor product problem with exponentially decaying eigenvalues λ_j,*

$$\exp(-\beta_1(j-1)) \le \lambda_j \le \exp(-\beta_2(j-1)) \quad \text{for all } j \in \mathbb{N},$$

for some positive numbers β_1, β_2. For $i = 1, 2$, define

$$B_{\mathrm{e}}^{(i)} := \liminf_{\substack{\varepsilon^{-1} + d \to \infty \\ \varepsilon < \sigma_i}} \frac{\ln T(\varepsilon^{-1}, d)}{m_{\mathrm{e}}^{(i)}(\varepsilon, d)},$$

where $\sigma_1 = e^{-\beta_1/2}$, $\sigma_2 = \sqrt{\lambda_2}$, and

$$m_{\mathrm{e}}^{(i)}(\varepsilon, d) := \lceil z_i \rceil \ln\left(1 + \frac{d}{\lceil z_i \rceil}\right) + d \ln\left(1 + \frac{\lceil z_i \rceil}{d}\right),$$

with

$$z_i = z_i(\varepsilon) := \frac{2}{\beta_i} \ln(\varepsilon^{-1}) - 1.$$

Then

$$S \text{ is } (T, \Omega^{\mathrm{unr}})\text{-tractable} \quad \text{iff} \quad B_{\mathrm{e}}^{(2)} \in (0, \infty].$$

Furthermore, $B_{\mathrm{e}}^{(2)} > 0$ is equivalent to $B_{\mathrm{e}}^{(1)} \in (0, \infty]$ and $B_2 \in (0, \infty]$ with B_2 given by (8.28) for $k = 2$.

If S is $(T, \Omega^{\mathrm{unr}})$-tractable then the exponent t^{tra} of tractability satisfies

$$\left(\min\{B_2, B_{\mathrm{e}}^{(1)}\}\right)^{-1} \le t^{\mathrm{tra}} \le \left(B_{\mathrm{e}}^{(2)}\right)^{-1}.$$

If $\beta_1 = \beta_2$ then

$$t^{\mathrm{tra}} = \left(B_{\mathrm{e}}^{(2)}\right)^{-1}.$$

Before we prove Theorem 8.20, we state an auxiliary lemma.

Lemma 8.21. *For $d \in \mathbb{N}$ and $x > -1$ let*

$$\mu_e(x, d) := \left|\left\{(i_1, \ldots, i_d) \in \mathbb{N}^d \mid \sum_{j=1}^{d} i_j < x + d + 1\right\}\right|.$$

Then

$$\mu_e(x, d) = \binom{\lceil x \rceil + d}{d}.$$

Proof. For $d = 1$ we have

$$\mu_e(x, 1) = |\{i \in \mathbb{N} \mid i < x + 2\}| = \lceil x \rceil + 1.$$

Assume by induction that

$$\mu_e(y, d) = \binom{\lceil y \rceil + d}{d}$$

for some $d \in \mathbb{N}$ and all $y > -1$. If $x > -1$ then

$$\mu_e(x, d + 1) = \sum_{k=1}^{\lceil x \rceil + 1} \mu_e(x + 1 - k, d) = \sum_{k=1}^{\lceil x \rceil + 1} \binom{\lceil x \rceil + 1 - k + d}{d}$$

$$= \sum_{\nu=0}^{\lceil x \rceil} \binom{\nu + d}{d} = \binom{\lceil x \rceil + d + 1}{d + 1}. \qquad \square$$

Proof of Theorem 8.20. Let $\mu_e(x, d)$ be defined as in Lemma 8.21. Then

$$\mu_e(z_1, d) = \left|\left\{(i_1, \ldots, i_d) \in \mathbb{N}^d \mid \prod_{j=1}^{d} \exp(-\beta_1(i_j - 1)) > \varepsilon^2\right\}\right| \leq n(\varepsilon, d).$$

Similarly, we get $n(\varepsilon, d) \leq \mu_e(z_2, d)$.

Let us first assume that S is $(T, \Omega^{\mathrm{unr}})$-tractable, i.e., that there exist positive t and C such that

$$n(\varepsilon, d) \leq C T(\varepsilon^{-1}, d)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega^{\mathrm{unr}}.$$

Let us assume that $\varepsilon < e^{-\beta_1/2}$, which implies that $\lceil z_1 \rceil \geq 1$. From this inequality we get due to Lemma 8.21

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_e^{(1)}(\varepsilon, d)} \geq \frac{\ln(C^{-1}) + \ln \binom{\lceil z_1 \rceil + d}{d}}{t m_e^{(1)}(\varepsilon, d)}.$$

Similarly as in the proof of Theorem 8.12 we use Stirling's formula for factorials, and conclude

$$\ln \binom{\lceil z_1 \rceil + d}{d} = m_e^{(1)}(\varepsilon, d) + \frac{1}{2} \ln \left(\frac{\lceil z_1 \rceil + d}{\lceil z_1 \rceil d}\right) + \mathcal{O}(1). \qquad (8.41)$$

We have

$$-\min\{\ln(d), \ln \lceil z_1 \rceil\} \leq \ln \left(\frac{\lceil z_1 \rceil + d}{\lceil z_1 \rceil d}\right) = \ln \left(\frac{1}{\lceil z_1 \rceil} + \frac{1}{d}\right) \leq \ln(2).$$

So it is easy to check that we get $B_e^{(1)} \geq 1/t$, implying $B_e^{(1)} > 0$ and $t^{\text{tra}} \geq (B_e^{(1)})^{-1}$. Furthermore, we get from Corollary 8.13 that $B_2 > 0$ and $t^{\text{tra}} \geq B_2^{-1}$.

Let us now show that $B_2 > 0$ and $B_e^{(1)} > 0$ imply $B_e^{(2)} > 0$. As a careful analysis reveals, we get

$$K := \liminf_{\varepsilon^{-1}+d \to \infty} \ \inf_{\varepsilon < e^{-\beta_1/2}} \frac{m_e^{(1)}(\varepsilon, d)}{m_e^{(2)}(\varepsilon, d)} > 0,$$

which gives us

$$\liminf_{\substack{\varepsilon^{-1}+d \to \infty \\ \varepsilon < e^{-\beta_1/2}}} \frac{\ln T(\varepsilon^{-1}, d)}{m_e^{(2)}(\varepsilon, d)} \geq B_e^{(1)} K > 0.$$

In the case $e^{-\beta_1/2} \leq \varepsilon < \sqrt{\lambda_2}$ both functions $\alpha(\varepsilon)$ and $z_2(\varepsilon)$ are bounded. Thus we have $m_2(\varepsilon, d) = \Theta(\ln(d)) = m_e^{(2)}(\varepsilon, d)$, where m_2 is given in Theorem 8.12. Hence

$$L := \liminf_{d \to \infty} \ \inf_{e^{-\beta_1/2} \leq \varepsilon < \sqrt{\lambda_2}} \frac{m_2(\varepsilon, d)}{m_e^{(2)}(\varepsilon, d)} > 0,$$

which yields

$$\liminf_{\substack{\varepsilon^{-1}+d \to \infty \\ e^{-\beta_1/2} \leq \varepsilon \leq \sqrt{\lambda_2}}} \frac{\ln T(\varepsilon^{-1}, d)}{m_e^{(2)}(\varepsilon, d)} \geq B_2 L > 0.$$

This means that $B_e^{(2)}$ is positive, as claimed.

Now let us assume that $B_e^{(2)} > 0$ and let $t_\delta := ((1-\delta)B_e^{(2)})^{-1}$ for a given $\delta \in (0, 1)$. Then there exists an $R(\delta)$ such that for any pair (ε, d) with $\varepsilon^{-1} + d > R(\delta)$ (and $\varepsilon < \sqrt{\lambda_2}$, but for convenience we will not mention this restriction in the rest of the proof) we get

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_e^{(2)}(\varepsilon, d)} > \left(1 - \frac{\delta}{2}\right) B_e^{(2)}.$$

We want to show that there exists a number C_δ such that

$$n(\varepsilon, d) \leq C_\delta T(\varepsilon^{-1}, d)^{t_\delta} \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega^{\text{unr}}.$$

Since $n(\varepsilon, d) \leq \mu_e(z_2, d)$, it is sufficient to verify the inequality

$$\frac{\ln T(\varepsilon^{-1}, d)}{m_e^{(2)}(\varepsilon, d)} \geq \frac{\ln(C_\delta^{-1}) + \ln\binom{\lceil z_2 \rceil + d}{d}}{t_\delta m_e^{(2)}(\varepsilon, d)}. \tag{8.42}$$

The left hand side is at least $(1 - \delta/2)B_e^{(2)}$. Using Stirling's formula (8.41) for z_2 instead of z_1, we see that the right hand side can be written as

$$(1-\delta)B_e^{(2)} + \frac{\ln(\frac{\lceil z_2 \rceil + d}{\lceil z_2 \rceil d})}{2t_\delta m_e^{(2)}(\varepsilon, d)} + \frac{\ln(C_\delta^{-1})}{t_\delta m_e^{(2)}(\varepsilon, d)} + \frac{\mathcal{O}(1)}{t_\delta m_e^{(2)}(\varepsilon, d)}.$$

The limes superior of all the summands, except of $(1-\delta)B_e^{(2)}$, goes to 0 as $\varepsilon^{-1} + d$ tends to infinity. Hence, there exists an $\widetilde{R}(\delta)$ such that for all pairs (ε, d) with $\varepsilon^{-1} + d > \widetilde{R}(\delta)$

inequality (8.42) holds. Choosing C_δ sufficiently large, we see therefore that (8.42) holds for all $(\varepsilon^{-1}, d) \in \Omega^{\mathrm{unr}}$. This shows that we have $(T, \Omega^{\mathrm{unr}})$-tractability and, since $\delta \in (0, 1)$ was arbitrary, the exponent of tractability t^{tra} satisfies $t^{\mathrm{tra}} \le (B_{\mathrm{e}}^{(2)})^{-1}$. As we already have seen, tractability implies $B_{\mathrm{e}}^{(1)} > 0$ and $B_2 > 0$.

Finally, if $\beta_1 = \beta_2$ then $B_{\mathrm{e}}^{(1)} = B_{\mathrm{e}}^{(2)}$, and therefore $(\min\{B_2, B_{\mathrm{e}}^{(1)}\})^{-1} \le t^{\mathrm{tra}} \le (B_{\mathrm{e}}^{(2)})^{-1}$ implies that $B_2 \ge B_{\mathrm{e}}^{(1)}$ and $t^{\mathrm{tra}} = (B_{\mathrm{e}}^{(2)})^{-1}$.   □

We illustrate Theorem 8.20 by taking again the tractability function $T(x, y) = x^{1+\ln y}$. For $\beta_1 = \beta_2 = \beta > 0$, we have $\lambda_2 = \exp(-\beta)$. It can be checked that

$$B_{\mathrm{e}}^{(2)} = \frac{\beta}{2} = \frac{\ln(\lambda_2^{-1})}{2}.$$

Thus the exponent of tractability is

$$t^{\mathrm{tra}} = (B_{\mathrm{e}}^{(2)})^{-1} = \frac{2}{\beta} = \frac{2}{\ln(\lambda_2^{-1})}.$$

We can simplify the necessary and sufficient conditions in Theorem 8.20 for $(T, \Omega^{\mathrm{unr}})$-tractability at the expense of getting good estimates on the exponent of tractability.

Corollary 8.22. *Let T be a tractability function. Let S be a linear tensor product problem with $0 < \lambda_2 < \lambda_1 = 1$, and with exponentially decaying eigenvalues λ_j,*

$$K_1 \exp(-\beta_1 j) \le \lambda_j \le K_2 \exp(-\beta_2 j) \quad \text{for all } j \in \mathbb{N},$$

for some positive numbers β_1, β_2, K_1 and K_2. Then S is $(T, \Omega^{\mathrm{unr}})$-tractable iff

$$\liminf_{\substack{\varepsilon^{-1}+d \to \infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{\ln T(\varepsilon^{-1}, d)}{\min\{d, \alpha(\varepsilon)\}(1 + |\ln(d/\alpha(\varepsilon))|)} \in (0, \infty]. \qquad (8.43)$$

Proof. Since $\lambda_j \le \min\{\lambda_2, K_2 \exp(-\beta_2 j)\}$ for $j \ge 2$, we can choose positive $\beta_1' \ge \beta_1, \beta_2' \le \beta_2$ such that

$$\exp(-\beta_1'(j - 1)) \le \lambda_j \le \exp(-\beta_2'(j - 1)) \quad \text{for all } j \in \mathbb{N}.$$

Thus we can apply Theorem 8.20. There we showed that $B_{\mathrm{e}}^{(2)} > 0$ is necessary and sufficient for $(T, \Omega^{\mathrm{unr}})$-tractability. For $1 \le \alpha(\varepsilon) \le d/2$ and large d, we have $m_2(\varepsilon, d)/2 \le \alpha(\varepsilon) \ln(d/\alpha(\varepsilon)) \le m_2(\varepsilon, d)$. Furthermore, one can also verify that

$$\liminf_{d \to \infty} \inf_{d/2 \le \alpha(\varepsilon)} \left(\frac{m_{\mathrm{e}}^{(2)}(\varepsilon, d)}{\min\{d, \alpha(\varepsilon)\}(1 + |\ln(d/\alpha(\varepsilon))|)} \right)^q > 0,$$

where $q \in \{-1, +1\}$. Thus (8.43) holds iff $B_{\mathrm{e}}^{(2)} \in (0, \infty]$, which proves the corollary.   □

8.4.3 Polynomial Decay of Eigenvalues

In this section we study tractability for linear tensor product problems with polynomially decaying eigenvalues for $d = 1$. We believe that such behavior of eigenvalues is typical and therefore the results of this section are probably more important than the results of the previous sections.

Theorem 8.23. *Let T be a tractability function. Let S be a linear tensor product problem with $1 = \lambda_1 > \lambda_2 > 0$ and $\lambda_j = \mathcal{O}(j^{-\beta})$ for all $j \in \mathbb{N}$ and some positive β. A sufficient condition for $(T, \Omega^{\mathrm{unr}})$-tractability of S is*

$$F := \liminf_{\substack{\varepsilon^{-1}+d \to \infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{\ln T(\varepsilon^{-1}, d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} \in (0, \infty].$$

If $F \in (0, \infty]$, then the exponent of tractability satisfies

$$B_2^{-1} \leq t^{\mathrm{tra}} \leq \max\left\{\frac{2}{\beta}, \frac{2}{\ln(\lambda_2^{-1})}\right\} F^{-1},$$

with B_2 given in (8.28) for $k = 2$.

Proof. Let C_1 be a positive number satisfying $\lambda_j \leq C_1 j^{-\beta}$ for all j. With $C_2 := C_1^{1/\beta}$ we have

$$n(\varepsilon, 1) = \max\{j \mid \lambda_j > \varepsilon^2\} \leq \max\{j \mid C_1 j^{-\beta} > \varepsilon^2\} \leq C_2 \varepsilon^{-2/\beta} \leq C_2 \varepsilon^{-p}$$

for all $p > 2/\beta$. From the identity

$$n(\varepsilon, d) = \sum_{i=1}^{\infty} n\left(\varepsilon / \sqrt{\lambda_i}, S_{d-1}\right)$$

it now follows by simple induction that

$$n(\varepsilon, d) \leq C_2 \left(\sum_{j=1}^{\infty} \lambda_j^{p/2}\right)^{d-1} \varepsilon^{-p} \quad \text{for all } p > 2/\beta. \tag{8.44}$$

Thus for each $d_0 \in \mathbb{N}$ and all $p > 2/\beta$ there exists a number $C(d_0, p)$ such that

$$n(\varepsilon, d) \leq C(d_0, p)\varepsilon^{-p} \quad \text{for all } d \leq d_0 \text{ and } \varepsilon \in (0, \sqrt{\lambda_2}).$$

Let now $\delta \in (0, 1)$ and $\varepsilon_\delta < \sqrt{\lambda_2}$ such that for all $\varepsilon \in (0, \varepsilon_\delta)$ and all $d \leq d_0$

$$\frac{\ln T(\varepsilon^{-1}, d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} \geq (1 - \delta)F,$$

where F is assumed to be positive. Then for $t = t(\delta, p, d_0) := p(1 - \delta)^{-1} F^{-1}$ and $C = C(d_0, p)$ we have

$$\ln(C T(\varepsilon^{-1}, d)^t) \geq \ln C + p(1 + \ln(d)) \ln(\varepsilon^{-1}) \geq \ln n(\varepsilon, d)$$

for all $d \leq d_0$ and $\varepsilon \in (0, \varepsilon_\delta)$. This implies that for each $t > (2/\beta)F^{-1}$ there exists a sufficiently large number $C = C_t$ such that

$$n(\varepsilon, d) \leq C T(\varepsilon^{-1}, d)^t \quad \text{for all } d \leq d_0 \text{ and } \varepsilon \in (0, \sqrt{\lambda_2}). \tag{8.45}$$

We now consider arbitrarily large d. Let us estimate the sum on the right hand side of inequality (8.44). For this purpose we choose $k \in \mathbb{N}$ such that $\lambda_2 > C_1 k^{-\beta}$. Since $\lambda_2 \leq C_1 2^{-\beta}$, we have obviously $k > 2$. We have

$$\sum_{j=1}^{\infty} \lambda_j^{p/2} \leq 1 + \lambda_2^{p/2} + \cdots + \lambda_k^{p/2} + C_1^{p/2} \sum_{j=k+1}^{\infty} j^{-\frac{p\beta}{2}},$$

and

$$\sum_{j=k+1}^{\infty} j^{-\frac{p\beta}{2}} \leq \int_k^{\infty} x^{-\frac{p\beta}{2}} \, dx = \frac{k^{-\frac{p\beta}{2}+1}}{(p\beta/2) - 1}.$$

Now we choose $p = p(d)$ such that

$$k \left(\lambda_2^{p/2} + \frac{(C_1 k^{-\beta})^{p/2}}{(p\beta/2) - 1} \right) = \frac{1}{d}.$$

From $k\lambda_2^{p/2} \leq 1/d$ we conclude

$$p \geq 2 \left(\frac{\ln d + \ln k}{\ln(\lambda_2^{-1})} \right).$$

From $\lambda_2 > C_1 k^{-\beta}$ we get

$$k \left(1 + \frac{1}{(p\beta/2) - 1} \right) \lambda_2^{p/2} \geq \frac{1}{d},$$

implying

$$p \leq 2 \left(\frac{\ln d + \ln k + \ln \left(1 + \frac{1}{(p\beta/2)-1} \right)}{\ln(\lambda_2^{-1})} \right).$$

Thus we have

$$p = \frac{2\ln(d)}{\ln(\lambda_2^{-1})} (1 + o_d(1)) \quad \text{as } d \to \infty.$$

Let now $\sigma \in (0, 1)$ and $d_\sigma \in \mathbb{N}$ such that $o_d(1) \leq \sigma$ and

$$\frac{\ln T(\varepsilon^{-1}, d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} \geq (1 + \sigma)^{-1} F$$

for all $d \geq d_\sigma$ and all $\varepsilon \in (0, \sqrt{\lambda_2})$. For these d and ε we have

$$n(\varepsilon, d) \leq C_2 \left(1 + \frac{1}{d} \right)^{d-1} \varepsilon^{-p} \leq e C_2 \exp \left(\frac{2\ln(d)}{\ln(\lambda_2^{-1})} (1 + \sigma) \ln(\varepsilon^{-1}) \right)$$

$$\leq C_3 \exp \left(\frac{2}{\ln(\lambda_2^{-1})} F^{-1} (1 + \sigma)^2 \ln T(\varepsilon^{-1}, d) \right),$$

where $C_3 := eC_2$. Hence for $\tau = \tau(\sigma, p, d_\delta) := 2(\ln(\lambda_2^{-1}))^{-1}(1+\sigma)^2 F^{-1}$ we get

$$n(\varepsilon, d) \leq C_3 T(\varepsilon^{-1}, d)^\tau \quad \text{for all } d \geq d_\sigma \text{ and } \varepsilon \in (0, \sqrt{\lambda_2}). \tag{8.46}$$

The estimates (8.45) and (8.46) show that we have (T, Ω^{unr})-tractability. Choosing $d_0 = d_\sigma$ in (8.45) and letting σ tend to 0 yields the claimed upper bound for t^{tra}.

Since our problem is at least as hard as the problem with only two positive eigenvalues $0 < \lambda_2 < \lambda_1 = 1$ for $d = 1$, the lower bound $t^{\text{tra}} \geq B_2^{-1}$ follows from Theorem 8.12 for $k = 2$. $\qquad\square$

The upper bound on the exponent t^{tra} in Theorem 8.23 is, in general, sharp. Indeed, assume that $\lambda_j = \Theta(j^{-\beta})$ and take $T(x, y) = x^{1+\ln y}$. Then $n(\varepsilon, 1) = \Theta(\varepsilon^{-2/\beta})$ which easily implies that $t^{\text{tra}} \geq 2/\beta$. In this case, we have $F = 1$ and $B_2 = \frac{1}{2}\ln(\lambda_2^{-1})$. This shows that the upper bound on t^{tra} in Theorem 8.23 is sharp and

$$t^{\text{tra}} = \max\left\{\frac{2}{\beta}, \frac{2}{\ln(\lambda_2^{-1})}\right\}.$$

Hence, for $\beta \geq \ln(\lambda_2^{-1})$ the exponent of tractability is the same as for the problem with only two positive eigenvalues $0 < \lambda_2 < \lambda_1 = 1$. For this tractability function, the problem S with polynomially decaying eigenvalues is as hard as the problem with only two positive eigenvalues. However, for $\beta < \ln\lambda_2^{-1}$, the exponent of tractability depends on β and the problem S is harder than the problem with only two positive eigenvalues.

Corollary 8.24. *Let $1 = \lambda_1 > \lambda_2 > 0$ and $\lambda_j = \mathcal{O}(j^{-\beta})$ for all $j \in \mathbb{N}$ and some fixed $\beta > 0$. Let $f_i : [1, \infty) \to (0, \infty)$, $i = 1, 2$, be non-decreasing functions such that*

$$\lim_{x+y\to\infty} \frac{f_1(x)f_2(y)}{x+y} = 0.$$

For $T(x, y) = \exp(f_1(x)f_2(y))$, we have (T, Ω^{unr})-tractability iff

$$a_i := \liminf_{x\to\infty} \frac{f_i(x)}{\ln x} \in (0, \infty] \quad \text{for } i = 1, 2.$$

If $a_1, a_2 \in (0, \infty]$, then the exponent of tractability satisfies

$$\frac{2}{a_1 a_2 \ln(\lambda_2^{-1})} \leq t^{\text{tra}} \leq \max\left\{\frac{2}{\beta}, \frac{2}{\ln(\lambda_2^{-1})}\right\} \frac{1}{\min\{a_1 b_2, b_1 a_2\}},$$

where

$$b_1 = \inf_{\varepsilon < \sqrt{\lambda_2}} \frac{f_1(\varepsilon^{-1})}{\ln(\varepsilon^{-1})} \quad \text{and} \quad b_2 = \inf_{d \in \mathbb{N}} \frac{f_2(d)}{1 + \ln(d)}.$$

Proof. We have already seen in Example 8.15 that even for two non-zero eigenvalues λ_1, λ_2 and $0 = \lambda_3 = \lambda_4 = \cdots$ the condition $a_1, a_2 > 0$ is necessary for S to be (T, Ω^{unr})-tractable, and that $t^{\text{tra}} \geq 2/(a_1 a_2 \ln(\lambda_2^{-1}))$.

Let us now assume that $a_1, a_2 \in (0, \infty]$. It is easy to see that

$$
\begin{aligned}
F &= \liminf_{\substack{\varepsilon^{-1}+d\to\infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{\ln T(\varepsilon^{-1}, d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} \\
&= \liminf_{\substack{\varepsilon^{-1}+d\to\infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{f_1(\varepsilon^{-1}) f_2(d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} = \min\{a_1 b_2, b_1 a_2\},
\end{aligned}
$$

and that $a_1, a_2 > 0$ implies $b_1, b_2 > 0$. Thus $F > 0$ and due to Theorem 8.23 we have $(T, \Omega^{\mathrm{unr}})$-tractability and the stated upper bound for t^{tra}. $\square$

We illustrate Corollary 8.24 again for $T(x, y) = x^{1+\ln y} = \exp((\ln x)(1 + \ln y))$. We now have $a_1 = a_2 = b_1 = b_2 = 1$. If we assume that $\lambda_j = \Theta(j^{-\beta})$ then, as we have already checked, $t^{\mathrm{tra}} = \max\{2/\beta, 2/\ln \lambda_2^{-1}\}$. Hence, the upper bound on t^{tra} in Corollary 8.24 is, in general, sharp. This proves that for tractability functions T of the form $T(x, y) = \exp(f_1(x) f_2(x))$, the exponent of tractability may depend on β, i.e., on how fast the eigenvalues decay to 0 for $d = 1$.

We now consider different tractability functions of the form

$$
T(x, y) = f_1(x) f_2(x) = \exp(\ln f_1(x) + \ln f_2(x))
$$

and show that for such functions the exponent of tractability does *not* depend on β. The following theorem generalizes a result from [293] which corresponds to $f_i(x) = \exp(\ln^{1+\alpha_i}(1 + x))$.

Theorem 8.25. *Let S be a linear tensor product problem with $1 = \lambda_1 > \lambda_2 > 0$ and $\lambda_j = \mathcal{O}(j^{-\beta})$ for all $j \in \mathbb{N}$. For $i = 1, 2$ let $f_i \colon [1, \infty) \to [1, \infty)$ be a non-decreasing function with*

$$
a_i := \liminf_{x\to\infty} \frac{\ln \ln f_i(x)}{\ln \ln x} < \infty.
$$

Then the function T defined by $T(x, y) = f_1(x) f_2(y)$ is a tractability function.
S is $(T, \Omega^{\mathrm{unr}})$-tractable iff

$$
a_1 > 1, \quad a_2 > 1, \quad (a_1 - 1)(a_2 - 1) \geq 1, \quad \text{and} \quad B_2 \in (0, \infty],
$$

where B_2 is given by (8.28) for $k = 2$.
If $a_1 > 1$, $a_2 > 1$ and $(a_1 - 1)(a_2 - 1) > 1$ then $B_2 = \infty$ and the exponent of tractability t^{tra} is 0.
If $a_1 > 1$, $a_2 > 1$, $(a_1 - 1)(a_2 - 1) = 1$ and $B_2 > 0$ then the exponent of tractability is

$$
t^{\mathrm{tra}} = B_2^{-1} = \left(\liminf_{\substack{\varepsilon^{-1}+d\to\infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{\ln f_1(\varepsilon^{-1}) + \ln f_2(d)}{\alpha(\varepsilon) \ln(d)} \right)^{-1}.
$$

Proof. Since $a_1, a_2 < \infty$, it is obvious that T is a tractability function. Let first S be (T, Ω)-tractable, i.e., there exist positive numbers C and t such that

$$n(\varepsilon, d) \leq C f_1(\varepsilon^{-1})^t f_2(d)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega^{\text{unr}}.$$

Due to (8.12) we have

$$n(\varepsilon, d) \geq \binom{d}{\alpha(\varepsilon)} \geq \left(\frac{d}{\alpha(\varepsilon)}\right)^{\alpha(\varepsilon)},$$

which implies

$$\alpha(\varepsilon) \ln\left(\frac{d}{\alpha(\varepsilon)}\right) \leq \ln(C) + t \ln f_1(\varepsilon^{-1}) + t \ln f_2(d). \tag{8.47}$$

Keeping ε fixed and letting d grow, we see that for any $\delta > 0$ there exists a $d' = d'(\delta, \varepsilon)$ such that for all $d \geq d'$ we have $\alpha(\varepsilon) \ln(d) \leq (t + \delta) \ln f_2(d)$, and therefore

$$1 + \frac{\ln \alpha(\varepsilon)}{\ln \ln(d)} \leq \frac{\ln \ln f_2(d)}{\ln \ln(d)} + \frac{\ln(t + \delta)}{\ln \ln(d)}.$$

Thus $a_2 \geq 1$. Let now ε vary and take $d = 2\alpha(\varepsilon)$. Since $\ln f_2(d) = o(d) = o(\alpha(\varepsilon))$, we get from (8.47) for arbitrary $\delta > 0$, for $\varepsilon' = \varepsilon'(\delta)$ sufficiently small, and for all $\varepsilon \leq \varepsilon'$ that $\alpha(\varepsilon) \ln(2) \leq (t + \delta) \ln f_1(\varepsilon^{-1})$. Since

$$\ln \alpha(\varepsilon) = \ln(2) + \ln \ln(\varepsilon^{-1}) - \ln \ln(\lambda_2^{-1}) + \mathcal{O}(1) \quad \text{as } \varepsilon \text{ tends to } 0,$$

the estimate $a_1 \geq 1$ easily follows. Let now $\eta > a_1 - 1$. Define

$$d = d(\varepsilon) = \alpha(\varepsilon)^{\alpha(\varepsilon)^\eta}.$$

Then (8.47) yields

$$(\alpha(\varepsilon)^{\eta+1} - \alpha(\varepsilon)) \ln(\alpha(\varepsilon)) \leq \ln(C) + t \ln f_1(\varepsilon^{-1}) + t \ln f_2(d).$$

Due to the choice of η and the fact that $\alpha(\varepsilon) = 2 \ln(\varepsilon^{-1})/\ln(\lambda_2^{-1}) + \mathcal{O}(1)$, the function $\ln f_1(\varepsilon^{-1})$ is of order $o(\alpha(\varepsilon)^{\eta+1})$. We thus have for arbitrary δ, for $\varepsilon(\delta)$ sufficiently small, and for all $\varepsilon \leq \varepsilon(\delta)$,

$$\alpha(\varepsilon)^{\eta+1} \ln(\alpha(\varepsilon)) \leq (t + \delta) \ln f_2(d),$$

leading to

$$\eta + 1 + \frac{\ln \ln(\alpha(\varepsilon))}{\ln(\alpha(\varepsilon))} \leq \frac{\ln(t + \delta)}{\ln(\alpha(\varepsilon))} + \frac{\ln \ln f_2(d)}{\ln \ln(d)} \frac{\ln \ln(d)}{\ln(\alpha(\varepsilon))}.$$

This implies

$$\eta + 1 \leq \left(\liminf_{d \to \infty} \frac{\ln \ln f_2(d)}{\ln \ln(d)}\right) \left(\lim_{\varepsilon^{-1} \to \infty} \frac{\eta \ln(\alpha(\varepsilon)) + \ln \ln(\alpha(\varepsilon))}{\ln(\alpha(\varepsilon))}\right) = a_2 \eta.$$

Thus $\eta(a_2 - 1) \geq 1$. Letting η tend to $a_1 - 1$ we get $(a_1 - 1)(a_2 - 1) \geq 1$. This proves that $a_1 > 1$ and $a_2 > 1$. Furthermore, due to Theorem 8.12, B_2 has to be positive or infinite for any tractable problems with two positive eigenvalues $0 < \lambda_2 < \lambda_1 = 1$.

Assume now that $a_1 > 1$, $a_2 > 1$, $(a_1 - 1)(a_2 - 1) \geq 1$, and $B_2 > 0$. Due to Theorem 8.23, to prove $(T, \Omega^{\mathrm{unr}})$-tractability it is enough to verify that

$$F = \liminf_{\substack{\varepsilon^{-1}+d \to \infty \\ \varepsilon < \sqrt{\lambda_2}}} \frac{\ln f_1(\varepsilon^{-1}) + \ln f_2(d)}{\ln(\varepsilon^{-1})(1 + \ln(d))} \in (0, \infty].$$

Assume we have an arbitrary sequence $\{(\varepsilon_m^{-1}, d_m)\}$ such that $\{\varepsilon_m^{-1} + d_m\}$ tends to infinity, $\varepsilon_m < \sqrt{\lambda_2}$, and the sequence $\{F_m\}$, where

$$F_m := \frac{\ln f_1(\varepsilon_m^{-1}) + \ln f_2(d_m)}{\ln(\varepsilon_m^{-1})(1 + \ln(d_m))},$$

converges to F. Then we find a sub-sequence $\{(\varepsilon_n^{-1}, d_n)\}$ for which

$$\{\ln \ln(d_n) / \ln \ln(\varepsilon_n^{-1})\}$$

converges to an element $x \in [0, \infty]$. For this sub-sequence we show that $F > 2B / \ln(\lambda_2^{-1})$. If the sequence $\{\varepsilon_n^{-1}\}$ or $\{d_n\}$ is bounded, then $\{F_n\}$ tends to infinity, since a_1 and a_2 are both strictly larger than 1. So we can assume that $\{\varepsilon_n^{-1}\}$ as well as $\{d_n\}$ tend to infinity. First, let us assume that $x \in [0, (a_1 - 1))$. Then $\ln(d_n) \leq \ln(\varepsilon_n^{-1})^{a_1 - 1 - \delta}$ for δ sufficiently small and sufficiently large $n \geq n(\delta)$. Thus

$$F \geq \liminf_{n \to \infty} \frac{\ln f_1(\varepsilon_n^{-1})}{\ln(\varepsilon_n^{-1})^{a_1 - \delta}} = \infty.$$

If $x \in ((a_2 - 1)^{-1}, \infty]$, we just change the roles of the parameters ε^{-1} and d to get

$$F \geq \liminf_{n \to \infty} \frac{\ln f_2(d_n)}{\ln(d_n)^{a_2 - \delta}} = \infty.$$

If $(a_1 - 1)(a_2 - 1) > 1$, then we have considered all possible values of x in $[0, \infty]$ since then $[0, (a_1 - 1)) \cup ((a_2 - 1)^{-1}, \infty] = [0, \infty]$, and we have shown $F = \infty$. Theorem 8.23 implies then that the exponent of tractability is $t^{\mathrm{tra}} = 0$ and therefore $B_2 = \infty$.

If $(a_1 - 1)(a_2 - 1) = 1$, we still have to consider the case $x = a_1 - 1$. Then

$$\ln(\alpha(\varepsilon_n)) = \ln \ln(\varepsilon_n^{-1}) + \ln(2) - \ln \ln(\lambda_2^{-1}) + \mathcal{O}(1) \in [(a_2 - 1) - \delta, (a_2 - 1) + \delta] \ln \ln(d_n)$$

for arbitrary δ and sufficiently large $n \geq n(\delta)$. Then $\alpha(\varepsilon_n) \leq (\ln d_n)^{a_2 - 1 + \delta} = o(d_n)$. Hence we have

$$F = \liminf_{n \to \infty} \frac{\ln f_1(\varepsilon_n^{-1}) + \ln f_2(d_n)}{\alpha(\varepsilon_n)(1 + \ln(d_n/\alpha(\varepsilon_n)))} \frac{\alpha(\varepsilon_n)(1 + \ln(d_n/\alpha(\varepsilon_n)))}{\ln(\varepsilon_n^{-1})(1 + \ln(d_n))} = B_2 \frac{2}{\ln(\lambda_2^{-1})} > 0.$$

To obtain the formula for the exponent t^{tra} we can use the bound on t^{tra} from Theorem 8.23. For $\beta \geq \ln \lambda_2^{-1}$ we get $t^{\text{tra}} = B_2^{-1}$. To obtain the same result for $\beta < \ln \lambda_2^{-1}$ we proceed as follows. In the proof of Theorem 8.23 we showed that for small positive δ there is a positive number $C_{\beta,\delta}$ depending only on β and δ such that

$$n(\varepsilon, d) \leq C_{\beta,\delta} \exp\left(-\max\left\{\frac{2+\delta}{\beta}, \frac{2(1+\delta)\ln(d)}{\ln(\lambda_2^{-1})}\right\} \ln(\varepsilon^{-1})\right)$$

for all $(\varepsilon^{-1}, d) \in \Omega^{\text{unr}}$.

To show that the last right side function is at most $C\left(f_1(\varepsilon^{-1}) f_2(d)\right)^t$ it is enough to check that

$$\frac{2(1+\delta)}{\ln(\lambda_2^{-1})} \ln(\varepsilon^{-1}) \ln(d) \leq t \left(\ln(f_1(\varepsilon^{-1})) + \ln(f_2(d))\right)$$

for large ε^{-1} and d. Or equivalently that

$$t \geq (1+\delta)\left(\liminf_{\substack{\varepsilon^{-1}+d\to\infty \\ \varepsilon<\sqrt{\lambda_2}}} \frac{\ln(f_1(\varepsilon^{-1})) + \ln(f_2(d))}{\alpha(\varepsilon)\ln(d)}\right)^{-1}.$$

The last limit inferior is achieved if $\alpha(\varepsilon)$ is a power of $\ln(d)$, and therefore it is the same as B_2. Since δ can be arbitrarily small we conclude that $t^{\text{tra}} \leq B_2^{-1}$. The lower bound on t^{tra} from Theorem 8.23 then implies $t^{\text{tra}} = B_2^{-1}$, as claimed. This completes the proof of Theorem 8.25. $\square$

Remark 8.26. Let the conditions of Theorem 8.25 hold and assume that $a_1 > 1, a_2 > 1$ and $(a_1 - 1)(a_2 - 1) = 1$. Then condition $B_2 \in (0, \infty]$ does not necessarily hold as the following example shows. Let $\delta : [1, \infty) \to [0, \infty)$ be a decreasing function with $\lim_{x\to\infty} \delta(x) = 0$. Define

$$f_i(x) = \exp\left(\ln(x)^{2-\delta(x)}\right) \quad \text{for } i = 1, 2.$$

Then we have obviously $a_1 = 2 = a_2$ and $(a_1 - 1)(a_2 - 1) = 1$. But

$$\left(\ln \lambda_2^{-1}\right)^{-1} B_2 \leq \liminf_{\substack{\varepsilon^{-1}+d\to\infty \\ \varepsilon^{-1}=d}} \frac{\ln(\varepsilon^{-1})^{2-\delta(\varepsilon^{-1})} + \ln(d)^{2-\delta(d)}}{\ln(\varepsilon^{-1})\ln(d)} = 2\liminf_{d\to\infty} \ln(d)^{-\delta(d)}$$

$$= 2\liminf_{d\to\infty} \exp\left(-\delta(d)\ln\ln(d)\right).$$

If we choose, e.g., $\delta(x) = (\ln\ln\ln(x))^{-1}$, then we see that $B_2 = 0$.

We stress again that the exponent of tractability in Theorem 8.25 does not depend on β and it is B_2^{-1} for all polynomial decaying eigenvalues with the same two largest eigenvalues $0 < \lambda_2 < \lambda_1 = 1$. However, B_2 depends on particular functions f_i satisfying the conditions of Theorem 8.25. We now show that B_2 can take any positive

value or even be infinite. Indeed, take $f_i(x) = \exp\left(c_i[\ln x]^{(1+\alpha_i)}\right)$ for positive c_i and α_i. Then $a_i = 1 + \alpha_i$. For $\alpha_1\alpha_2 = 1$ it can be checked that

$$B_2 = c_2(1+\alpha_2)\left(\frac{c_1\alpha_1}{c_2}\right)^{1/(1+\alpha_1)}\frac{\ln(\lambda_2^{-1})}{2}. \tag{8.48}$$

Taking, $c_2 = c_1 = c$ and varying c for fixed α_i, we see that B_2 can be any positive number with the same limits a_i.

On the other hand, for $f_i(x) = \exp\left(\ln(e + \ln x)[\ln x]^{1+\alpha_i}\right)$, and $\alpha_1\alpha_2 = 1$ we get $a_i = 1 + \alpha_i$ as before, but $B_2 = \infty$.

We also stress that in Theorem 8.25 we assume that the eigenvalues decay *at least* polynomially. This assumption holds, in particular, for finitely many positive or exponentially decaying eigenvalues. We summarize this discussion in the following remark.

Remark 8.27. As long as a tractability function T is of product form, $T(x, y) = f_1(x)f_2(x)$, then $(T, \Omega^{\mathrm{unr}})$-tractability of S as well as the exponent of tractability depend *only* on the functions f_1, f_2 and the second eigenvalue λ_2 as long as the eigenvalues λ_j decay at least polynomially. Hence, if we have two problems, one with only two positive eigenvalues $0 < \lambda_2 < \lambda_1 = 1$, and the second with the same two eigenvalues and the rest of them are non-negative and decaying polynomially, then these two problems lead to the same tractability conditions and to the same exponents of tractability.

We stress that this property does *not* hold for more general tractability functions. For instance, if we consider $T(x, y) = \exp(g_1(x)g_2(y))$, i.e, when $\ln T$ is of product form, then the exponent of tractability may depend on the rate of decay of eigenvalues. This holds, for instance, for $T(x, y) = \exp(\ln(x)(1+\ln(d)))$ as shown after Corollary 8.24.

8.5 Comparison

We briefly compare tractability results for the restricted and unrestricted domains. We consider linear tensor product problems S with $\varepsilon_0^2 < \lambda_2 < \lambda_1 = 1$. Then

- Strong $(T, \Omega^{\mathrm{unr}})$-tractability of S as well as strong $(T, \Omega^{\mathrm{res}})$-tractability of S does not hold regardless of the tractability function T.

- Consider finitely many, say k, positive eigenvalues as in Section 8.4.1.

  It is easy to see from (8.30) that for (ε, d) with $d \le d^*$, the information complexity $n(\varepsilon, d)$ is uniformly bounded in ε^{-1}. Therefore the more interesting case is when $(\varepsilon^{-1}, d) \in [1, \varepsilon_0^{-1}) \times \mathbb{N}$. Then $n(\varepsilon, d) = \Theta(d^{\alpha(\varepsilon)})$ with the factors in the Theta-notation only dependent on ε_0, λ_2 and k. So we have a polynomial dependence on d which obviously implies weak tractability. It follows from

Theorem 8.3 that $(T, \Omega^{\mathrm{res}})$-tractability of S holds iff

$$B_{\mathrm{res}} := \liminf_{d \to \infty} \inf_{1 \le \alpha(\varepsilon) \le \alpha(\varepsilon_0)} \frac{\ln T(\varepsilon^{-1}, d)}{\alpha(\varepsilon) \ln(d)} \in (0, \infty], \qquad (8.49)$$

and the exponent of tractability is $1/B_{\mathrm{res}}$.

In particular, we have polynomial tractability, i.e., when $T(x, y) = xy$, with the exponent

$$\alpha(\varepsilon_0) = \left\lceil \frac{2 \ln(\varepsilon_0^{-1})}{\ln(\lambda_2^{-1})} \right\rceil - 1.$$

This exponent can be arbitrarily large if ε_0 is small or λ_2 close to one. On the other hand, it is interesting that the exponent does not depend on the total number k of positive eigenvalues.

As we already said, for the unrestricted domain Ω^{unr} we do *not* have polynomial tractability of S. This agrees with the fact that the exponent of polynomial tractability for the restricted domain goes to infinity as ε_0 approaches 0, and for the unrestricted domain formally $\varepsilon_0 = 0$.

- Consider exponentially decaying eigenvalues $\lambda_j = \exp(-\beta(j - 1))$ for a positive β. Then Theorem 8.10 states that $(T, \Omega^{\mathrm{res}})$-tractability of S holds iff

$$A_{\mathrm{e,res}} := \liminf_{x \to \infty} \frac{\ln T(x, 1)}{\ln \ln(x)} \in (0, \infty] \quad \text{and} \quad B_{\mathrm{res}} \in (0, \infty],$$

where B_{res} is given by (8.49). Furthermore, if $A_{\mathrm{e,res}} = \infty$ then the exponent of tractability is B_{res}^{-1}.

Hence, we again have polynomial tractability, and indeed since $A_{\mathrm{e,res}} = \infty$ and $\lambda_2 = \exp(-\beta)$, the exponent of polynomial tractability is

$$\alpha(\varepsilon_0) = \left\lceil \frac{2 \ln \varepsilon_0^{-1}}{\beta} \right\rceil - 1.$$

As we know, for the unrestricted domain Ω^{unr} we do not have polynomial tractability.

Take now $T(x, y) = x^{1+\ln y}$. Then $A_{\mathrm{e,res}} = \infty$ and $B_{\mathrm{res}} = \beta/2$. Furthermore, as we already know, $B_{\mathrm{e}}^{(2)} = \beta/2$. So we have $(T, \Omega^{\mathrm{res}})$-tractability as well as $(T, \Omega^{\mathrm{unr}})$-tractability with the same exponents $2/\beta$. Hence, there is no much difference between the restricted and unrestricted domains for this particular tractability function.

Note also the difference in the exponents for the last two tractability functions and for the restricted domain. For polynomial tractability, the exponent depends on ε_0 and goes to infinity as ε_0 approaches 0. For the second tractability function, the exponent does not depend on ε_0.

- Consider polynomially decaying eigenvalues $\lambda_j = \Theta(j^{-\beta})$ for a positive β. Then Theorem 8.10 states that (T, Ω^{res})-tractability of S holds iff

$$A_{\text{p,res}} := \liminf_{x \to \infty} \frac{\ln T(x, 1)}{\ln(x)} \in (0, \infty] \quad \text{and} \quad B_{\text{res}} \in (0, \infty].$$

If this holds then the exponent of tractability is

$$t^{\text{tra}} = \max\{2/(\beta A_{\text{p,res}}), 1/B_{\text{res}}\}.$$

Let us consider polynomial tractability, i.e., $T(x, y) = xy$. Then we have $A_{\text{p,res}} = 1$ and, as stated above, $B_{\text{res}} = \alpha(\varepsilon_0)^{-1}$. Due to Theorem 8.10 we have (T, Ω^{res})-tractability with $t^{\text{tra}} = \max\{2/\beta, \alpha(\varepsilon_0)\}$ but, as already mentioned, no (T, Ω^{unr})-tractability.

Take now $T(x, y) = \exp(\ln^2 x)\exp(\ln^2 y)$. Then $A_{\text{p,res}} = B_{\text{res}} = \infty$, and S is (T, Ω^{res})-tractable with $t^{\text{tra}} = 0$. For the unrestricted case, we conclude from (8.48) that S is (T, Ω^{unr})-tractable with $t^{\text{tra}} = (\ln(\lambda_2^{-1}))^{-1}$. Hence, we have tractability in both cases but the exponents are quite different.

Let now $T(x, y) = x^{1+\ln y}$. Then $A_{\text{p,res}} = 1$ and $B_{\text{res}} = \beta/2$. Thus S is (T, Ω^{res})-tractable with $t^{\text{tra}} = 2/\beta$, see also Theorem 8.10. As already stated, we have also (T, Ω^{unr})-tractability with the exponent of tractability $t^{\text{tra}} = \max\{2/\beta, 2/\ln(\lambda_2^{-1})\}$.

8.6 Notes and Remarks

NR 8:1. As already mentioned, this chapter is basically equivalent to the two papers [69], [70].

Appendix A
Reproducing Kernel Hilbert Spaces of Sobolev Type

Sobolev spaces are used in many applications. Examples include the solution of important computational problems such as differential and integral equations, and problems in financial mathematics. There are many variants of Sobolev spaces. In this book, we usually study Sobolev spaces that are also Hilbert spaces. Moreover, we consider spaces which are often called *Sobolev spaces of dominating mixed smoothness*. A recent survey about such spaces can be found in the paper of Schmeisser [209].

We wish to stress an important difference between the theory of function spaces and the study of tractability. In the theory of function spaces, one often studies properties that do not change if a norm is replaced by an equivalent norm. This is different from the study of tractability since tractability very much depends on the specific norm. The reader will notice that some of the spaces considered here are identical as vector spaces with equivalent norms but lead to different tractability results. The reason is that the corresponding numbers for the equivalence of norms usually depend exponentially on the number of variables.

A.1 Korobov Spaces

Korobov spaces are probably the most important spaces for the study of computational problems for periodic smooth functions. These spaces have also many interesting applications for non-periodic functions due to interesting relations and estimates between the complexity for the periodic and non-periodic computational problems such as multivariate integration, see Volume II.

The Korobov space $H_d^{\text{Kor}} = H_{d,\alpha}$ is a separable Hilbert space that consists of complex-valued functions defined on $[0, 1]^d$. The parameter $\alpha \geq 0$ measures the smoothness of these functions. For $\alpha = 0$, the Korobov space $H_d^{\text{Kor}} = L_2([0, 1]^d)$ is the same as the L_2 space of square Lebesgue integrable functions. For $\alpha > 0$, the Korobov space H_d^{Kor} is a proper subset of $L_2([0, 1]^d)$. For $\alpha > 1/2$ it consists of functions that are periodic in each variable with period 1 and enjoy some smoothness property measured by α.

For $h = [h_1, h_2, \ldots, h_d] \in \mathbb{Z}^d$ and $f \in L_2([0, 1]^d)$, let $\hat{f}(h)$ denote the Fourier coefficient of f,

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

where $\mathrm{i} = \sqrt{-1}$ and $h \cdot x = h_1 x_1 + h_2 x_2 + \cdots + h_d x_d$.

We consider a little more general case of the Korobov spaces than usually studied

and define the inner product and the norm in $H_{d,\alpha}$ depending on two positive parameters β_1 and β_2. We then discuss various values of β_j and relate them to the cases studied before.

For $\alpha \geq 0$, define

$$\varrho_{d,\alpha}(h) = \prod_{j=1}^{d} \left(\beta_1^{-1} \delta_{0,h_j} + \beta_2^{-1} (1 - \delta_{0,h_j}) |h_j|^{2\alpha} \right),$$

where δ is the Kronecker delta function. For $h = 0$, we have $\varrho_{d,\alpha}(0) = \beta_1^{-d}$. For $\alpha = 0$, we have

$$\varrho_{d,0}(h) = \beta_1^{-|\{j\,|\,h_j=0\}|} \beta_2^{-(d-|\{j\,|\,h_j=0\}|)}.$$

The Korobov space H_d^{Kor} consists of functions for which

$$\|f\|_{H_d^{\mathrm{Kor}}} = \left(\sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}(h) |\hat{f}(h)|^2 \right)^{1/2} < \infty.$$

The inner product is defined for $f, g \in H_{d,\alpha}$ as

$$\langle f, g \rangle_{H_{d,\alpha}} = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}(h) \hat{f}(h) \overline{\hat{g}(h)}.$$

As in [123], we have adopted an unconventional notation in which the parameter α in the traditional notation is replaced by 2α. This is done, as we shall see, to simplify the expression for the Korobov norm in terms of derivatives. The traditional notation was used in e.g., [45], [95], [172], [218].

Korobov spaces have been studied for multivariate integration and approximation in many papers for $\beta_1 = \beta_2 = 1$, and with our α replaced by $\alpha/2$. There are too many papers to cite, and we only give the reference to the book of Sloan and Joe [214]. The case of $\beta_1 = 1$ and $\beta_2 = (2\pi)^{-2\alpha}$, was studied in [123]. As we shall see in a moment, the latter choice of β_j simplifies the relations between the Korobov norm and the norm involving derivatives of functions.

In the theory of function spaces, Korobov spaces are sometimes called *periodic Sobolev spaces of dominating mixed smoothness*, see e.g., Sickel and Ullrich [212] and the survey Schmeisser [209]. Also the notion of *functions with bounded mixed derivative* is very common, see the book of Temlyakov [232].

For $h \in \mathbb{Z}^d$, define

$$e_h(x) = \varrho_{d,\alpha}(h)^{-1/2} \exp\left(2\pi \mathrm{i} h \cdot x\right).$$

It is easy to check that for $\tau \in \mathbb{Z}^d$ we have $\hat{e}_h(\tau) = \varrho_{d,\alpha}(h)^{-1/2} \delta_{h,\tau}$. This implies that

$$\langle e_h, e_\tau \rangle_{H_{d,\alpha}} = \delta_{h,\tau} \quad \text{for all } h, \tau \in \mathbb{Z}^d.$$

Hence, $\{e_h\}_{h \in \mathbb{Z}^d}$ is an orthonormal basis of $H_{d,\alpha}$. Note that each e_h is of the product form,

$$e_h(x) = \prod_{j=1}^{d} e_{h_j}(x_j)$$

with

$$e_{h_j}(x_j) = \left(\beta_1^{-1}\delta_{0,h_j} + \beta_2^{-1}(1 - \delta_{0,h_j})|h_j|^{2\alpha}\right)^{-1/2} \exp\left(2\pi i h_j x_j\right).$$

This easily yields that $H_{d,\alpha}$ is the tensor product Hilbert space

$$H_{d,\alpha} = H_{1,\alpha} \otimes H_{1,\alpha} \otimes \cdots \otimes H_{1,\alpha}, \quad d \text{ times}.$$

Here, $H_{1,\alpha}$ is the Hilbert space of univariate functions with the inner product

$$\langle f, g \rangle_{H_{1,\alpha}} = \sum_{h \in \mathbb{Z}} \varrho_{1,\alpha}(h) \hat{f}(h)\overline{\hat{g}(h)},$$

and $\{e_h\}_{h \in \mathbb{Z}}$ is its orthonormal basis.

The smoothness parameter α moderates the rate of decay of the Fourier coefficients of the functions and guarantees the existence of some derivatives of the functions. To see this more clearly, assume for a moment that $d = 1$ and let $\alpha = r \geq 1$ be an integer. Then

$$\|f\|_{H_{1,r}}^2 = \beta_1^{-1}|\hat{f}(0)|^2 + \beta_2^{-1} \sum_{h \in \mathbb{Z}, h \neq 0} |h|^{2r}|\hat{f}(h)|^2 < \infty.$$

On the other hand, we know that

$$f(x) = \sum_{h \in \mathbb{Z}} \hat{f}(h) \exp(2\pi i h \cdot x).$$

From this it follows that f is r times differentiable, and its kth derivative is absolutely continuous for $k = 1, 2, \ldots, r - 1$, whereas the rth derivative belongs to L_2. Indeed, for $k = 1, 2, \ldots, r$, we have

$$f^{(k)}(x) = (2\pi i)^k \sum_{h \in \mathbb{Z}, h \neq 0} h^k \hat{f}(h) \exp(2\pi i h \cdot x).$$

For $k \leq r - 1$, the last series is pointwise convergent since

$$|f^{(k)}(x)| \leq (2\pi)^k \sum_{h \in \mathbb{Z}, h \neq 0} |\hat{f}(h)|[\varrho_{1,r}(h)]^{1/2}|h^k|/[\varrho_{1,r}(h)]^{1/2}$$

$$\leq (2\pi)^k \left(\sum_{h \in \mathbb{Z}} \varrho_{1,r}(h)|\hat{f}(h)|^2\right)^{1/2} \left(\beta_2 \sum_{h \in \mathbb{Z}, h \neq 0} |h^{2k}|/|h|^{2r}\right)^{1/2}$$

$$= 2(2\pi)^k \beta_2^{1/2}\|f\|_{H_{1,r}} \zeta(2(r - k))^{1/2} < \infty,$$

with ζ being, as always, the Riemann zeta function. For $k = r$ this argument yields that $f^{(r)}$ belongs to $L_2[0, 1]$.

Absolute continuity for $k < r - 1$ is obvious since $f^{(k)}$ is even Lipschitz. For $k = r - 1$ it is easy to check that $f^{(r-1)}(x) = f^{(r-1)}(a) + \int_a^x f^{(r)}(t)\,dt$ holds for any $a \in [0, 1]$ which is equivalent to absolute continuity.

Orthonormality of the functions $\{\exp(2\pi i h \cdot x)\}$ in $L_2([0, 1[)$ implies that

$$\int_0^1 |f^{(r)}(x)|^2 \, dx = (2\pi)^{2r} \sum_{h \in \mathbb{Z}, h \neq 0} |h|^{2r} |\hat{f}(h)|^2.$$

Since $\hat{f}(0) = \int_0^1 f(x)\,dx$, this means that

$$\|f\|_{H_{1,r}}^2 = \beta_1^{-1} \left| \int_0^1 f(x)\,dx \right|^2 + \beta_2^{-1}(2\pi)^{-2r} \int_0^1 |f^{(r)}(x)|^2 \, dx.$$

For $\beta_1 = 1$ and $\beta_2 = (2\pi)^{-2r}$ we have an especially pleasing relation,

$$\|f\|_{H_{1,r}}^2 = \left| \int_0^1 f(x)\,dx \right|^2 + \int_0^1 |f^{(r)}(x)|^2 \, dx. \tag{A.1}$$

Hence, for $d = 1$ and any arbitrary positive β_1 and β_2, the space $H_{1,r}$ consists of periodic functions whose derivatives up to the $(r-1)$st are absolutely continuous and the rth derivative belongs to L_2.

Assume now that $d \geq 1$. As always, $[d] = \{1, 2, \ldots, d\}$. For $h \in \mathbb{Z}^d$ define $\mathfrak{u}_h = \{j \in [d] \mid h_j \neq 0\}$. For any $\mathfrak{u} \subseteq [d]$, let

$$\mathbb{Z}_\mathfrak{u} = \{h \in \mathbb{Z}_d \mid \mathfrak{u}_h = \mathfrak{u}\}.$$

It is obvious that the sets $\mathbb{Z}_\mathfrak{u}$ are disjoint and $\mathbb{Z}_d = \bigcup_{\mathfrak{u} \subseteq [d]} \mathbb{Z}_\mathfrak{u}$. We can then rewrite the inner product of $H_{d,\alpha}$ as

$$\langle f, g \rangle_{H_{d,\alpha}} = \sum_{\mathfrak{u} \subseteq [d]} \sum_{h \in \mathbb{Z}_\mathfrak{u}} \beta_1^{-(d-|\mathfrak{u}|)} \beta_2^{-|\mathfrak{u}|} \hat{f}(h) \overline{\hat{g}(h)} \prod_{j \in \mathfrak{u}} |h_j|^{2\alpha}.$$

For an integer $\alpha = r \geq 1$, define the differentiation operator

$$D_{\mathfrak{u},r} f = \frac{\partial^{r|\mathfrak{u}|}}{\prod_{j \in \mathfrak{u}} \partial x_j^r} f.$$

For $\mathfrak{u} = \emptyset$, we have $D_{\emptyset,r} f = f$. We also need to define the integration operator,

$$I_{-\mathfrak{u}} f(x) = \int_{[0,1]^{d-|\mathfrak{u}|}} f(x) \, dx_{-\mathfrak{u}},$$

where we integrate over variables not in the set $\mathfrak{u}$ and variables in $\mathfrak{u}$ are intact. For $\mathfrak{u} = [d]$, we have $I_{-[d]} f = f$. Finally, let

$$V_{\mathfrak{u},r} f = D_{\mathfrak{u},r} I_{-\mathfrak{u}} f.$$

In particular, we have $V_{\emptyset,r} f = \int_{[0,1]^d} f(x)\, dx$, and $V_{[d],r} f = f^{(r,r,\dots,r)}$.

We claim that

$$\langle f, g\rangle_{H_{d,r}} = \sum_{\mathfrak{u}\subseteq[d]} \beta_1^{-(d-|\mathfrak{u}|)}\beta_2^{-|\mathfrak{u}|}(2\pi)^{-2r|\mathfrak{u}|}\int_{[0,1]^d} V_{\mathfrak{u},r}f(x)\overline{V_{\mathfrak{u},r}g(x)}\,dx.$$

Indeed, using again $f(x) = \sum_{h\in\mathbb{Z}^d} \hat{f}(h)\exp(2\pi ih\cdot x)$ we obtain

$$D_{\mathfrak{u},r}f(x) = \sum_{h\in\mathbb{Z}^d} \hat{f}(h)(2\pi i)^{r|\mathfrak{u}|}\left(\prod_{j\in\mathfrak{u}} h_j^r\right)\exp(2\pi ih\cdot x).$$

Observe that if $h_j = 0$ for $j\in\mathfrak{u}$ then the corresponding term in the last sum is 0. Therefore we can sum up only terms with h such that $\mathfrak{u}\subseteq\mathfrak{u}_h$, i.e.,

$$D_{\mathfrak{u},r}f(x) = \sum_{h\in\mathbb{Z}^d,\mathfrak{u}\subseteq\mathfrak{u}_h} \hat{f}(h)(2\pi i)^{r|\mathfrak{u}|}\left(\prod_{j\in\mathfrak{u}} h_j^r\right)\exp(2\pi ih\cdot x).$$

This leads to

$$\begin{aligned}
V_{\mathfrak{u},r}f(x) &= \sum_{h\in\mathbb{Z}^d,\mathfrak{u}\subseteq\mathfrak{u}_h} \hat{f}(h)(2\pi i)^{r|\mathfrak{u}|}\left(\prod_{j\in\mathfrak{u}} h_j^r\right)\int_{[0,1]^{d-|\mathfrak{u}|}}\exp(2\pi ih\cdot x)\,dx_{-\mathfrak{u}}\\
&= \sum_{\substack{h\in\mathbb{Z}^d,\mathfrak{u}\subseteq\mathfrak{u}_h\\ h_j=0\text{ for all }j\notin\mathfrak{u}}} \hat{f}(h)(2\pi i)^{r|\mathfrak{u}|}\left(\prod_{j\in\mathfrak{u}} h_j^r\right)\exp\left(2\pi i\sum_{j\in\mathfrak{u}} h_j x_j\right)\\
&= \sum_{h\in\mathbb{Z}_{\mathfrak{u}}} \hat{f}(h)(2\pi i)^{r|\mathfrak{u}|}\left(\prod_{j\in\mathfrak{u}} h_j^r\right)\exp\left(2\pi i\sum_{j\in\mathfrak{u}} h_j x_j\right).
\end{aligned}$$

Hence

$$\int_{[0,1]^d} V_{\mathfrak{u},r}f(x)\overline{V_{\mathfrak{u},r}g(x)}\,dx = (2\pi)^{2r|\mathfrak{u}|}\sum_{h\in\mathbb{Z}_{\mathfrak{u}}} \hat{f}(h)\overline{\hat{g}(h)}\left(\prod_{j\in\mathfrak{u}} h_j^{2r}\right),$$

from which our claim follows. Again for

$$\beta_1 = 1 \quad\text{and}\quad \beta_2 = (2\pi)^{-r}$$

we have a pleasing relation

$$\langle f, g\rangle_{H_{d,r}} = \sum_{\mathfrak{u}\in[d]}\int_{[0,1]^d} V_{\mathfrak{u},r}f(x)\overline{V_{\mathfrak{u},r}g(x)}\,dx.$$

This formula for $f = g$ and $d = 1$ agrees with (A.1). For $d = 2$, it takes the form

$$\begin{aligned}
\|f\|_{H_{2,r}}^2 &= \left|\int_{[0,1]^2} f(x_1,x_2)\,dx_1\,dx_2\right|^2 + \int_0^1\left|\int_0^1 \frac{\partial f(x_1,x_2)}{\partial x_1}\,dx_2\right|^2 dx_1\\
&\quad + \int_0^1\left|\int_0^1 \frac{\partial f(x_1,x_2)}{\partial x_2}\,dx_1\right|^2 dx_2 + \int_{[0,1]^2}\left|\frac{\partial^2 f(x_1,x_2)}{\partial x_1\partial x_2}\right|^2 dx_1\,dx_2.
\end{aligned}$$

The Korobov space $H_{d,\alpha}$ is a reproducing kernel Hilbert spaces iff $\alpha > 1/2$. Indeed, for $\alpha > 1/2$ and $x, y \in [0, 1]^d$ define the kernel

$$K_{d,\alpha}(x, y) = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}^{-1}(h) \exp(2\pi i h \cdot (x - y)). \qquad \text{(A.2)}$$

Since

$$\varrho_{d,\alpha}^{-1}(h) = \prod_{j=1}^{d} \left(\beta_1^{-1} \delta_{0,h_j} + \beta_2^{-1}(1 - \delta_{0,h_j}|h_j|^{2\alpha})^{-1}\right)$$

$$= \prod_{j=1}^{d} \left(\beta_1 \delta_{0,h_j} + \beta_2(1 - \delta_{0,h_j})|h_j|^{-2\alpha}\right),$$

we easily check that K_d is well defined and

$$|K_{d,\alpha}(x, y)| \leq K_{d,\alpha}(x, x) = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}^{-1}(h) = \prod_{j=1}^{d} (\beta_1 + 2\beta_2 \zeta(2\alpha)) < \infty.$$

Note that to claim that $\zeta(2\alpha)$ is well defined we must assume that $\alpha > 1/2$.

To show that $K_{d,\alpha}$ is a reproducing kernel of $H_{d,\alpha}$ observe that

$$\widehat{K_{d,\alpha}(\cdot, x)}(h) = \varrho_{d,\alpha}^{-1}(h) \exp\left(-2\pi i h \cdot x\right),$$

and therefore

$$\langle f, K_{d,\alpha}(\cdot, x)\rangle_{H_{d,\alpha}} = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}(h) \hat{f}(h) \overline{\widehat{K_{d,\alpha}(\cdot, x)}(h)}$$

$$= \sum_{h \in \mathbb{Z}^d} \hat{f}(h) \exp\left(2\pi i h \cdot x\right) = f(x),$$

as needed.

On the other hand, if $H_{d,\alpha}$ is a reproducing kernel Hilbert space then

$$L_x(f) = f(x) \quad \text{for all } f \in H_{d,\alpha},$$

is a continuous linear functional for any $x \in [0, 1]^d$. Indeed,

$$f(x) = \sum_h \varrho_{d,\alpha}^{1/2}(h) \hat{f}(h) \varrho_{d,\alpha}^{-1/2}(h) \exp(2\pi i h \cdot x)$$

yields

$$|f(x)|^2 \leq \|f\|_{H_{d,\alpha}}^2 \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha}^{-1}(h) = \|f\|_{H_{d,\alpha}}^2 \prod_{j=1}^{d} (\beta_1 + 2\beta_2 \zeta(2\alpha)).$$

Note that the last estimate is sharp. Hence it is finite iff $\alpha > 1/2$, as claimed.

The reproducing kernel $K_{d,\alpha}$ can also be written in the product form as

$$K_{d,\alpha}(x, y) = \prod_{j=1}^{d} K_{1,\alpha}(x_j, y_j)$$

with

$$K_{1,\alpha}(x_j, y_j) = \sum_{h \in \mathbb{Z}} \varrho_{1,\alpha}^{-1}(h) \exp\left(2\pi i h(x_j - y_j)\right)$$

$$= \beta_1 + 2\beta_2 \sum_{h=1}^{\infty} \frac{\cos\left(2\pi h(x_j - y_j)\right)}{h^{2\alpha}}.$$

If $\alpha = r$ is a positive integer, then the reproducing kernel $K_{d,\alpha}$ is related to the Bernoulli polynomial B_{2r} of degree $2r$ since B_{2r} has the Fourier series

$$B_{2r}(x) = (-1)^{r+1}(2r)! \sum_{h \in \mathbb{Z}, h \neq 0} (2\pi h)^{-2r} \exp\left(2\pi i h x\right)$$

$$= \frac{2(-1)^{r+1}(2r)!}{(2\pi)^{2r}} \sum_{h=1}^{\infty} \frac{\cos(2\pi h x)}{h^{2r}}.$$

Then

$$K_{d,\alpha}(x, y) = \prod_{j=1}^{d} \left(\beta_1 + \beta_2(-1)^{r+1}(2\pi)^{2r} \frac{B_{2r}(\{x_j - y_j\})}{(2r)!}\right),$$

where $\{x_j - y_j\}$ denotes the fractional part of $x_j - y_j$.

A.1.1 Weighted Korobov Spaces

We define the weighted Korobov space $H_{d,\gamma}^{\mathrm{Kor}} = H_{d,\alpha,\gamma}$ for $\alpha \geq 0$ and a sequence of non-negative weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, where $d \in \mathbb{N}$ and $\mathfrak{u}$ is, as always, an arbitrary subset of $[d] := \{1, 2, \ldots, d\}$.

To obtain the weighted Korobov space we only need to generalize $\varrho_{d,\alpha}(h)$ which are used for the inner product and norm of the Korobov space H_d^{Kor}. For $h \in \mathbb{Z}^d$, we already used

$$\mathfrak{u}_h = \{j \in [d] \mid h_j \neq 0\}$$

as the set of indices of h with non-zero components. Let

$$\varrho_{d,\alpha,\gamma}(h) = \gamma_{d,\mathfrak{u}_h}^{-1} \varrho_{d,\alpha}(h) = \frac{1}{\gamma_{d,\mathfrak{u}_h} \beta_1^{d-|\mathfrak{u}_h|} \beta_2^{|\mathfrak{u}_h|}} \prod_{j \in \mathfrak{u}_h} |h_j|^{2\alpha}.$$

For $\gamma_{d,\mathfrak{u}_h} = 0$ we formally set $\varrho_{d,\alpha,\gamma}(h) = \infty$.

Then we proceed as in the previous section with the only difference that we take $\varrho_{d,\alpha,\gamma}(h)$ instead of $\varrho_{d,\alpha}(h)$. Hence, the weighted Korobov space $H_{d,\gamma}^{\mathrm{Kor}}$ consists of complex-valued functions defined on $[0,1]^d$ for which

$$\|f\|_{H_{d,\gamma}^{\mathrm{Kor}}} = \Big(\sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha,\gamma}(h)|\hat{f}(h)|^2 \Big)^{1/2} < \infty.$$

Here, we adopt the notation that if $\varrho_{d,\alpha,\gamma}(h) = \infty$ then we have $\hat{f}(h) = 0$ and interpret the product $\infty\, 0$ as 0. Hence, if $\gamma_{d,\mathfrak{u}} = 0$ then $\hat{f}(h) = 0$ for all indices h for which $\mathfrak{u}_h = \mathfrak{u}$.

The inner product for $f, g \in H_{d,\alpha,\gamma}$ is, of course, defined as

$$\langle f, g \rangle_{H_{d,\alpha,\gamma}} = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha,\gamma}(h)\,\hat{f}(h)\overline{\hat{g}(h)}.$$

Clearly, for $\gamma_{d,\mathfrak{u}} \equiv 1$ we have $H_{d,\gamma}^{\mathrm{Kor}} = H_d^{\mathrm{Kor}}$.

For $\alpha = 0$, the weighted Korobov space is a weighted $L_2([0,1]^d)$ which is algebraically the same as $L_2([0,1]^d)$ if all $\gamma_{d,\mathfrak{u}} > 0$. For $\alpha > 1/2$, the weighted Korobov space is a separable Hilbert space consisting of functions which are periodic in each variable with period 1.

For $h \in \mathbb{Z}^d$ and $\gamma_{d,\mathfrak{u}_h} > 0$, define

$$e_{h,\gamma}(x) = \varrho_{d,\alpha,\gamma}(h)^{-1/2} \exp\left(2\mathrm{i}\pi h \cdot x\right)$$
$$= \gamma_{d,\mathfrak{u}_h}^{1/2}\, \beta_1^{(d-|\mathfrak{u}_h|)/2} \beta_2^{|\mathfrak{u}_h|/2} \prod_{j \in \mathfrak{u}_h} \frac{\exp\left(2\pi \mathrm{i} h_j x_j\right)}{|h_j|^\alpha}.$$

It is easy to check that $\{e_{h,\gamma}\}_{h \in \mathbb{Z}^d}$ is an orthonormal basis of $H_{d,\alpha,\gamma}$. Note that $e_{h,\gamma}$ depends only on variables belonging to $\mathfrak{u}_h$.

For general weights, the space $H_{d,\alpha,\gamma}$ is *not* a tensor product Hilbert space. However, for product weights, $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ for non-empty $\mathfrak{u}$, it is a tensor product Hilbert space of the form

$$H_{d,\alpha,\gamma} = H_{1,\alpha,\gamma_{d,1}} \otimes H_{1,\alpha,\gamma_{d,2}} \otimes \cdots \otimes H_{1,\alpha,\gamma_{d,d}}$$

with the Hilbert space $H_{1,\alpha,\gamma_{d,j}}$ of univariate functions and the inner product

$$\langle f, g \rangle_{H_{1,\alpha,\gamma_{d,j}}} = \sum_{h \in \mathbb{Z}} \varrho_{1,\alpha,\gamma_{d,j}}(h)\,\hat{f}(h)\overline{\hat{g}(h)}$$
$$= \beta_1^{-1}\hat{f}(0)\overline{\hat{g}(0)} + \gamma_{d,j}^{-1}\beta_2^{-1} \sum_{h \in \mathbb{Z}, h \neq 0} \hat{f}(h)\overline{\hat{g}(h)}|h|^{2\alpha}.$$

For general weights $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, functions in $H_{d,\alpha,\gamma}$ are at least as smooth as functions in $H_{d,\alpha}$, and for non-zero weights $\gamma_{d,\mathfrak{u}}$ they are of the same smoothness. Of

course, if all $\gamma_{d,\mathfrak{u}} = 0$ then $H_{d,\alpha,\gamma} = \{0\}$, and if all $\gamma_{d,\mathfrak{u}} = 0$ for $\mathfrak{u} \neq \emptyset$ and $\gamma_{d,\emptyset} > 0$, then $H_{d,\alpha,\gamma}$ is the set of constant functions. So these special weights increase the smoothness of functions.

The weighted Korobov space is a reproducing kernel Hilbert space if $\alpha > 1/2$ and the condition $\alpha > 1/2$ is also necessary if at least one of the weights $\gamma_{d,\mathfrak{u}}$ is non-zero for $\mathfrak{u} \neq \emptyset$. Indeed, for $x, y \in [0, 1]^d$ define

$$K_{d,\alpha,\gamma}(x, y) = \sum_{h \in \mathbb{Z}^d} \varrho_{d,\alpha,\gamma}^{-1}(h) \exp\left(2\pi i h \cdot (x - y)\right). \tag{A.3}$$

Observe that if $\gamma_{d,\mathfrak{u}} = 0$ for all $\mathfrak{u} \neq \emptyset$ then $K_{d,\alpha,\gamma}(x, y) = \gamma_{d,\emptyset} \beta_1^d$ is well defined independently of α. If at least one of $\gamma_{d,\mathfrak{u}} > 0$ for $\mathfrak{u} \neq \emptyset$ then

$$\begin{aligned}
K_{d,\alpha,\gamma}(x, y) &= \sum_{h \in \mathbb{Z}^d} \gamma_{d,\mathfrak{u}_h} \beta_1^{d-|\mathfrak{u}_h|} \beta_2^{|\mathfrak{u}_h|} \prod_{j \in \mathfrak{u}_h} \frac{\exp\left(2\pi h_j(x_j - y_j)\right)}{|h_j|^{2\alpha}} \\
&= \sum_{\mathfrak{u} \subseteq [d]} \sum_{h \in \mathbb{Z}^d, \mathfrak{u}_h = u} \gamma_{d,\mathfrak{u}} \beta_1^{d-|\mathfrak{u}|} \beta_2^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \frac{\exp\left(2\pi i h_j(x_j - y_j)\right)}{|h_j|^{2\alpha}} \\
&= \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \beta_1^{d-|\mathfrak{u}|} (2\beta_2)^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \sum_{h=1}^{\infty} \frac{\cos\left(2\pi h(x_j - y_j)\right)}{h^{2\alpha}}.
\end{aligned}$$

For $x = y$ we obtain

$$K_{d,\alpha,\gamma}(x, x) = \sum_{u \in [d]} \gamma_{d,\mathfrak{u}} \beta_1^{d-|\mathfrak{u}|} (2\beta_2)^{|\mathfrak{u}|} \zeta(2\alpha)^{|\mathfrak{u}|}.$$

Hence, it is well defined iff $\alpha > 1/2$.

As in the previous section, if $\alpha = r$ is a positive integer then

$$K_{d,r,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} (-1)^{(r+1)|\mathfrak{u}|} (2\pi)^{2r|\mathfrak{u}|} \beta_1^{d-|\mathfrak{u}|} \beta_2^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \frac{B_{2r}(\{x_j - y_j\})}{(2r)!}.$$

Proceeding as in the previous section, we can check that $K_{d,\alpha,\gamma}$ is the reproducing kernel of the weighted Korobov space $H_{d,\alpha,\gamma}$. For general weights, $K_{d,\alpha,\gamma}$ cannot be written in the product form. However, for the product weights $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ for non-empty $\mathfrak{u}$, we have

$$K_{d,\alpha,\gamma}(x, y) = \prod_{j=1}^{d} K_{1,\alpha,\gamma_{d,j}}(x_j, y_j)$$

with

$$K_{1,\alpha,\gamma_{d,j}}(x_j, y_j) = \beta_1 + 2\gamma_{d,j}\beta_2 \sum_{h=1}^{\infty} \frac{\cos\left(2\pi(x_j - y_j)\right)}{h^{2\alpha}}.$$

If $\alpha = r$ is a positive integer then

$$K_{d,r,\gamma}(x, y) = \prod_{j=1}^{d} \left(\beta_1 + \gamma_{d,j} \beta_2 (-1)^{r+1} (2\pi)^{2r} \frac{B_{2r}(\{x_j - y_j\})}{(2r)!} \right)$$

with the Bernoulli polynomial B_{2r}.

A.2 Sobolev Spaces

In this section, we restrict ourselves only to Sobolev spaces which are relevant for financial applications, and this corresponds to the study of multivariate integration and low discrepancy points. These spaces are reproducing kernel Hilbert spaces related to tensor product spaces and they consist of functions with relatively low smoothness.

The standard Sobolev spaces are unweighted which corresponds to $\gamma_{d,\mathfrak{u}} = 1$. As we know, polynomial tractability results in many cases can be only obtained for weighted spaces with sufficiently decaying weights. That is why we adopt the standard Sobolev spaces to general weights and obtain weighted Sobolev spaces. The first weighted Sobolev space was introduced in [216], and further variants of weighted Sobolev spaces were given and analyzed in [218].

We consider here three weighted variants of the standard Sobolev spaces of non-periodic and periodic functions defined over the d-dimensional unit cube $[0, 1]^d$ whose first mixed derivatives are square integrable. These variants differ by the choice of a norm. The univariate norms are of the form

$$\|f\| = \left[A^2(f) + \gamma^{-1} \|f'\|^2_{L_2([0,1])} \right]^{1/2}$$

with three different choices of $A(f)$:

$$A(f) = \|f\|_{L_2([0,1])}, \quad A(f) = f(a), \quad A(f) = \int_0^1 f(t)\, \mathrm{d}t,$$

for some positive γ and $a \in [0, 1]$. The case $\gamma = 0$ is also allowed and is obtained by passing with positive γ to 0. For $\gamma = 0$, the three Sobolev spaces considered here consist of only constant functions.

A.2.1 The First Weighted Sobolev Space

For $d = 1$ and $\gamma > 0$, the space $H_{1,\gamma}$ is the Sobolev space of absolutely continuous real functions defined over $[0, 1]$ whose first derivatives belong to $L_2([0, 1])$. The inner product in the space $H_{1,\gamma}$ is defined as

$$\langle f, g \rangle_1 = \int_0^1 f(x)g(x)\, \mathrm{d}x + \gamma^{-1} \int_0^1 f'(x)g'(x)\, \mathrm{d}x \quad \text{for all } f, g \in H_{1,\gamma}.$$

Following Thomas-Agnan [236] and [280], we show that $\{e_k\}_{k \in \mathbb{N}}$ is an orthogonal basis of $H_{1,\gamma}$, where

$$e_k(x) = \cos(\pi(k-1)x) \quad \text{for all } x \in [0,1].$$

Note that e_k does not depend on γ, however, its norm is a function of γ given by

$$\|e_k\|_1 = \left(\delta_{1,k} + (1 - \delta_{1,k})\sqrt{2}\right)\sqrt{1 + \gamma^{-1}[\pi(k-1)]^2},$$

with the Kronecker delta function $\delta_{1,k}$.

Indeed, to show that $\{e_k\}_{k \in \mathbb{N}}$ is an orthogonal basis, it is enough to show that $\langle e_k, e_j \rangle_1 = 0$ for all $k \neq j$, which obviously holds, and that $f \in H_{1,\gamma}$ and $\langle f, e_k \rangle_1 = 0$ for all $k \in \mathbb{N}$ imply that $f = 0$. Observe that $e_k'(0) = e_k'(1) = 0$ for all $k \in \mathbb{N}$, and integration by parts yields

$$\langle f', e_k' \rangle_{L_2} = \int_0^1 e_k'(x) \, df(x) = -\int_0^1 f(x) e_k''(x) \, dx = [\pi(k-1)]^2 \langle f, e_k \rangle_{L_2},$$

with $L_2 = L_2([0,1])$. Hence,

$$\langle f, e_k \rangle_1 = \left(1 + \gamma^{-1}[\pi(k-1)]^2\right) \langle f, e_k \rangle_{L_2}.$$

Then $\langle f, e_k \rangle_1 = 0$ implies that $\langle f, e_k \rangle_{L_2} = 0$ for all $k \in \mathbb{N}$. Since $f \in L_2$ and $\{e_k\}_{k \in \mathbb{N}}$ is an orthonormal basis of L_2, we conclude that $f = 0$, as claimed.

The reproducing kernel K_1 of this space was found by Thomas-Agnan [236] relatively late in 1996, and has the intriguing form

$$K_\gamma(x,y) = \frac{\sqrt{\gamma}}{\sinh \sqrt{\gamma}} \cosh\left(\sqrt{\gamma}(1 - \max(x,y))\right) \cosh\left(\sqrt{\gamma}\min(x,y)\right). \quad \text{(A.4)}$$

The reproducing kernel K_γ can also be written as

$$K_\gamma(x,y) = \sum_{k=1}^{\infty} \frac{e_k(x)}{\|e_k\|_1} \frac{e_k(y)}{\|e_k\|_1} \quad \text{for all } x, y \in [0,1].$$

For $d \geq 2$ and a product weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, with $\gamma_{d,\emptyset} = 1$ and $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$ for non-empty $\mathfrak{u} \subseteq [d]$ and positive $\gamma_{d,j}$, the space $H_{d,\gamma} = H(K_{d,\gamma})$ is the d-fold tensor product of $H_{1,\gamma_{d,j}}$, and is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x,y) = \prod_{j=1}^{d} K_{\gamma_{d,j}}(x_j, y_j) \quad \text{for all } x, y \in [0,1]^d. \quad \text{(A.5)}$$

This is the Sobolev space of d-variate real functions defined over $[0,1]^d$ with the inner product

$$\langle f, g \rangle_{H_{d,\gamma}} = \sum_{\mathfrak{u} \subseteq [d]} \prod_{j \in \mathfrak{u}} \gamma_{d,j}^{-1} \int_{[0,1]^d} \frac{\partial^{|\mathfrak{u}|} f}{\partial x_{\mathfrak{u}}}(x) \frac{\partial^{|\mathfrak{u}|} g}{\partial x_{\mathfrak{u}}}(x) \, dx. \quad \text{(A.6)}$$

As always, $x = [x_1, x_2, \ldots, x_d]$, and $x_{\mathfrak{u}}$ denotes the vector with $|\mathfrak{u}|$ components given by $(x_{\mathfrak{u}})_j = x_j$ for all $j \in \mathfrak{u}$. For $\mathfrak{u} = \emptyset$, the integrand is $f(x)g(x)$.

Due to the tensor product of $H_{d,\gamma}$, the sequence $\{e_k\}_{k \in \mathbb{N}^d}$ is an orthogonal basis of $H_{d,\gamma}$, where

$$e_k(x) = \prod_{j=1}^{d} \cos\left(\pi(k_j - 1)x_j\right) \quad \text{for all } x \in [0,1]^d. \tag{A.7}$$

As for $d = 1$, this sequence does not depend on γ, however, its norm is still a function of γ,

$$\|e_k\|_{H_{d,\gamma}} = \prod_{j=1}^{d} \left(\delta_{1,k_j} + (1 - \delta_{1,k_j})\sqrt{2}\right)\sqrt{1 + \gamma_{d,j}^{-1}\left[\pi(k_j - 1)\right]^2}.$$

The periodic variant of this weighted Sobolev space is when we take the tensor product of univariate periodic functions from $H_{1,\gamma_{d,j}}$ with the additional assumption that $f(0) = f(1)$. Then, see [218], the kernel is

$$K_{d,\gamma}(x, y) = \prod_{j=1}^{d} \left(K_{\gamma_{d,j}}(x_j, y_j) - a_j\left(\sinh(b_j(x_j - 1/2))\sinh(b_j(y_j - 1/2))\right)\right), \tag{A.8}$$

where $a_j = \sqrt{\gamma_{d,j}}/\sinh\sqrt{\gamma_{d,j}}$ and $b_j = \sqrt{\gamma_{d,j}}$.

We now briefly discuss an arbitrary weight sequence $\gamma_{d,\mathfrak{u}}$. Note that the space $H_{d,\gamma}$ can be now defined as a Hilbert space with inner product given by (A.6) with $\prod_{j \in \mathfrak{u}} \gamma_{d,j}$ replaced by a positive $\gamma_{d,\mathfrak{u}}$. That is, the inner product is now

$$\langle f, g \rangle_{H_{d,\gamma}} = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \int_{[0,1]^d} \frac{\partial^{|\mathfrak{u}|} f}{\partial x_{\mathfrak{u}}}(x) \frac{\partial^{|\mathfrak{u}|} g}{\partial x_{\mathfrak{u}}}(x) \, \mathrm{d}x. \tag{A.9}$$

Note that the spaces $H_{d,\gamma}$ for arbitrary positive weights $\gamma_{d,\mathfrak{u}}$ are equivalent, and the sequence $\{e_k\}_{k \in \mathbb{N}^d}$ given by (A.7) is an orthogonal basis for all $H_{d,\gamma}$. However, the norm of e_k depends on γ and is now given by

$$\|e_k\|_{H_{d,\gamma}} = 2^{-|\{j \in [d] | k_j > 1\}|/2}\left(1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2\right)^{1/2}.$$

The space $H_{d,\gamma}$ is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x, y) = \sum_{k \in \mathbb{N}^d} \frac{e_k(x)}{\|e_k\|_{H_{d,\gamma}}} \frac{e_k(y)}{\|e_k\|_{H_{d,\gamma}}} \quad \text{for all } x, y \in [0,1]^d.$$

Hence, for general weights, we do not have a closed form of the reproducing kernel which holds for product weights, see (A.5).

Observe that $\|e_k\|_{L_2([0,1]^d} = 2^{-|\{j \in [d] | k_j > 1\}|/2}$, and therefore we have

$$\frac{\|e_k\|^2_{L_2([0,1]^d)}}{\|e_k\|^2_{H_{d,\gamma}}} = \left(1 + \sum_{\emptyset \neq \mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2 \right)^{-1}. \qquad (A.10)$$

This formula will be useful when we study the multivariate approximation problem for the space $H_{d,\gamma}$.

Finally, we add that it is possible to consider also weights with some $\gamma_{d,\mathfrak{u}} = 0$. In this case we must restrict indices $k \in \mathbb{N}^d$ such that $\prod_{j \in \mathfrak{u}} \left[\pi(k_j - 1)\right]^2 = 0$. That is, $k_j = 1$ for some $j \in \mathfrak{u}$. Then we have the expression $0/0$ which should be interpreted as 0. In terms of functions, the weight $\gamma_{d,\mathfrak{u}} = 0$ means that

$$\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f = 0$$

for all $f \in H_{d,\gamma}$. Observe that this implies that

$$\frac{\partial^{|\mathfrak{v}|}}{\partial x_{\mathfrak{v}}} f = 0 \quad \text{for all } \mathfrak{u} \subseteq \mathfrak{v},$$

and without loss of generality we can assume that also $\gamma_{d,\mathfrak{v}} = 0$ for all sets $\mathfrak{v}$ which contain $\mathfrak{u}$. In this way, we can consider finite-order, or finite-diameter weights also for the space $H_{d,\gamma}$. In particular, if $\gamma_{d,\emptyset} = 1$ and all the rest of the weights are 0, then $H_{d,\gamma}$ consists only of constant functions.

A.2.2 The Second Weighted Sobolev Space

The second weighted Sobolev space $H_{d,\gamma}$ is algebraically the same as the first space but has a different inner product. This space depends on a vector $a \in [0,1]^d$ and sometimes it is called the weighted Sobolev space *anchored* at a, whereas the first weighted Sobolev space is sometimes referred to as the *unanchored* space.

As before, for $d = 1$ and $\gamma > 0$, the space $H_{1,\gamma}$ consists of absolutely continuous real functions defined over $[0, 1]$ with integrable first derivatives. The inner product is now

$$\langle f, g \rangle_1 = f(a)g(a) + \gamma_1^{-1} \int_0^1 f'(x)g'(x)\,\mathrm{d}x \quad \text{for all } f, g \in H_{1,\gamma},$$

where $a \in [0, 1]$. Its reproducing kernel is

$$K_\gamma(x, y) = 1 + \tfrac{1}{2}\gamma\big[|x - a| + |y - a| - |x - y|\big] \quad \text{for all } x, y \in [0, 1]. \quad (A.11)$$

Note that $K_\gamma(x, y) = 0$ for all $x \leq a \leq y$ and $y \leq a \leq x$. This means that K_γ is *decomposable* at a if $0 < a < 1$. As we shall see in Volume II this is a very useful

property which allows us to obtain lower bounds on the worst case errors of arbitrary algorithms for approximation of linear functionals.

The most standard choices of a are 0 and 1. For $a = 0$, we obtain $K_\gamma(x, y) = 1 + \gamma \min(x, y)$, and for $a = 1$, we obtain $K_\gamma(x, y) = 1 + \gamma \min(1 - x, 1 - y)$.

For $d \geq 2$, $a \in [0, 1]^d$ and an arbitrary weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, the space $H_{d,\gamma} = H(K_{d,\gamma})$ is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \frac{\gamma_{d,\mathfrak{u}}}{2^{|\mathfrak{u}|}} \prod_{j \in \mathfrak{u}} \left[|x_j - a_j| + |y_j - a_j| - |x_j - y_j| \right] \qquad (A.12)$$

for all $x, y \in [0, 1]^d$. The inner product of $H_{d,\gamma}$ is

$$\langle f, g \rangle_{H_{d,\gamma}} = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \int_{[0,1]^{|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} f}{\partial x_{\mathfrak{u}}}(x_{\mathfrak{u}}, a) \frac{\partial^{|\mathfrak{u}|} g}{\partial x_{\mathfrak{u}}}(x_{\mathfrak{u}}, a) \, dx_{\mathfrak{u}}. \qquad (A.13)$$

Here, $(x_{\mathfrak{u}}, a)$ denotes the vector of d components such that $(x_{\mathfrak{u}}, a)_j = x_j$ for all $j \in \mathfrak{u}$, and $(x_{\mathfrak{u}}, a)_j = a_j$ for all $j \notin \mathfrak{u}$. Hence, for $\mathfrak{u} = \emptyset$ we have $(x_{\mathfrak{u}}, a) = a$, and for $\mathfrak{u} = [d]$ we have $(x_{\mathfrak{u}}, a) = x$. For $\mathfrak{u} = \emptyset$, the integral is replaced by $f(a)g(a)$.

The space $H_{d,\gamma}$ is used for the study of the L_2-discrepancy anchored at a, see Volume II. For $a = 0$ or $a = 1$, this space is related to the Wiener sheet measure and the average case setting, see also Volume II.

The periodic variant of the space $H_{d,\gamma}$ is obtained as before by assuming that for $d = 1$ we impose the periodicity condition $f(0) = f(1)$. There is a general procedure how to find the reproducing kernel after such periodization, see Section 2.2 of [218]. Applying this procedure, we find out that the kernel for the periodic case is changed to

$$\widetilde{K}_{d,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \frac{\gamma_{d,\mathfrak{u}}}{2^{|\mathfrak{u}|}} \prod_{j \in \mathfrak{u}} \widetilde{K}_{a_j}(x_j, y_j) \quad \text{for all } x, y \in [0, 1]^d, \qquad (A.14)$$

where

$$\widetilde{K}_a(x, y) = |x - a| + |y - a| - |x - y| - 2(x - a)(y - a).$$

For product weights, $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$, the space $H_{d,\gamma}$ is the d-fold tensor product of $H_{1,\gamma_{d,j}}$ and its reproducing kernel has the product form

$$K_{d,\gamma}(x, y) = \prod_{j=1}^{d} \left(1 + \frac{\gamma_{d,j}}{2} \left[|x_j - a_j| + |y_j - a_j| - |x_j - a_j| \right] \right)$$

for all $x, y \in [0, 1]^d$, whereas its periodic counterpart takes the form

$$K_{d,\gamma}(x, y) = \prod_{j=1}^{d} \left(1 + \frac{\gamma_{d,j}}{2} \widetilde{K}_{a_j}(x_j, y_j) \right) \quad \text{for all } x, y \in [0, 1]^d.$$

A.2.3 The Third Weighted Sobolev Space

The third weighted Sobolev space $H_{d,\gamma}$ is algebraically the same as the first and the second space but has a different inner product which does not depend on any parameter. That is why this space, as the first one, is also sometimes called the *unanchored* Sobolev space.

As before, for $d = 1$ and $\gamma > 0$, the space $H_{1,\gamma}$ consists of absolutely continuous real functions defined over $[0, 1]$ with integrable first derivatives. The inner product is now

$$\langle f, g \rangle_1 = \int_0^1 f(x)\,\mathrm{d}x \int_0^1 g(x)\,\mathrm{d}x \; + \; \gamma_1^{-1} \int_0^1 f'(x)g'(x)\mathrm{d}x \quad \text{for all } f, g \in H_{1,\gamma}.$$

Its reproducing kernel is

$$K_\gamma(x, y) = 1 + \tfrac{\gamma}{2}\big[B_2(|x - y|) + 2(x - \tfrac{1}{2})(y - \tfrac{1}{2})\big] \quad \text{for all } x, y \in [0, 1], \quad \text{(A.15)}$$

where B_2 is the Bernoulli polynomial of degree 2, that is, $B_2(x) = x^2 - x + \tfrac{1}{6}$. Note that $B_2(|x - y|) = B_2(\{x - y\})$, where $\{\cdot\}$ denotes the fractional part. That is why we can use these two formulas interchangeably.

For $d \geq 2$ and an arbitrary weight sequence $\gamma = \{\gamma_{d,\mathfrak{u}}\}$, the space $H_{d,\gamma} = H(K_{d,\gamma})$ is a reproducing kernel Hilbert space with the kernel

$$K_{d,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \frac{\gamma_{d,\mathfrak{u}}}{2^{|\mathfrak{u}|}} \prod_{j \in \mathfrak{u}} \big[B_2(|x_j - y_j|) + 2(x_j - \tfrac{1}{2})(y_j - \tfrac{1}{2})\big] \quad \text{(A.16)}$$

for all $x, y \in [0, 1]^d$. The inner product of $H_{d,\gamma}$ is

$$\langle f, g \rangle_d = \sum_{\mathfrak{u} \subseteq \{1,2,\dots,d\}} \prod_{j \in \mathfrak{u}} \gamma_j^{-1} \int_{[0,1]^{|\mathfrak{u}|}} \left(\int_{[0,1]^{d-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} f}{\partial x_{\mathfrak{u}}}(x)\,\mathrm{d}x_{-\mathfrak{u}} \right)$$
$$\times \left(\int_{[0,1]^{d-|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|} g}{\partial x_{\mathfrak{u}}}(x)\,\mathrm{d}x_{-\mathfrak{u}} \right) \mathrm{d}x_{\mathfrak{u}}. \quad \text{(A.17)}$$

Here, $x_{-\mathfrak{u}}$ denotes the vector $x_{[d]-\mathfrak{u}}$ of $d - |\mathfrak{u}|$ components. For $\mathfrak{u} = \emptyset$, the term of the last sum is $\gamma_{d,\emptyset}^{-1} \int_{[0,1]^d} f(x)\,\mathrm{d}x \int_{[0,1]^d} g(x)\,\mathrm{d}x$, whereas for $\mathfrak{u} = [d]$, the corresponding term is $\gamma_{d,[d]}^{-1} \int_{[0,1]^d} \partial^d/\partial x f(x)\,\mathrm{d}x \int_{[0,1]^d} \partial^d/\partial x g(x)\,\mathrm{d}x$.

The periodic variant of the space H_d is obtained as before by assuming that for $d = 1$ we impose the periodicity condition $f(0) = f(1)$. Then the kernel is changed to

$$\tilde{K}_{d,\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \frac{\gamma_{d,\mathfrak{u}}}{2^{|\mathfrak{u}|}} \prod_{j \in \mathfrak{u}} B_2(|x_j - y_j|) \quad \text{for all } x, y \in [0, 1]^d. \quad \text{(A.18)}$$

In this case, it is the same as for the Korobov space of Section with $\alpha = 1$ and $\beta_1 = 1$, $\beta_2 = (2\pi)^{-1}$.

For product weights, $\gamma_{d,\mathfrak{u}} = \prod_{j\in\mathfrak{u}} \gamma_{d,j}$, the space $H_{d,\gamma}$ is the d-fold tensor product of $H_{1,\gamma_{d,j}}$ and its reproducing kernel has the product form

$$K_{d,\gamma}(x, y) = \prod_{j=1}^{d} \left(1 + \frac{\gamma_{d,j}}{2} \left[B_2(|x_j - y_j|) + 2(x_j - \tfrac{1}{2})(y_j - \tfrac{1}{2})\right]\right)$$

for all $x, y \in [0, 1]^d$, whereas its periodic counterpart takes the form

$$\tilde{K}_{d,\gamma}(x, y) = \prod_{j=1}^{d} \left(1 + \frac{\gamma_{d,j}}{2} B_2(|x_j - y_j|)\right) \quad \text{for all } x, y \in [0, 1]^d.$$

Appendix B
Gaussian Measures

Here we list major properties of Gaussian measures that are needed in our book. A more detailed study of Gaussian measures can be found, for instance, in the books of Kuo [118] and Vakhania, Tarieladze and Chobanyan [251].

B.1 Gaussian Measures on Banach Spaces

A probability measure μ defined on Borel sets of $\mathbb{R}$ is called a *Gaussian measure* if it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and has the density

$$\varrho(x) = (2\pi t)^{-1/2} \exp(-(x - \alpha)^2/(2t)),$$

where $\alpha \in \mathbb{R}$ and $t > 0$. The parameters α and t are the *mean* and the *variance* of μ. Hence

$$\alpha = \int_{\mathbb{R}} y\mu(\mathrm{d}y), \quad t = \int_{\mathbb{R}} (y - \alpha)^2\mu(\mathrm{d}y) = \int_{\mathbb{R}} y^2\mu(\mathrm{d}y) - \alpha^2.$$

In this way, mean and variance determine a Gaussian measure on $\mathbb{R}$. For a Borel set A of $\mathbb{R}$, its Gaussian measure is

$$\mu(A) = \frac{1}{\sqrt{2\pi t}} \int_A \exp(-(x - \alpha)^2/(2t))\mathrm{d}x.$$

It is also possible to take $t = 0$ in which case the Gaussian measure becomes a point measure $\mu = \delta_\alpha$, i.e.,

$$\mu(A) = 1 \text{ if } \alpha \in A \quad \text{and} \quad \mu(A) = 0 \text{ if } \alpha \notin A$$

for any Borel set A of $\mathbb{R}$.

Let now $k \in \mathbb{N}$. For finite dimensional spaces $\mathbb{R}^k$, a probability measure μ defined on Borel sets of $\mathbb{R}^k$ is called a *Gaussian measure* if μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^k$ and has the density

$$\varrho_k(x) = (2\pi)^{-k/2}\det(M)^{-1/2} \exp\left(-\tfrac{1}{2}\left\langle M^{-1}(x - \alpha), x - \alpha\right\rangle\right),$$

where $\alpha \in \mathbb{R}^k$ and M is a $k \times k$ symmetric and positive definite matrix, whereas $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product, $\langle x, y \rangle = \sum_{j=1}^{k} x_j y_j$. The vector α is called *mean* and the matrix M is called the *correlation matrix* of μ. Hence

$$\langle \alpha, x \rangle = \int_{\mathbb{R}^k} \langle x, y \rangle \, \mu(\mathrm{d}y) \quad \text{for all } x \in \mathbb{R}^k,$$

and

$$\langle Mx, z \rangle = \int_{\mathbb{R}^k} \langle y - \alpha, x \rangle \langle y - \alpha, z \rangle \, \mu(\mathrm{d}y) \quad \text{for all } x, y \in \mathbb{R}^k.$$

This means that mean and correlation matrix determine a Gaussian measure on $\mathbb{R}^k$. For a Borel set A of $\mathbb{R}^k$, its Gaussian measure is

$$\mu(A) = \frac{1}{(2\pi)^{d/2} \det(M)^{-1/2}} \int_A \exp\left(-\tfrac{1}{2} \langle M^{-1}(x - \alpha), x - \alpha \rangle\right) \, \mathrm{d}x.$$

For a diagonal matrix $M = \mathrm{diag}(t_1, t_2, \ldots, t_k)$, with $t_j > 0$, the last formula simplifies to

$$\mu(A) = \frac{1}{\prod_{j=1}^{k} (2\pi t_j)^{1/2}} \int_A \exp\left(-\tfrac{1}{2} \sum_{j=1}^{k} (x_j - \alpha_j)^2 / t_j\right) \mathrm{d}x.$$

It is also possible to take $t_j = 0$ for some j by limiting process $t_j \to 0$. For example, if all $t_j = 0$, we obtain a point measure $\mu = \delta_\alpha$, and

$$\mu(A) = 1 \text{ if } \alpha \in A \quad \text{and} \quad \mu(A) = 0 \text{ if } \alpha \notin A$$

for any Borel set A of $\mathbb{R}^k$.

For the complex numbers these definitions are modified as follows. A probability measure ν is called *standard Gaussian measure on* $\mathbb{C}$ if it has the density

$$\varrho(z) = \pi^{-1} \exp(-|z|^2).$$

Note that this formally corresponds to the real Gaussian measure on $\mathbb{R}^2$ with $\alpha = 0$ and $M = \mathrm{diag}\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$.

A measure μ on $\mathbb{C}$ is called *Gaussian* if there exist $\alpha \in \mathbb{C}$ and $t \geq 0$ such that μ is the image of ν under the mapping $z \mapsto \alpha + \sqrt{t}z$. Then, again, the parameter α is the mean and we obtain

$$\alpha = \int_{\mathbb{C}} z \mu(\mathrm{d}z), \quad t = \int_{\mathbb{C}} |z|^2 \mu(\mathrm{d}z) - |\alpha|^2.$$

We now define a Gaussian measure μ on a separable Banach space F. If $\dim(F) = \infty$ then Gaussian measures cannot be defined by means of densities. Instead a different approach is used. Gaussian measures, as well as arbitrary probability measures, can be uniquely defined through their *characteristic functionals*. The characteristic functional ψ_μ of a probability measure μ defined on Borel sets of F is given by

$$\psi_\mu(L) = \int_F \exp\left(\mathrm{i}\,\mathrm{Re}\left(L(f)\right)\right) \mu(\mathrm{d}f)$$

for all $L \in F^*$, where $\mathrm{i} = \sqrt{-1}$. The Re stands for the real part and, of course, this is not needed in the case of a Banach space over the reals.

A measure μ defined on Borel sets of F is Gaussian iff the 1-dimensional images μL^{-1} are Gaussian for all $L \in F^*$. That is, for any Borel set of $\mathbb{R}$, we have

$$\mu L^{-1}(A) = \mu(\{f \in F \mid L(f) \in A\}) = \frac{1}{\sqrt{2\pi t_L}} \int_A \exp(-(x - \alpha_L)^2/(2t_L)) \, dx$$

for some α_L and t_L depending on L. In the complex case, this formula is modified by

$$\mu L^{-1}(A) = \mu(\{f \in F \mid L(f) \in A\}) = \frac{1}{\pi t_L} \int_A \exp(-|z - \alpha_L|^2/t_L) \, dz.$$

Equivalently, a measure μ is Gaussian iff its characteristic functional is of the form

$$\psi_\mu(L) = \exp(iL(m_\mu) - \tfrac{1}{2}L(C_\mu L))$$

for all $L \in F^*$ for some $m_\mu \in F$ and a linear operator $C_\mu : F^* \to F$. The element m_μ is called the *mean* element of μ and is uniquely defined by the condition

$$L(m_\mu) = \int_F L(f)\mu(df) \quad \text{for all } L \in F^*.$$

The operator C_μ is called the *correlation operator* of μ and is uniquely defined by the condition

$$L_1(C_\mu L_2) = \int_F \overline{L_1(f - m_\mu)} L_2(f - m_\mu)\mu(df) \quad \text{for all } L_1, L_2 \in F^*.$$

The correlation operator is linear, symmetric in the sense that

$$L_1(C_\mu L_2) = \overline{L_2(C_\mu L_1)},$$

and non-negative definite, i.e., $L(C_\mu L) \geq 0$ for all $L_1, L_2, L \in F^*$. If $m_\mu = 0$ then C_μ is also called the *covariance* operator of μ.

For general separable Banach spaces, the complete characterization of the correlation operators is not known. If, however, F is a separable Hilbert space then the correlation operators are fully characterized by being symmetric, nonnegative definite, and having a *finite trace*. That is, C_μ is a correlation operator of a Gaussian measure μ on F iff $C_\mu = C_\mu^* \geq 0$ and

$$\text{trace}(C_\mu) = \int_F \|f\|^2 \mu(df) = \sum_{j=1}^{\dim(F)} \langle C_\mu \eta_j, \eta_j \rangle < \infty,$$

where $\{\eta_j\}$ is a complete orthonormal system of F. In particular, one can choose $\{\eta_j\}$ as eigenelements of C_μ, $C_\mu \eta_j = \lambda_j \eta_j$ for $j = 1, 2, \ldots$. Thus, the trace of C_μ is the sum of the eigenvalues of C_μ which is finite,

$$\text{trace}(C_\mu) = \sum_{j=1}^{\dim(F)} \lambda_j < \infty.$$

Without loss of generality we can assume that the eigenvalues λ_j are non-increasing, $\lambda_j \geq \lambda_{j+1}$ for all $j \in \mathbb{N}$. The eigenvalue λ_j can be interpreted as the importance of the direction η_j of the space F. For a finite dimensional space F, one can take all λ_j equal, i.e., $\lambda_j = \lambda > 0$ for $j = 1, 2, \ldots, \dim(F)$. For an infinite dimensional space F, we must have decaying eigenvalues. Thus all directions of the space F can be equally important only if $\dim(F) < \infty$. This property is sometimes described that there is no fair Gaussian measure in infinite dimensional spaces.

We mention a result concerning the measure of a ball $B_q = \{x \in X \mid \|x\| \leq q\}$ with respect to a zero-mean Gaussian measure on the separable Hilbert space X, see [242] for a proof. For all $q > 0$,

$$\mu(B_q) \geq 1 - 5 \exp\left(\frac{-q^2}{2 \operatorname{trace}(C_\mu)}\right).$$

Hence, if q^2 is large relative to the trace of the correlation operator then the Gaussian measure of the ball B_q is close to one. This property will be used in the average and probabilistic settings.

In this book we mainly use Gaussian measures on real Banach spaces, however only minor changes are needed to consider the complex case. We provide an example with the *Wiener measure*. For the real case, we consider the space $C([0,1])$ of real continuous functions with the sup norm, $\|f\| = \sup_{x \in [0,1]} |f(x)|$. Then the Wiener measure w is Gaussian with the zero-mean and with the covariance kernel,

$$K(x,y) := \int_{C([0,1])} f(x) f(y) w(\mathrm{d}f) \quad \text{for all } x, y \in [0,1],$$

given by $K(x,y) = \min(x,y)$. Note that $f(0) = 0$ with probability 1 for the Wiener measure.

The complex Wiener space is a triple $(C^{\mathbb{C}}([0,1]), \mathcal{S}', \mu_{\mathbb{C}})$. Here $C^{\mathbb{C}}([0,1])$ is the space of complex valued continuous functions $f : [0,1] \to \mathbb{C}$ with the norm $\|f\| = \sup_{x \in [0,1]} |f(x)|$. By $\mathcal{B}$ we denote the Borel σ-algebra induced by the norm. The canonical process is

$$Z_t(\omega) = X_t(\omega) + \mathrm{i}Y_t(\omega) = \omega(t), \quad \omega \in C^{\mathbb{C}}([0,1]).$$

The probability measure $\mu_{\mathbb{C}}$ is the unique probability measure on $\mathcal{B}$ such that

- X and Y are independent real processes,

- $X_0 = Y_0 = 0$ with probability 1,

- X and Y have independent zero-mean Gaussian increments,

- $E((X_t - X_s)^2) = E((Y_t - Y_s)^2) = t - s$ for $0 \leq s \leq t \leq 1$.

B.2 Gaussian Measures and Reproducing Kernel Hilbert Spaces

Let F be a separable Banach space of real functions defined on D which is a subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}$. We assume that for each $x \in D$, the linear functional L_x is continuous, where $L_x(f) = f(x)$ for all $f \in F$.

We equip the space F with a zero-mean Gaussian measure with correlation operator $C_\mu : F^* \to B$. Then

$$K_\mu(x, y) = L_x(C_\mu L_y) = \int_F f(x) f(y) \mu(\mathrm{d}f) \quad \text{for all } x, y \in D,$$

is called the *covariance kernel* of μ.

The covariance kernel K_μ has the same properties as a reproducing kernel of a Hilbert space. Indeed, it is symmetric, $K_\mu(x, y) = K_\mu(y, x)$, and semi-positive definite, i.e., the matrix $(K_\mu(x_i, x_j))_{i,j=1,\dots,m}$ is symmetric and semi-positive definite for all m and $x_i, x_j \in D$. This simply follows from the fact that for arbitrary real a_j we have

$$\sum_{i,j=1}^{m} a_i a_j K_\mu(x_i, x_j) = \int_F \left(\sum_{j=1}^{m} a_j f(x_j) \right)^2 \mu(\mathrm{d}f) \geq 0.$$

Let

$$\tilde{H} = C_\mu(F^*) = \{ C_\mu L \mid L \in F^* \}$$

be a subset of F consisting of functions $g_L = C_\mu L$ which we can obtain by varying $L \in F^*$. Note that

$$g_L(x) = L_x(C_\mu L) = \int_F L_x(f) L(f) \mu(\mathrm{d}f) = \int_F f(x) L(f) \mu(\mathrm{d}f) \quad \text{for all } x \in D.$$

Clearly, $\tilde{H}$ is a linear subspace of F. We equip $\tilde{H}$ with the inner product

$$\langle g_{L_1}, g_{L_2} \rangle_{\tilde{H}} = L_1(C_\mu L_2) = \int_F L_1(f) L_2(f) \mu(\mathrm{d}f) \quad \text{for all } L_1, L_2 \in F^*.$$

This defines a semi-inner product. To prove that this is indeed an inner product assume that there exists $L \in F^*$ such that $\langle g_L, g_L \rangle_{\tilde{H}} = L(C_\mu L) = 0$. Since

$$|L_1(C_\mu L)| \leq (L_1(C_\mu L_1))^{1/2} (L(C_\mu L)^{1/2} = 0$$

then $L_1(C_\mu L) = 0$ for all $L_1 \in F^*$. In particular, $L_x(C_\mu L) = (C_\mu L)(x) = 0$ for all $x \in D$, and $g_L = C_\mu L = 0$, as needed.

Note that for $g_{L_x} = C_\mu L_x$ we have

$$g_{L_x}(y) = L_y(C_\mu L_x) = K_\mu(y, x) \quad \text{for all } y \in D.$$

Hence, $g_{L_x} = K_\mu(\cdot, x) \in \tilde{H}$ for all $x \in D$. Furthermore,

$$\langle g_L, K_\mu(\cdot, x) \rangle_{\tilde{H}} = L(C_\mu L_x) = L_x(C_\mu L) = (C_\mu L)(x) = g_L(x)$$

and this holds for all $g_L \in \widetilde{H}$ and all $x \in D$. This proves that K_μ is a reproducing kernel of $\widetilde{H}$.

The space $\widetilde{H}$ is, in general, not a Hilbert space. We define $H = H(K_\mu)$ as the completion of $\widetilde{H}$, and obtain a reproducing kernel Hilbert space with the reproducing kernel K_μ. Its inner product is denoted by $\langle \cdot, \cdot \rangle_H$.

In this way, a Banach space F and a zero-mean Gaussian measure yield a reproducing kernel Hilbert space $H(K_\mu)$. We illustrate the construction of the space $H(K_\mu)$ for the class of continuous functions $F = C([0, 1]^d)$ equipped with the Wiener sheet measure $\mu = w$ which is a zero-mean Gaussian measure with the covariance function

$$K_w(x, y) = \prod_{j=1}^d \min(x_j, y_j) \quad \text{for all } x, y \in [0, 1]^d.$$

Note that just now $f(x) = 0$ with probability 1 if x has at least one component equal to 0.

The space $H(K_w)$ is the Sobolev space which is the tensor product of d copies of the space of univariate functions that are absolutely continuous, vanish at 0, and whose first derivatives are in $L_2([0, 1])$. The inner product in $H(K_w)$ is of the form

$$\langle g_1, g_2 \rangle_{H(K_w)} = \int_{[0,1]^d} \frac{\partial^d}{\partial x_1 \partial x_2 \ldots \partial x_d} g_1(x) \frac{\partial^d}{\partial x_1 \partial x_2 \ldots \partial x_d} g_2(x) \, dx,$$

and functions g in $H(K_w)$ satisfy the boundary conditions $g(x) = 0$ if x has at least one component equal to 0.

It is also possible to remove the boundary conditions if we equip the space $F = C([0, 1]^d)$ with the weighted Wiener measure $\mu = w_\gamma$. This is a zero-mean Gaussian measure with the covariance kernel

$$K_{w_\gamma}(x, y) = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}} \prod_{j \in \mathfrak{u}} \min(x_j, y_j) \quad \text{for all } x, y \in [0, 1]^d.$$

Then the space $H(K_{w_\omega})$ is a Sobolev space with the inner product

$$\langle g_1, g_2 \rangle_{H(K_{w_\omega})} = \sum_{\mathfrak{u} \subseteq [d]} \frac{1}{\gamma_{d,\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \frac{\partial^{|\mathfrak{u}|}}{\partial x_\mathfrak{u}} g_1(x_\mathfrak{u}, 0) \frac{\partial^{|\mathfrak{u}|}}{\partial x_\mathfrak{u}} g_2(x_\mathfrak{u}, 0) \, dx.$$

As we shall see in Volume II, approximation of a continuous linear functional L in the average case setting for the class Λ^{std} and for the space F equipped with the zero-mean Gaussian measure μ is equivalent to approximation of a continuous linear functional $\langle \cdot, g_L \rangle_H$ in the worst case setting for the same class Λ^{std} and for the unit ball of $H(K_\mu)$. In particular, the problem of multivariate integration in the average case setting for the space $C([0, 1]^d)$ and the Gaussian measure w or w_γ is equivalent to the problem of multivariate integration in the worst case setting for the space Sobolev space $H(K_w)$ or $H(K_{w_\gamma})$, respectively.

Appendix C
List of Open Problems

1. Integration and approximation for the classes $F_{d,r}$, Section 3.3.

2. Integration and approximation for the classes $F_{d,r(d)}$ and $F_{d,\infty}$, Section 3.3.

3. Integration for a finite dimensional space F_d of trigonometric polynomials, Section 3.3.

4. Integration for weighted Korobov spaces, Section 3.3.

5. Approximation of C^∞-functions from the classes $F_{d,p}$, Section 3.3.

6. Construction of points with small star-discrepancy, Section 3.3.

7. On bounds for the star-discrepancy, Section 3.3.

8. Diagonal problems for C^r-functions from the class $F_{d,\gamma,r}$, Section 3.3.

9. Construction of good points for Gaussian integration for the isotropic Wiener measure, Section 3.3.

10. Tractability for approximation with folded Wiener sheet measures with increasing smoothness, Section 3.3.

11. Tractability for approximation with folded Wiener sheet measures with varying smoothness, Section 3.3.

12. Tractability for a modified error criterion, Section 3.3.

13. Tractability in the randomized setting for integration over weighted Sobolev spaces, Section 3.3.

14. Tractability in the randomized setting for integration over periodic weighted Sobolev spaces, Section 3.3.

15. Tractability in the randomized setting for general linear operators, Section 3.3.

16. On the power of adaption for linear problems, Section 4.2.1.

17. On the power of adaption for linear operators on convex sets, Section 4.2.1.

18. On the asymptotic optimality of linear algorithms for Sobolev embeddings for Λ^{std}, Section 4.2.4.

19. On the existence of optimal measurable algorithms, Section 4.3.3.

20. On the power of adaption for linear problems in the randomized setting, Section 4.3.3.

21. On the (almost) optimality of linear algorithms for linear problems in the randomized setting, Section 4.3.3.

22. How good are linear randomized algorithms for linear problems?
    Section 4.3.3.

23. How good are linear randomized algorithms for linear problems defined over Hilbert spaces? Section 4.3.3.

24. On the optimality of measurable algorithms in the randomized setting, Section 4.3.3.

25. On Sobolev embeddings in the randomized setting, Section 4.3.3.

26. Weak tractability of linear tensor product problems in the worst case setting with $\lambda_1 = 1$ and $\lambda_2 < 1$, Section 5.2.

27. Tractability of linear weighted tensor product problems for the absolute error criterion, Section 5.3.4.

28. Weak tractability for linear tensor product problems in the average case setting, Section 6.2.

29. Tractability of linear weighted product problems in the average case setting for the absolute error criterion, Section 6.3.

30. Weak tractability for linear weighted tensor product problems in the average case setting, Section 6.3.

Bibliography

The numbers at the end of each item refer to the pages on which the respective work is cited.

[1] P. Atkins and R. Friedman, *Molecular quantum mechanics*, Oxford University Press, Oxford, 1997. 3

[2] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68**, 337–404, 1950. 57, 193

[3] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, Princeton University, draft of a book, available on the internet. 150

[4] V. F. Babenko and S. V. Borodachov, On optimization of approximate integration over a d-dimensional ball, *East J. on Approx.* **9**, 95–109, 2003. 94

[5] V. F. Babenko and S. V. Borodachov, On the construction of optimal cubature formulae which use integrals over hyperspheres, *J. Complexity* **23**, 346–358, 2007. 94

[6] I. Babuška and W. C. Rheinboldt, Error estimates for adaptive finite element computations, *SIAM J. Numer. Anal.* **15**, 736–754, 1978. 113

[7] N. S. Bakhvalov, On approximate computation of integrals, *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* **4**, 3–18, 1959, in Russian. 15, 24, 71, 75, 127, 140, 142, 153

[8] N. S. Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions. *USSR Comput. Maths. Math. Phys.* **11**, 244–249, 1971. 111, 115

[9] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton NJ, 1957. 4

[10] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Boston, 2004. 57

[11] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968. 115

[12] D. Bini and V. Pan, *Polynomial and Matrix Computation*, Birkhäuser Verlag, Basel, 1994. 4, 99

[13] L. Blum, M. Shub and S. Smale, On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines, *Bull. Amer. Math. Soc.* **21**, 1–46, 1989. 4, 98, 100

[14] L. Blum, F. Cucker, M. Shub and S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, 1998. 4, 5, 98

[15] C. G. E. Boender and H. E. Romeijn, Stochastic methods, in: *Handbook of Global Optimization*, R. Horst, P. M. Pardalos, (eds.), 829–869, Kluwer, Dordrecht, 1995. 3

[16] B. Bojanov, Optimal quadrature formulas, *Russian Math. Surveys* **60**, 1035–1055, 2005. 94

[17] J. M. Borwein and P. B. Borwein, *Pi and the AGM: a study in analytic number theory and computational complexity*, Wiley-Interscience, New York, 1987. 100

[18] G. E. P. Box and M. E. Muller, A note on the generation of random normal deviates, *Ann. Math. Statist.*, **29**, 610–611, 1958. 13, 56

[19] R. P. Brent, Multiple-precision zero-finding methods and the complexity of elementary function evaluation, in: *Analytic Computational Complexity*, J. F. Traub (ed.), 151–176, Academic Press, New York, 1975. 100

[20] R. P. Brent, Fast multiple-precision evaluation of elementary functions, *J. ACM* **23**, 242–251, 1976. 100

[21] C. Le Bris, Computational chemistry from the perspective of numerical analysis, *Acta Numerica* **14**, 363–444, 2005.

[22] H. Bückner, Bemerkungen zur numerischen Quadratur, I, II, *Math. Nachr.* **3**, 142–145, 146–151, 1950. 152

[23] P. Bürgisser, M. Clausen and M. A. Shokrollahi, *Algebraic Complexity Theory*, Springer Verlag, Berlin, 1997. 4, 99

[24] A. M. Caetano, About approximation numbers in function spaces, *J. Approx. Theory* **94**, 383–395. 123

[25] R. E. Caflisch, W. Morokoff and A. Owen, Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comp. Finance* **1**, 27–46, 1997. 43

[26] J. M. Calvin, A lower bound on complexity of optimization on the Wiener space, *Theoret. Comput. Sci.* **383**, 132–139, 2007. 132

[27] E. J. Candes and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory* **52**, 5406–5425, 2006. 117

[28] Ming Chu, There exists a problem whose computational complexity is any given function of the information complexity, *J. Complexity* **10**, 445–450, 1994. 108

[29] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978. 126

[30] Z. Ciesielski, On Lévy's Brownian motion with several-dimensional time, *Lecture Notes in Mathematics* **472**, 29–56, Springer-Verlag, New York, 1975. 51, 57

[31] A. Cohen, W. Dahmen and R. DeVore, Adaptive wavelet methods II—beyond the elliptic case, *Found. Comput. Math.* **2**, 203–245, 2002. 96, 113

[32] D. Coppersmith and S. Winograd, Matrix multiplication via arithmetic progression, *J. Symbolic Comput.*, 9, 251–280, 1990. 4, 99

[33] J. Creutzig, Finite-diameter weights, private communications, 2007. 197, 240

[34] J. Creutzig and P. Wojtaszczyk, Linear vs. nonlinear algorithms for linear problems, *J. Complexity* **20**, 807–820, 2004. 112, 116, 117, 152, 153

[35] S. Dahlke, W. Dahmen and R. A. DeVore, Nonlinear approximation and adaptive techniques for solving elliptic operator equations, In: *Multiscale Wavelet Methods for PDEs*, W. Dahmen, A. J. Kurdila, P. Oswald (eds.), 237–283, Academic Press, San Diego, 1997. 96, 113

[36] S. Dahlke, E. Novak and W. Sickel, Optimal approximation of elliptic problems by linear and nonlinear mappings I. *J. Complexity* **22**, 29–49, 2006. 96, 126

[37] S. Dahlke, E. Novak and W. Sickel, Optimal approximation of elliptic problems by linear and nonlinear mappings II. *J. Complexity* **22**, 549–603, 2006. 96, 110, 113, 126

[38] S. Dahlke, E. Novak and W. Sickel, Optimal approximation of elliptic problems by linear and nonlinear mappings III. *J. Complexity* **23**, 614–648, 2007. 96, 126

[39] R. A. DeVore, Nonlinear approximation, *Acta Numerica* **7**, 51–150, 1998. 95, 96

[40] R. A. DeVore, R. Howard and C. Micchelli, Optimal nonlinear approximation, *Manuscripta Math.* **63**, 469–478, 1989. 96

[41] R. A. DeVore, G. Kyriazis, D. Leviatan and V. M. Tikhomirov, Wavelet compression and nonlinear n-widths. *Adv. Comput. Math.* **1**, 197–214, 1993. 96

[42] J. Dick, On the convergence order of the component-by-component construction of good lattice rules, *J. Complexity* **20**, 493–522, 2004. 13, 78

[43] J. Dick and F. Y. Kuo, Reducing the construction cost of the component-by-component construction of good lattice rules, *Math. Comp.* **73**, 1967–1988, 2004. 13, 78

[44] J. Dick and F. Y. Kuo, Constructing good lattice rules with millions of points, in *Monte Carlo and Quasi-Monte Carlo Methods 2002*, H. Niederreiter, ed., Springer-Verlag, 181–197, 2004. 13, 78

[45] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights, *Numer. Math.* **103**, 63–97. 2006. 9, 78, 195, 196, 342

[46] B. Doerr and M. Gnewuch, Construction of low-discrepancy point sets of small size by bracketing covers and dependent randomized rounding, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 299–312, Springer, Berlin, 2008. 79

[47] B. Doerr, M. Gnewuch, P. Kritzer and F. Pillichshammer, Component-by-component construction of low-discrepancy point sets of small size, *Monte Carlo Meth. Appl.* **14** (2008), 129–149. 79

[48] B. Doerr, M. Gnewuch and A. Srivastav, Bounds and constructions for the star-discrepancy via δ-covers, *J. Complexity* **21**, 691–709, 2005. 39, 79

[49] D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* **52**, 1289–1306, 2006. 117

[50] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, LNiM **1651**, Springer-Verlag, Berlin, 1997. 35

[51] R. M. Dudley, A course on empirical processes (École d'Été de Probabilités de Saint-Flour XII-1982), LNiM 1097, Springer-Verlag, New York, 2–141, 1984. 39

[52] D. Dung and V. Q. Thanh (1996): On nonlinear n-widths. *Proc. of the AMS* **124**, 2757–2765. 96

[53] D. Dung (2000): Continuous algorithms in n-term approximation and non-linear widths. *J. Approx. Theory* **102**, 217–242. 96

[54] D. E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers and Differential Operators*, Cambridge University Press, Cambridge, 1996. 123

[55] B. Efron and C. Stein, The Jackknife estimate of variance, *Ann. Statist.* **9**, 586–596, 1981. 41

[56] K. V. Emelyanov and A. M. Ilin, On the number of arithmetic operations, necessary for the approximate solution of Fredholm integral equations of the second kind, Zh. Vychisl. Mat. i Mat. Fiz. **7**, 905–910, 1967. 153

368   Bibliography

[57] K. Eriksson, D. Estep, P. Hansbo and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica 1995, Cambridge University Press, pp. 105–158. 113

[58] G. Fang and L. Qian, Approximation characteristics for diagonal operators in different computational settings, *J. Approx. Theory* **140**, 178–190, 2006. 147

[59] H.-J. Flad, W. Hackbusch, R. Schneider, Best N-term approximation in electronic structure calculations II, Jastrow factors, *Math. Model. Numer. Anal.* **41**, 261–279, 2007. 3

[60] S. Gal and C. A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* **10**, 105–120, 1980. 111

[61] W. Gao, P. Ye and H. Wang, Optimal error bound of restricted Monte Carlo integration on isotropic Sobolev classes, *Progr. Natur. Sci.* (English Ed.) **16**, 588–593, 2006. 141

[62] Th. Gerstner and M. Griebel, Numerical integration using sparse grids, *Num. Algorithms* **18**, 209–232, 1998. 3

[63] J. Gilewicz, V. N. Konovalov and D. Leviatan, Widths and shape-preserving widths of Sobolev-type classes of s-monotone functions, *J. Approx. Theory* **140**, 101–126, 2006. 114

[64] J. Glimm and A. Jaffe, *Quantum Physics*, Springer Verlag, New York, 1987. 10, 196

[65] E. D. Gluskin, Norms of random matrices and diameters of finite-dimensional sets, *Mat. Sb. (N.S.)* **120 (162)**, 180–189, 286, 1983. 124

[66] I. S. Gradshtein and I. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New York, 1980. 53

[67] M. Gnewuch, Bounds for the average L_p-extreme and the L^∞-extreme discrepancy, *Electronic J. Combinatorics* **12**, Research Paper 54, 2005. 39

[68] M. Gnewuch, Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy, *J. Complexity* **24**, 154–172, 2008. 39

[69] M. Gnewuch and H. Woźniakowski, Generalized tractability for multivariate problems, Part I: Linear tensor product problems and linear information, *J. Complexity* **23**, 262–295, 2007. 5, 154, 286, 340

[70] M. Gnewuch and H. Woźniakowski, Generalized tractability for multivariate problems, Part II: Linear tensor product problems, linear information, unrestricted tractability, Technical Report, Department of Computer Science, Columbia University, New York 2007, Report cucs-042-07. 5, 7, 286, 340

[71] M. Gnewuch and H. Woźniakowski, Generalized tractability for linear functionals, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 359–381, Springer, Berlin, 2008. 5, 7

[72] M. Griebel, Sparse grids and related approximation schemes for higher dimensional problems, in: *Foundations of Computational Mathematics*, Santander 2005, L. M. Pardo et al. (eds.), 106–161, London Math. Soc., Lecture Note Series 331, Cambridge University Press, 2006. 8

[73] A. Griewank and Ph. L. Toint, Local convergence analysis for partitioned quasi-Newton updates, *Numer. Math.* **39**, 429–448, 1982. 10

[74] J. M. Hammersley, Monte Carlo methods for solving multivariate problems, *Ann. New York Acad. Sci.* **86**, 844–874, 1960. 40

[75] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* **2**, 84–90, 1960. 40

[76] D. Haroske and H. Triebel, Entropy numbers in weighted function spaces and eigenvalue distributions of some degenerate pseudodifferential operators I, *Math. Nachr.* **167**, 131–156, 1994. 121

[77] D. Haroske and H. Triebel, *Distributions, Sobolev Spaces, Elliptic Equations*, European Mathematical Society, 2008. 119

[78] D. Haussler, Sphere packing numbers for subsets of the Boolean *n*-cube with bounded Vapnik-Červonenkis dimension, *J. Combinatorial Theory* **A 69**, 217–232, 1995. 39

[79] S. Heinrich, Lower bounds for the complexity of Monte Carlo function approximation, *J. Complexity* **8**, 277–300, 1992. 74, 85, 140, 145, 146, 285

[80] S. Heinrich, Random approximation in numerical analysis, in: Functional Analysis, K. D. Bierstedt et al. (eds.), 123–171, Dekker, New York, 1994. 71, 75, 126, 127, 140, 153

[81] S. Heinrich, Complexity of Monte Carlo algorithms, in *The Mathematics of Numerical Analysis*, Lectures in Applied Mathematics **32**, pp. 405–419, AMS-SIAM Summer Seminar, Park City, American Math. Society, Providence, RI, 1996. 127, 140, 153

[82] S. Heinrich, Monte Carlo complexity of global solution of integral equations, *J. Complexity* **14**, 151–175, 1998. 153

[83] S. Heinrich, Some open problems concerning the star-discrepancy, *J. Complexity* **19**, 416–419, 2003. 79

[84] S. Heinrich, Monte Carlo approximation of weakly singular integral operators, *J. Complexity* **22**, 192–219, 2006. 153

[85] S. Heinrich, The randomized information complexity of elliptic PDE, *J. Complexity* **22**, 220–249, 2006. 153

[86] S. Heinrich, Randomized approximation of Sobolev embeddings, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 445–459, Springer, Berlin. 123, 146

[87] S. Heinrich, Randomized approximation of Sobolev embeddings II, in preparation. 123, 147

[88] S. Heinrich, Randomized approximation of Sobolev embeddings III, in preparation. 147

[89] S. Heinrich and P. Mathé, The Monte Carlo complexity of Fredholm integral equations, *Math. Comp.* **60**, 257–278, 1993. 153

[90] S. Heinrich, E. Novak and H. Pfeiffer, How many random bits do we need for Monte Carlo integration? In *Monte Carlo and Quasi-Monte Carlo Methods 2002*, H. Niederreiter (ed.), Springer 2004, pp. 27–49. 141

[91] S. Heinrich, E. Novak, G. W. Wasilkowski and H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, *Acta Arithmetica* **96**, 279–302, 2001. 39, 40

[92] P. Hertling, Topological complexity with continuous operations, *J. Complexity* **12**, 315–338, 1996. 98, 107

[93] F. J. Hickernell, G. W. Wasilkowski and H. Woźniakowski, Tractability of linear multi-variate problems in the average case setting, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 461–494, Springer, Berlin, 2008. 174, 239, 282

[94] F. J. Hickernell and H. Woźniakowski, Integration and approximation in arbitrary dimension, *Adv. Comput. Math.*, **12**, 25–58, 2000. 56, 86, 194, 239, 282

[95] F. J. Hickernell and H. Woźniakowski, Tractability of multivariate integration for periodic functions, *J. Complexity* **17**, 660–682, 2001. 78, 342

[96] N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996. 101

[97] A. Hinrichs, Covering numbers, Vapnik-Cervonenkis classes and bounds for the star discrepancy, *J. Complexity* **20**, 477–483, 2004. 40

[98] E. Hlawka, Über die Diskrepanz mehrdimensionaler Folgen mod 1, *Math. Z.* **77**, 273–284, 1961. 36

[99] Tak-San Ho and H. Rabitz, Reproducing kernel Hilbert space interpolation methods as a paradigm of high-dimensional model representations: application to multidimensional potential energy surface construction, *J. Chem. Physics* **119**, 6433–6442, 2003. 8, 41

[100] F. L. Huang and S. Zhang, Approximation of infinitely differentiable multivariate functions is not strongly tractable, *J. Complexity* **23**, 73–81, 2007. 77

[101] I. P. Huerta, Adaption helps for some nonconvex classes, *J. Complexity* **4**, 333–352, 1986. 115

[102] T. Jackowski and H. Woźniakowski, Complexity of approximation with relative error criterion in worst, average, and probabilistic settings, *J. Complexity* **3**, 114–135, 1987. 66

[103] J. W. Jaromczyk and G. W. Wasilkowski, Numerical stability of a convex hull algorithm for simple polygons, *Algorithmica* **10**, 457–472, 1993. 99

[104] J. W. Jaromczyk and G. W. Wasilkowski, Computing convex hull in floating point arithmetic, *Computational Geometry: Theory and Applications* **4**, 283–292, 1994. 99

[105] B. Z. Kacewicz, Optimal solution of ordinary differential equations, *J. Complexity* **3**. 451-465. 1987. 3

[106] B. S. Kashin, Widths of certain finite-dimensional sets and classes of smooth functions, *Izv. Math.* **41**, 334–351, 1977. 124

[107] B. S. Kashin, Widths of Sobolev classes of small-order smoothness, *Vestnik Mosc. Univ. ser. mat. mekh. astron.* **5**, 50–54, 1981. 124

[108] Ker-I Ko, *Complexity Theory of Real Functions*, Birkhäuser, Boston, 1991. 105

[109] J. Kiefer, Sequential minimax search for a maximum, *Proc. Amer. Math. Soc.* **4**, 502–505, 1953 152

[110] J. Kiefer, Optimum sequential search and approximation methods under minimum regularity assumptions, *J. Soc. Indust. Appl. Math.* **5**, 105–136, 1957. 114

[111] A. Kiełbasiński, Basic concepts in numerical error analysis (suggested by linear algebra problems), in Mathematical models and numerical methods, *Banach Center Publications* **3**, PWN, Warsaw, 223–232, 1978. 101, 102

[112] A. N. Kolmogorov, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, *Ann. of Math.* **37**, 107–110, 1936. 153

[113] M. A. Kon and E. Novak, The adaption problem for approximating linear operators, *Bull. Amer. Math. Soc.* **23**, 159–165, 1990. 111

[114] N. P. Korneichuk, Optimization of active algorithms for recovery of monotonic functions from Hölder's class, *J. Complexity* **10**, 265–269, 1994. 114

[115] M. Kowalski, K. A. Sikorski and F. Stenger, *Selected Topics in Approximation and Computation*, Oxford University Press, New York, 1995. 95

[116] P. Kritzer and F. Pillichshammer, Constructions of general polynomial lattices for multivariate integration, *Bull. Austral. Math. Soc.* **76**, 93–110, 2007. 79

[117] S. N. Kudryavtsev. The best accuracy of reconstruction of finitely smooth functions from their values at a given number of points. *Izv. Math.* **62**(1), 19–53, 1998. 126

[118] Hui-Hsiung Kuo, *Gaussian Measures in Banach Spaces*, Lecture Notes in Mathematics **463**, Springer-Verlag, Berlin, 1975. 49, 357

[119] F. Y. Kuo, Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces, *J. Complexity* **19**, 301–320, 2003. 13, 78

[120] F. Y. Kuo and I. H. Sloan, Quasi-Monte Carlo methods can be efficient for integration over product spheres, *J. Complexity* **21**, 196–210, 2005. 239

[121] 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*, H. Niederreiter and D. Talay (eds.), 289–330, Springer, Berlin, 2006. 13

[122] F. Y. Kuo, I. H. Sloan and H. Woźniakowski, Lattice rule algorithms for multivariate approximation in the average case setting, *J. Complexity* **24**, 283-323, 2008. 13, 283

[123] F. Y. Kuo, I. H. Sloan and H. Woźniakowski, Periodization strategy may fail in high dimensions, *Numerical Algorithms* **46**, 369-391, 2007. 342

[124] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, Multivariate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces: Part I, *J. Approx. Theory* **152**, 135–160, 2008. 13

[125] F. Y. Kuo, G. W. Wasilkowski and H. Woźniakowski, Multivariate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces: Part II, to appear in *J. Approx. Theory* 13

[126] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski, On decompositions of multivariate functions, in progress. 198, 239

[127] M. Kwas, An optimal Monte Carlo algorithm for multivariate Feynman-Kac path integrals, *J. Math. Phys.* **46**, 103511, 2005. 3, 239

[128] M. Kwas and Y. Li, Worst case complexity of multivariate Feynman-Kac path integration, *J. Complexity* **19**, 730–743, 2003. 3, 239

[129] G. Larcher, Talk at Dagstuhl-Seminar "Algorithms and Complexity for Continuous Problems", 1998. 39

[130] F. M. Larkin, Gaussian measure in Hilbert space and application in numerical analysis, *Rocky Mount. J. Math.* **2**, 372–421, 1972. 153

[131] D. Lee and G. W. Wasilkowski, Approximation of linear functionals on a Banach space with a Gaussian measure, *J. Complexity* **2**, 12–43, 1986. 153

[132] I. Levine, *Quantum Chemistry* , 5th ed., Prentice-Hall, 2000. 3

[133] P. Lévy, *Processus Stochastiques et Monvement Brownian*, Gauthier-Villars, Paris, 1947. 51

[134] Y. Li, Applicability of Smolyak's algorithm to certain Banach spaces of functions, *J. Complexity* **18**, 792–814, 2002. 239

[135] M. A. Lifshits and E. V. Tulyakova, Curse of dimension in approximation of random fields, *Probab. Math. Stat.* **26**, 83–98, 2006. 86, 140

[136] M. A. Lifshits and M. Zani, Approximation complexity of additive random fields, *J. Complexity* **24**, 362–379, 2008. 86, 140

[137] R. Linde, *s*-numbers of diagonal operators and Besov embeddings, Rend. Circ. Mat. Palermo (2) Suppl. No. 10, 83–110, 1986. 124

[138] G. G. Lorentz, M. v. Golitschek and Y. Makovoz, *Constructive Approximation. Advanced Problems*, Springer-Verlag, Berlin, 1996. 124, 126

[139] G. G. Magaril-Ilyaev and K. Yu. Osipenko, Optimal reconstruction of functionals from inaccurate data, *Math. Notes* **50**, 1274–1279, 1992. 116

[140] P. Mathé, *s*-numbers in information-based complexity, *J. Complexity* **6**, 41–66, 1990. 116, 117, 144

[141] P. Mathé, Random approximation of Sobolev embeddings, *J. Complexity* **7**, 261–281, 1991. 74, 85, 140, 146, 285

[142] P. Mathé and E. Novak, Simple Monte Carlo and the Metropolis algorithm, *J. Complexity* **23**, 673–696, 2007. 153

[143] K. Meer and C. Michaux, A survey on real structural complexity theory, *Bull. Belg. Math. Soc.* **4**, 113–148, 1997. 5, 98, 100

[144] A. Messiah, *Quantum Mechanics*, vol. 1 and 2, North-Holland, Amsterdam, 1961/62. 3

[145] N. Metropolis and S. Ulam, The Monte Carlo method, *J. Amer. Statist. Assoc.* **44**, 335–341, 1949. 67

[146] C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, in: *Optimal Estimation in Approximation Theory*, C. A. Micchelli, T. J. Rivlin (eds.), Plenum, New York, 1977. 116

[147] C. A. Micchelli and G. Wahba, Design problems for optimal surface interpolation, in *Approximation Theory and Applications*, (Z. Ziegler, ed.), 329–347, Academic Press, New York, 1981. 61

[148] W. Miller, Computational complexity and numerical stability, *SIAM J. Comput.* **4**, 97–107, 1975. 104

[149] R. v. Mises, Zur mechanischen Quadratur, *Z. Angew. Math. Mech.* **13**, 53–56, 1933. 152

[150] G. M. Molchan, On some problems concerning Brownian motion in Lévy's sense, *Theory Probab. Appl.* **12**, 682–690, 1967. 57

[151] W. J. Morokoff, Generating quasi-random paths for stochastic processes, *SIAM Review* **40**, 765–788, 1998. 3

[152] Th. Müller-Gronbach and K. Ritter, Minimal errors for strong and weak approximation of stochastic differential equations, in: *Monte Carlo and Quasi-Monte Carlo Methods 2006*, A. Keller, S. Heinrich, H. Niederreiter (eds.), 53–82, Springer, Berlin, 2008. 3

[153] C. A. Neff and J. H. Reif, An efficient algorithm for the complex roots problem, *J. Complexity* **12**, 81–115, 1996. 105

[154] A. Nemirovsky, Polynomial time methods in convex programming, in: *The Mathematics of Numerical Analysis, 1995 AMS–SIAM Summer Seminar in Applied Mathematics*, J. Renegar, M. Shub and S. Smale (eds.), AMS Lect. Appl. Math. **32**, 543–589, 1996. 3

[155] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience, New York, 1983. 3, 140, 153

[156] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics, SIAM, Philadelphia, 1994. 3

[157] A. Neumayer, Molecular modeling of proteins and mathematical prediction of protein structure, *SIAM Review* **39**, 407–460, 1997. 3

[158] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992. 20, 35, 113

[159] S. M. Nikolskij, On the problem of approximation estimate by quadrature formulas (in Russian), *Usp. Mat. Nauk* **5**, 165–177, 1950. 152

[160] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, LNiM **1349**, Springer-Verlag, Berlin, 1988. 3, 71, 75, 86, 126, 127, 140, 146, 153

[161] E. Novak, Optimal linear randomized methods for linear operators in Hilbert spaces, *J. Complexity* **8**, 22–36, 1992. 74, 145, 147

[162] E. Novak, Optimal recovery and *n*-widths for convex classes of functions, *J. Approx. Theory* **80**, 390–408, 1995. 114

[163] E. Novak, The adaption problem for nonsymmetric convex sets, *J. Approx. Theory* **82**, 123–134, 1995. 114

[164] E. Novak, The real number model in numerical analysis, *J. Complexity* **11**, 57–73, 1995. 4, 98, 100, 140, 141

[165] E. Novak, On the power of adaption, *J. Complexity* **12**, 199–237, 1996. 110, 114, 152, 153

[166] E. Novak, The Bayesian approach to numerical problems: results for zero finding. Proc. of the IMACS-GAMM Int. Symp. on Numerical Methods and Error Bounds, Akademie-Verlag, Berlin, 1996, 164–171. 131

[167] E. Novak, Intractability results for positive quadrature formulas and extremal problems for trigonometric polynomials, *J. Complexity* **15**, 299–316, 1999. 18, 77

[168] E. Novak and H. Pfeiffer, Coin tossing algorithms for integral equations and tractability, *Monte Carlo Meth. Appl.* **10**, 491–498, 2004. 141, 153

[169] E. Novak and K. Ritter, A stochastic analog to Chebyshev centers and optimal average case algorithms, *J. Complexity* **5**, 60–79, 1989. 129, 144

[170] E. Novak and K. Ritter, Global optimization using hyperbolic cross points, in: *State of the Art in Global Optimization: Computational Methods and Applications*, C. A. Floudas and P. M. Pardalos (eds.), 19–33, Kluwer, Dordrecht, 1996. 3

[171] E. Novak, K. Ritter and H. Woźniakowski, Average Case Optimality of a Hybrid Secant-Bisection Method, *Math. Comp.* **64**, 1517-1539, 1995. 130, 131

[172] E. Novak, I. H. Sloan and H. Woźniakowski, Tractability of approximation for weighted Korobov spaces on classical and quantum computers, *Found. Comput. Math.* **4**, 121–156, 2004. 342

[173] E. Novak and H. Triebel, Function spaces in Lipschitz domains and optimal rates of convergence for sampling, *Constr. Approx.* **23**, 325–350, 2006. 124, 126, 127

[174] E. Novak and H. Woźniakowski, Topological complexity of zero finding, *J. Complexity* **12**, 380–400, 1996. 5, 98, 100, 107

[175] E. Novak and H. Woźniakowski, On the cost of uniform and nonuniform algorithms, *Theoretical Comp. Sc.* **219**, 301–318, 1999. 5, 98, 100

[176] E. Novak and H. Woźniakowski, Complexity of linear problems with a fixed output basis, *J. Complexity* **16**, 333–362, 2000. 5, 98, 100, 236

[177] E. Novak and H. Woźniakowski, When are integration and discrepancy tractable?, in: *Foundation of Computational Mathematics*, Oxford, 1999, R. A. DeVore, A. Iserles and E. Süli (eds.), 211–266, Cambridge University Press, Cambridge, 2001. 13, 19, 194

[178] E. Novak and H. Woźniakowski, Intractability results for integration and discrepancy, *J. Complexity* **17**, 388–441, 2001. 72

[179] E. Novak and H. Woźniakowski, Optimal order of convergence and (in-) tractability of approximation of smooth functions, submitted for publication. 86

[180] 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**, 903–920, 2006. 13

[181] D. Nuyens and R. Cools, Fast algorithms for component-by-component construction of rank-1 lattice rules with a non-prime number of points, *J. Complexity* **22**, 4–28, 2006. 13

[182] A. M. Ostrowski, On two problems in abstract algebra connected with Horner's rule, Studies presented to R. von Mises, Academic Press, 40–48, 1954. 99

[183] P. Oswald, On the degree of nonlinear spline approximation in Besov-Sobolev spaces, *J. Approx. Theory* **61**, 131–157, 1990. 96

[184] E. W. Packel, Do linear problems have linear optimal algorithms? *SIAM Rev.* **30**, 388–403, 1988. 116

[185] V. Ya. Pan, On some methods of computing polynomial values, *Problemy Kibernetiki*, **7**, 21–30, 1962, (in Russian), translation in *Problems of Cybernetics*, **7**, 20–30, 1962. 99

[186] V. Ya. Pan, *How to multiply matrices fast?* Lecture Notes in Computer Science, **179**, Springer-Verlag, Berlin, 1984. 4, 99

[187] C. H. Papadimitriou and J. Tsitsiklis, Intractable problems in control theory, *SIAM J. Control Optim.*, **24**, 639–654, 1986. 108

[188] A. Papageorgiou and G. W. Wasilkowski, On the average complexity of multivariate problems, *J. Complexity* **6**, 1–23, 1990. 59, 81, 86, 153

[189] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York. 115, 128

[190] H. Pfeiffer, *Monte Carlo Methods with Few Random Bits*, Shaker Verlag, Aachen, 2005. 141, 153

[191] A. Pietsch, *Eigenvalues and s-Numbers*, Cambridge University Press, Cambridge, 1987. 116, 118, 126

[192] A. Pietsch, *History of Banach Spaces and Linear Operators*, Birkhäuser, Boston, 2007. 118, 126

[193] A. Pinkus, *n-Widths in Approximation Theory*, Springer, Berlin, 1985. 116, 118, 126, 153

[194] L. Plaskota, A note on varying cardinality in the average case setting, *J. Complexity* **9**, 458–470, 1993. 137

[195] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge University Press, Cambridge, UK, 1996. 100, 105

[196] L. Plaskota and G. W. Wasilkowski, Adaption allows efficient integration of functions with unknown singularities, *Numer. Math.* **102**, 123–144, 2005. 115

[197] L. Plaskota, G. W. Wasilkowski and H. Woźniakowski, A new algorithm and worst case complexity for Feynman-Kac path integration, *J. Computational Physics* **164**, 335–353, 2000. 3, 12

[198] L. Plaskota, G. W. Wasilkowski and Y. Zhao, The power of adaption for approximating functions with singularities, to appear in *Math. Comp.* 115

[199] F. P. Preparata and M. I. Shamos, *Computational Geometry*, Springer Verlag, Berlin, 1993. 4, 99

[200] H. Rabitz and Ö. F. Alis, General foundations of high-dimensional model representations, *J. Math. Chemistry* **25** 197–233, 1999. 8, 41

[201] G. Rote, The convergence rate of the sandwich algorithm for approximating convex functions, *Computing* **48**, 337–361, 1992. 114

[202] K. F. Roth, On irregularities of distributions, *Mathematika* **1**, 73–79, 1954. 40

[203] D. Rudolf, Explicit error bounds for lazy reversible Markov chain Monte Carlo, *J. Complexity*, to appear. 153

[204] J. Rust, J. F. Traub and H. Woźniakowski, Is there a curse of dimensionality for contraction fixed points in the worst case?, *Econometrica* **70**, 285–329, 2002. 8

[205] V. S. Rychkov, On restrictions and extensions of the Besov and Triebel-Lizorkin spaces with respect to Lipschitz domains, *J. London Math. Soc.* **60**, 237–257, 1999. 126

[206] E. Salamin, Computation of pi using arithmetic-geometric mean, *Math. Comp.* **30**, 565–570, 1976. 100

[207] A. Sard, Best approximate integration formulas; best approximation formulas, *Amer. J. Math.* **71**, 80–91, 1949. 152

[208] S. Schäffler, Unconstrained global optimization using stochastic integral equations, *Optimization* **35**, 43–60, 1995. 3

[209] H.-J. Schmeisser, Recent developments in the theory of function spaces with dominating mixed smoothness, in: Nonlinear Analysis, Function Spaces and Applications, Vol. 8, Proc. Spring School Praha, May 30–June 6, 2006, J. Rákosník, Academy of Sciences of the Czech Republic, Praha 2007. 341, 342

[210] A. Schönhage, A. Grotefeld and E. Vetter, *Fast Algorithms, A Multiple Turing Machine Implementation*, Bibliographisches Institut, Mannheim, 1994. 105

[211] G. E. Shilov and Fan Dyk Tin, *Integral measure and derivative on linear spaces*, Nauka, Moscow, 1967 (in Russian). 261

[212] W. Sickel and T. Ullrich, Smolyak's algorithm, sampling on sparse grids and function spaces of dominating mixed smoothness, *East J. on Approx.* **13**, 387–425, 2007. 342

[213] I. H. Sloan, F. Y. Kuo and S. Joe, On the step-by-step construction of quasi-Monte Carlo integration rules that achieves strong tractability error bounds in weighted Sobolev spaces, *Math. Comp.* **71**, 1609–1640, 2002. 13

[214] I. H. Sloan and S. Joe, *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford, 1994. 20, 342

[215] I. H. Sloan and H. Woźniakowski, An intractability result for multiple integration, *Math. Comp.* **66**, 1119–1124, 1997. 21, 22

[216] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity* **14**, 1–33, 1998. 9, 19, 194, 240, 350

[217] I. H. Sloan and H. Woźniakowski, Tractability of integration for weighted Korobov spaces, *J. Complexity* **17**, 697–721, 2001. 78

[218] I. H. Sloan and H. Woźniakowski, Tractability of integration in non-periodic and periodic weighted tensor product Hilbert spaces, *J. Complexity* **18**, 479–499, 2002. 342, 350, 352, 354

[219] I. H. Sloan and H. Woźniakowski, When does Monte Carlo depend polynomially on the number of variables? In *Monte Carlo and quasi-Monte Carlo methods 2002*, 407–437, Springer, Berlin, 2004. 72, 84, 153

[220] S. Smale, On the topology of algorithms, *J. Complexity* **3**, 81–89, 1987. 98, 107

[221] S. Smale, Mathematical problems for the next century, *Math. Intell.* **20**, 7–15, 1998. 4

[222] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulation* **55**, 271–280, 2001. 41, 43

[223] Gy. Sonnevend, Sequential algorithms of optimal order for the uniform recovery of functions with monotone $(r-1)$ derivatives, *Analysis Math.* **10**, 311–335, 1984. 114

[224] E. M. Stein, *Singular integrals and differentiable properties of functions*, Princeton University Press, Princeton, 1970. 57

[225] R. Stevenson, Optimality of a standard adaptive finite element method, *Found. Comput. Math.* **7**, 245–269, 2007. 96, 113

[226] V. Strassen, Gaussian elimination is not optimal, *Numer. Math.* **13**, 354–356, 1969. 4, 99

[227] A. G. Sukharev, Optimal numerical integration formulas for some classes of functions, *Sov. Math. Dokl.* **20**, 472–475, 1979. 16

[228] A. G. Sukharev, On the existence of optimal affine methods for approximating linear functionals, *J. Complexity* **2**, 317–322, 1986. 116

[229] A. V. Suldin, Wiener measure and its applications to approximation methods, I (in Russian), *Izv. Vyssh. Ucheb. Zaved. Mat.* **13**, 145–158, 1959. 153

[230] A. V. Suldin, Wiener measure and its applications to approximation methods, II (in Russian), *Izv. Vyssh. Ucheb. Zaved. Mat.* **18**, 165–179, 1960. 153

[231] M. Talagrand, Sharper bounds for Gaussian and empirical processes, *Ann. Probability* **22**, 28–76, 1994. 39

[232] V. N. Temlyakov, *Approximation of Periodic Functions*, Nova Science, New York, 1994. 126, 342

[233] V. N. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.* **3**, 33–107, 2003. 96

[234] S. Tezuka, Financial applications of Monte Carlo and Quasi-Monte Carlo methods, in: *Random and Quasi-random Point Sets*, P. Hellekalek, G. Larcher (eds.), LNiM **138**, Springer-Verlag, Berlin, 1998. 3

[235] E. Thiémard, Computing bounds for the star discrepancy, *Computing* **65**, 169–186, 2000. 79

[236] C. Thomas-Agnan, Computing a family of reproducing kernels for statistical applications, *Numerical Algorithms* **13**, 21–32, 1996. 221, 351

[237] V. M. Tikhomirov, *Some Problems in Approximation Theory*, Moscow State University, Moscow, 1976 (in Russian). 60

[238] V. M. Tikhomirov, *Analysis II*, Encyclopaedia of Mathematical Sciences, Vol. 14, Springer-Verlag, Berlin, 1990. 124, 126

[239] J. F. Traub, A continuous model of computation, *Physics Today* May 1999, 39–44. 5, 100

[240] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information, Uncertainty, Complexity*, Addison-Wesley, Reading MA, 19983. 90

[241] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, Average case optimality for linear problems, *J. Theor. Comput. Sci.* **29**, 1–25, 1984. 145

[242] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988. 4, 26, 63, 64, 65, 66, 71, 75, 82, 91, 92, 110, 113, 116, 117, 118, 127, 129, 134, 136, 137, 140, 146, 152, 153, 261, 263, 360

[243] J. F. Traub and A. G. Werschulz, *Complexity and Information*, Cambridge University Press, Cambridge, 1998. 3, 8, 36, 86

[244] J. F. Traub and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980. 4, 26, 108, 110, 111, 116, 118

[245] J. F. Traub and H. Woźniakowski, Complexity of linear programming, *Oper. Res. Lett.* **1**, 59–62, 1982. 4

[246] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1978. 25, 126

[247] H. Triebel, Function spaces in Lipschitz domains and on Lipschitz manifolds. Characteristic functions as pointwise multipliers, *Revista Mat. Complutense* **15**, 475–524, 2002. 126

[248] H. Triebel, Sampling numbers and embedding constants, *Proc. Steklov Inst. Math.* **248**, 268–277, 2005. 123

[249] H. Triebel, *Theory of Function Spaces III*, Birkhäuser Verlag, Basel, 2006. 120, 123, 125, 126

[250] H. Triebel, Wavelet para-bases and sampling numbers in function spaces on domains, *J. Complexity* **23**, 468–497, 2007. 126

[251] N. N. Vakhania, V. I. Tarieladze and S. A. Chobanyan, *Probability Distributions on Banach Spaces*, Reidel Publishing, Kluwer, Dordrecht, 1987. 49, 357

[252] V. A. Vassiliev, Topological complexity of root-finding algorithms, in: *The Mathematics of Numerical Analysis*, J. Renegar, M. Shub, S. Smale (eds.), 831–856, Lectures in Applied Mathematics 32, AMS, 1996. 98, 107

[253] J. Vybíral, Sampling numbers and function spaces, *J. Complexity* **23**, 773–792, 2007. 123, 125, 126, 153

[254] J. Vybíral, Widths of embeddings in function spaces, to appear in *J. Complexity* 124, 125, 153

[255] D. H. Wagner, Survey of measurable selection theorems: an update, In: Measure Theory, D. Kölzow (ed.), Lecture Notes in Math. **794**, 176–216, Springer, 1980. 144

[256] X. Wang and K.T. Fang, Effective dimensions and quasi-Monte Carlo integration, *J. Complexity* **19**, 101–124, 2003. 43

[257] X. Wang, I. H. Sloan and J. Dick, On Korobov lattice rules in weighted spaces, *J. Numer. Anal.*, **42**, 1760–1779, 2004. 78

[258] G. W. Wasilkowski, Local average error, Report, Columbia University Computer Science Department, 1983. 129

[259] G. W. Wasilkowski, Some nonlinear problems are as easy as the approximation problem, Comput. Math. Appl. **10**, 351–363, 1984. 86

[260] G. W. Wasilkowski, Information of varying cardinality, *J. Complexity* **2**, 204–228, 1986. 61, 136, 153

[261] G. W. Wasilkowski, Optimal algorithms for linear problems with Gaussian measures, *Rocky Mount. J. Math.* **16**, 727–749, 1986. 63, 153

[262] G. W. Wasilkowski, Randomization for continuous problems, *J. Complexity* **5**, 195–218, 1989. 74, 140, 153

[263] G. W. Wasilkowski, Integration and approximation of multivariate functions: Average case complexity with isotropic Wiener measure, *Bull. Amer. Math. Soc.* **28**, 308-314, 1993; full version in *J. Approx. Theory* **77**, 212–227, 1994. 56

[264] G. W. Wasilkowski, On polynomial-time property for a class of randomized quadratures, *J. Complexity* **20**, 624–637, 2004. 84, 153

[265] G. W. Wasilkowski and F. Gao, On the power of adaptive information for functions with singularities, *Math. Comp.* **58**, 285–304, 1992. 115

[266] G. W. Wasilkowski and H. Woźniakowski, Can adaption help on the average? *Numer. Math.* **44**, 169–190, 1984. 145

[267] G. W. Wasilkowski and H. Woźniakowski, Average case optimal algorithms in Hilbert spaces. *J. Approx. Theory* **47**, 17–25, 1986. 153

[268] G. W. Wasilkowski and H. Woźniakowski, There exists a linear problem with infinite combinatory complexity, *J. Complexity* **9**, 326–337, 1993. 108

[269] G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multi-variate tensor product problems, *J. Complexity* **11**, 1–56, 1995. 238, 282

[270] G. W. Wasilkowski and H. Woźniakowski, On tractability of path integration, *J. Math. Phys.* **37**, 2071–2088, 1996. 3

[271] G. W. Wasilkowski and H. Woźniakowski, Weighted tensor-product algorithms for linear multivariate problems, *J. Complexity* **15**, 402–447, 1999. 13, 194, 201, 208, 209, 239, 240

[272] G. W. Wasilkowski and H. Woźniakowski, Finite-order weights imply tractability of linear multivariate problems, *J. Approx. Theory* **130**, 57–77, 2004. 239, 240

[273] G. W. Wasilkowski and H. Woźniakowski, Polynomial-time algorithms for multivariate linear problems with finite-order weights; worst case setting, *Found. Comput. Math.* **5**, 451–491, 2005. 174, 196, 239, 240

[274] K. Weihrauch, A refined model of computation for continuous problems, *J. Complexity* **14**, 102–121, 1998. 5, 100, 105

[275] K. Weihrauch, *Computable Analysis*, Springer-Verlag, Berlin, 2000. 105

[276] H. Wendland, Local polynomial reproduction and moving least squares approximation, *IMA J. Numerical Analysis* **21**, 285–300, 2001. 126

[277] A. G. Werschulz, *The Computational Complexity of Differential and Integral Equations: an Information-Based Approach*, Oxford University Press, New York, 1991. 86, 113, 126

[278] A. G. Werschulz, The complexity of two point boundary value problems with piece-wise analytic data, *J. Complexity* **10**, 367–383, 1994. 115

[279] A. G. Werschulz and H. Woźniakowski, Are linear algorithms always good for linear problems? *Aequationes Math.* **31**, 202–212, 1986. 108, 116, 117

[280] A. G. Werschulz and H. Woźniakowski, Tractability of multivariate approximation for a weighted unanchored Sobolev space: Smoothness sometimes hurts, in progress. 156, 221, 240, 351

[281] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, N. J., 1963. 101

[282] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford Univ. Press, London and New York, 1965. 101, 102

[283] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag, Berlin, 1995. 3

[284] J. O. Wojtaszczyk, Multivariate integration in $C^\infty([0,1]^d)$ is not strongly tractable, *J. Complexity* **19**, 638–643, 2003. 76

[285] H. Woźniakowski, Numerical Stability for Solving Nonlinear Equations, *Numer. Math.* **27**, 373–390, 1977. 102

[286] H. Woźniakowski, Average complexity for linear operators over bounded domains, *J. Complexity* **3**, 57–80, 1987. 263

[287] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity* **10**, 96–128, 1994. 5, 150, 154, 238, 282

[288] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, *J. of Computing and Information* **4**, 1–19, 1994. 5, 32, 179, 239, 295, 307

[289] H. Woźniakowski, Probabilistic setting of information-based complexity, *J. Complexity* **2** 255–269, 1986. 153

[290] H. Woźniakowski, Why does information-based complexity use the real number model? *Theoretical Comp. Sc.* **219**, 451–466, 1999. 5, 100, 152

[291] H. Woźniakowski, Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals, in: *Monte Carlo and Quasi-Monte Carlo Methods 1998*, H. Niederreiter and J. Spanier (eds.), 114–136, Springer-Verlag, Berlin, 1999. 19

[292] H. Woźniakowski, Tractability of tensor product linear operators in weighted Hilbert spaces, in: Festschrift for N. V. Vakhania, *Georgian Math. Journal* **8**, 415–426, 2001. 240

[293] H. Woźniakowski, On generalized tractability of multivariate problems, in: *Approximation Theory*, a volume dedicated to Borislav Bojanov, 312–324, Prof. M. Drinov Acad. Publ. House, Sofia, 2004. 287, 290, 293, 334

[294] H. Woźniakowski, Tractability of multivariate problems for weighted spaces of functions, in: Approximation and Probability, Banach Center Publications **72**, 407–427, Polish Academy of Sciences, Warszawa 2006. 13

[295] H. Yserentant, Sparse grid spaces for the numerical solution of the electronic Schrödinger equation, *Numer. Math.* **101**, 381–389, 2005. 3

[296] S. K. Zaremba, Some applications of multidimensional integration by parts, *Ann. Polon. Math.* **21**, 85–96, 1968. 36

Index