High-Dimensional Problems: Multivariate Linear Tensor Product Problems, Multivariate Numerical Integration, and Geometric Discrepancy

Habilitationsschrift
vorgelegt von

Michael Gnewuch

Kiel
2009

Preface

This cumulative habilitation thesis consists of a summary and an appendix, which contains the thirteen research articles [GnW07, GnW09, GnW08, GLSS07, Gne05, DGS05, Gne08a, Gne08b, DoG08, DGKP08, DGW09, Gne09, GSW09]. The articles are listed in the order of their appearance in the appendix.

In the summary the mathematical background of the problems treated in this thesis is described, and the research articles are discussed in detail.

Acknowledgement

I would like to thank David Bundy, Benjamin Doerr, and Henryk Woźniakowski for reading excerpts of earlier versions of the summary.

A considerable part of my work on the summary and the research articles collected here was done while I was visiting the University of Warsaw, the Max Planck Institute for Mathematics in the Sciences in Leipzig, the Max Planck Institute for Computer Science in Saarbrücken, the University of Jena, Columbia University New York, and the University of New South Wales in Sydney. I would like to thank the colleagues in these places for their hospitality.

Furthermore, I would like to acknowledge support from the DFG under the grants SR7/10-1, GN91/1-1, GN91/2-1, and GN91/3-1.

Last, but not least I would like to thank my collaborators for many inspiring discussions and for the work they contributed to the articles collected in this thesis.

Kiel, June 2009 Michael Gnewuch

Contents

Part 1

Summary

1.1 Introduction

High-dimensional mathematical problems appear in many applications in the natural sciences, in economics, and in finance. Examples of such problems are the solution of the Schrödinger equation for many-particle systems, the modeling and prediction of proteins or the pricing of complex financial derivatives.

Many high-dimensional problems suffer from the "curse of dimensionality". This phrase was coined by Bellman [Bel57] and describes the experience that the computational resources needed to solve a high-dimensional problem (within a prescribed error tolerance) often scale exponentially in the dimension. An important part of mathematical research is dedicated to the question of how to break the curse of dimensionality.

The mathematically posed high-dimensional problems we are concerned with in this habilitation thesis are *linear tensor product problems*, *multivariate numerical integration*, and problems in *geometric discrepancy theory*. As we shall explain later in more detail, the problems in geometric discrepancy we discuss here are intimately related to multivariate numerical integration, and fall under the heading *quasi-Monte Carlo integration*. Furthermore, multivariate numerical integration with respect to product measures (which is the case we are focusing on) is a particular linear tensor product problem.

The two aspects we are especially interested in are the computational complexity of these problems, and the design of efficient algorithms to solve them.

1.1.1 Computational Complexity

Computational complexity is a measure of the minimal computational resources needed to solve a mathematical problem. The computational complexity depends not on the specific computer or algorithm we have at our disposal to solve the problem, but only on the model of computation we use and the problem itself.

Before discussing the results of the articles collected in this thesis in detail, we want to describe the notions of computational complexity and the models of computation that we use in the different articles. This seems to be appropriate since it is not mentioned

explicitly in the articles. In particular, we want to explain why the notion of information complexity (which in the setting of discrepancy theory corresponds to the so-called inverse of the discrepancy) is most important for us.

Depending on the nature of the specific problem we study in the articles collected in this thesis, the underlying model of computation is the *Turing number model* (in [GSW09]) or the *real number model with oracle* (in almost all the other articles). Since the Turing number model is well-known in discrete mathematics and computer science, we do not feel the need to describe it here and refer to introductory texts in computer science or to [GaJ79, Hro03, Weg05]. The Turing number model is the underlying model of computation when we deal with the complexity of calculating the discrepancy of given sets and of related problems in computational geometry in [GSW09].

The real number model is, e.g., described in [BSS89], and its extension to include oracles is treated in [Nov95, Pla96]. We use the real number model with oracle for our complexity studies of linear tensor product problems and multivariate numerical integration. Before we mention the most important characteristics of this model of computation, let us describe those problems in the abstract language of *information-based complexity* (IBC) (see, e.g., [TWW88, TrW98]): For given m, $d \in \mathbb{N}$ we have a set F_d of functions defined on a fixed domain in $\mathbb{R}^{md}$, a normed linear space G_d, and a solution operator S_d mapping F_d into G_d. For a given ε we want to approximate S_d by an algorithm U_d with error at most ε. The sets F_d we study here will be normed linear spaces, the error criterion will be the worst-case error on the norm unit ball of F_d, normalized by dividing by the operator norm of S_d. In the case of multivariate numerical integration, e.g., G_d is nothing but the set of real numbers $\mathbb{R}$ and $S_d(f)$ is the integral of f with respect to some measure.

Since computers can only store finitely many numbers, the input of our algorithm cannot consist of general multivariate functions, but only of partial information about these functions. We assume that the partial information about a given $f \in F_d$ is of the form $N(f) = (L_1(f), \ldots, L_n(f))$, where $L_1, \ldots, L_n$ belong to a class Λ_d of admissible information functionals. In the case of linear tensor product problems we study the class $\Lambda_d = \Lambda_d^{\mathrm{all}}$ of all continuous linear functionals on F_d (*"linear information"*). In the case of multivariate numerical integration we restrict ourselves to the class $\Lambda_d = \Lambda_d^{\mathrm{std}}$ of function evaluations (*"standard information"*). In general, this information can be chosen adaptively, i.e., $n = n(f)$, $L_1(f) =: y_1$, $L_i(f; y_1, \ldots, y_{i-1}) =: y_i$ for $i = 2, \ldots, n(f)$, and $N(f) = (y_1, \ldots, y_{n(f)})$. An admissible algorithm is of the form $U_d(f) = \phi(N(f))$ for some mapping $\phi : N(F_d) \to G_d$, which belongs to an admissible class of mappings Φ.

In this abstract formulation it is easy to see the difference between our continuous multivariate problems and combinatorial problems like the traveling salesperson problem (TSP), which can be treated well within the Turing machine model:

The most important difference is that the information about a mathematical input $f \in F_d$ that can be utilized by a specific algorithm is only *partial*, since in the general case F_d is an infinite-dimensional space and information operators $N : F_d \to \mathbb{R}^n$ are therefore not injective, but many-to-one. This implies that in general we cannot hope to achieve $U_d(f) = S_d(f)$, since U_d is constant on the set $N^{-1}(N(f))$, while S_d may vary on

$N^{-1}(N(f))$. Thus the only hope we have is for a given $\varepsilon > 0$ to approximate $S_d(f)$ by $U_d(f)$ to within ε.

Notice that for combinatorial problems it is assumed that algorithms can dispose of the *full* information about the mathematical input: In the case of the TSP the (actual) input consists of the vertices of a graph, its edges, and the edge weights (at least if we confine ourselves to rational weights). We will see later that similarly the problem of calculating the star discrepancy of a given set as considered in [GSW09] can be viewed as a discrete enumeration problem with complete information.

In IBC it is assumed that information is not only partial, but also *priced*. This is motivated from problems in scientific computing: If we want to evaluate, e.g., some function that describes some state in physics, meteorology or geoscience, we usually have to make expensive measurements. Also the numerical evaluation of complex functions or functions given only implicitly (e.g., as solutions of certain optimization problems) may be quite costly.

This motivates the assumption that each information evaluation (i.e., evaluation of some admissible functional) has cost $c > 0$. This is one important attribute of the real number model with oracle: Each information evaluation, that is each evaluation of a functional $L \in \Lambda_d$ at some $f \in F_d$, has cost c. This cost is independent of $L \in \Lambda_d$ and $f \in F_d$. Another important attribute of the model is that combinatory operations such as arithmetic operations in $\mathbb{R}$, addition and scalar multiplication in G_d, comparisons and evaluation of elementary functions can be performed with the same cost; for simplicity we assume with unit cost. These features are an abstraction of fixed precision floating point arithmetic. As indicated by our discussion above, it is reasonable to assume that $c \gg 1$.

A deeper discussion why the real number model is appropriate for scientific computation can be found in [NoW08, Sect. 4.3.1], advantages and disadvantages of the real number model and the Turing number model are discussed in [Tra99].

We define the cost $\mathrm{cost}(U_d, f)$ of computation of our algorithm U_d for a mathematical input f by counting the number $n_{\mathrm{comb}}(f)$ of combinatory operations of the algorithm and adding the number $n_{\mathrm{info}}(f) = n(f)$ of information operations it uses multiplied by the cost of a single information operation:

$$\mathrm{cost}(U_d, f) = c n_{\mathrm{info}}(f) + n_{\mathrm{comb}}(f). \tag{1.1}$$

Note that this definition of cost does not explicitly take into account memory usage, an issue that may lead to some problems in practice, see, e.g., the comment in [DGW09, Sect. 3]. Furthermore, costs that are due to precomputation (as, e.g., the precomputation of weights and integration points of cubature formulas for numerical integration) are not considered here.

The *worst case error* of the algorithm U_d is given by

$$e^{\mathrm{wor}}(U_d) = \sup_{\|f\|_{F_d} \le 1} \|S_d(f) - U_d(f)\|_{G_d},$$

where $\|\cdot\|_{F_d}$ and $\|\cdot\|_{G_d}$ are the norms in F_d and G_d. As already mentioned, the error criterion we use is rather the *normalized error* given by

$$e^{\mathrm{nor}}(U_d) = e^{\mathrm{wor}}(U_d)/\|S_d\|$$

than the worst case error; here $\|S_d\|$ denotes the operator norm of S_d. The reason for this is that the norm $\|S_d\|$ can be seen as "initial error" (cf., e.g., [GnW07, Sect. 2.1]), and dividing the worst case error by it scales the problem properly. In [GnW07, GnW09] we do this scaling implicitly, since the assumptions made there lead to $\|S_d\| = 1$.

The *total complexity* $\mathrm{comp}(\varepsilon, S_d, \Lambda_d)$ is now defined by

$$\mathrm{comp}(\varepsilon, S_d, \Lambda_d) = \inf_{e^{\mathrm{nor}}(U_d) \le \varepsilon} \sup_{f \in F_d} \mathrm{cost}(U_d, f),$$

where the infimum is taken over all admissible algorithms U_d. The *information complexity* $n(\varepsilon, S_d, \Lambda_d)$ is defined by

$$n(\varepsilon, S_d, \Lambda_d) = \min\{n \in \mathbb{N} \mid \exists U_{d,n} : e^{\mathrm{nor}}(U_{d,n}) \le \varepsilon\};$$

here $U_{d,n}$ is an algorithm of the form $\varphi \circ N$ such that N uses at most n admissible information functionals from Λ_d.

Due to (1.1) the information complexity is always a lower bound for the total complexity. For the continuous multivariate problems we discuss here it is known that adaptive information does not help and that linear algorithms (using non-adaptive information) are optimal, see, e.g., [TWW88, Sect. 4.5] or [NoW08, Sect. 4.2]. This means for upper bounds on the total complexity it suffices to study linear algorithms of the form

$$U_d(f) = \varphi(N(f)) = \sum_{k=1}^{n} a_k L_k(f)$$

for $a_1, \ldots, a_n \in G_d$ and non-adaptive information $N(f) = [L_1(f), \ldots, L_n(f)]$. Since we view the determination of $a_1, \ldots, a_n$ as precomputation, the combinatory costs of such a linear algorithm U_d are $2n - 1$ and its information complexity is n. Since $c \gg 1$, the cost $\mathrm{cost}(U_d, f)$ is dominated by the information cost cn_{info}. That is why in the continuous problems investigated here we do not care about combinatorial costs but confine ourselves to the study of the information complexity (or, in the case of the discrepancy problem, of the inverse of the discrepancy). Recall that for a discrete problem with complete information no information complexity arises and that its computational complexity is due to the combinatory operations necessary to solve it.

1.1.2 Generalized Tractability

Since we are interested in arbitrarily large dimension, it is not sufficient to determine solely the dependence of the information complexity on the approximation error ε, but it is necessary to study the explicit dependence on both parameters ε and d. This issue was addressed in [Woź94a] by introducing the notion of tractability. A multivariate problem $S = (S_d)_{d \in \mathbb{N}}$ is called *tractable* if there exist positive constants C,p and q such that

$$n(\varepsilon, S_d, \Lambda_d) \le C \varepsilon^{-p} d^q \quad \text{for all } (\varepsilon^{-1}, d) \in [1, \infty) \times \mathbb{N}. \tag{1.2}$$

A multivariate problem $S = (S_d)_{d\in\mathbb{N}}$ is called *strongly tractable* if there exist positive constants C and p such that

$$n(\varepsilon, S_d, \Lambda_d) \le C\varepsilon^{-p} \quad \text{for all } (\varepsilon^{-1}, d) \in [1,\infty)\times\mathbb{N}. \tag{1.3}$$

In [GnW07, GnW08, GnW09] these notions are relaxed considerably. Let Ω be an unbounded subset of $[1,\infty)\times\mathbb{N}$, and let $T : [1,\infty)\times\mathbb{N} \to [0,\infty)$ be a function that is non-decreasing with respect to both variables. We require T not to grow exponentially in Ω if $\varepsilon^{-1} + d$ approaches infinity. Such a domain Ω is called *tractability domain*, the function T is called *tractability function*. In [GnW07], we call a multivariate problem $S = (S_d)_{d\in\mathbb{N}}$ (T,Ω)-*tractable* if there exist positive numbers C and t such that

$$n(\varepsilon, S_d, \Lambda_d) \le CT(\varepsilon^{-1}, d)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega. \tag{1.4}$$

The infimum t^{tra} over all t satisfying (1.4) is called the *exponent of* (T,Ω)-*tractability*. A multivariate problem $S = (S_d)_{d\in\mathbb{N}}$ is called *strongly* (T,Ω)-*tractable* if there exist positive numbers C and t such that

$$n(\varepsilon, S_d, \Lambda_d) \le CT(\varepsilon^{-1}, 1)^t \quad \text{for all } (\varepsilon^{-1}, d) \in \Omega. \tag{1.5}$$

The infimum t^{str} over all t satisfying (1.5) is called the *exponent of strong* (T,Ω)-*tractability*.

In particular, the notion of domain of tractability and the general form of the tractability function extend an earlier relaxation of the definition of tractability made in [Woź04], where T was required to be of product form $T(x,y) = f_1(x)f_2(y)$.

The motivation for these generalizations is described in detail in [GnW07]. One reason is the following: Many models used in applications in the sciences or in finance are rather weak and imprecise; this is particularly the case in mathematical finance. For those models it makes not much sense to try to obtain an approximate solution with an error less than, e.g., $\varepsilon_0 = 0.01$. This would be addressed by the choice of the tractability domain $\Omega = [1,100)\times\mathbb{N}$.

In the articles [GnW08, GnW09] also *weak tractability* of multivariate problems is studied. A multivariate problem $S = (S_d)_{d\in\mathbb{N}}$ is called *weakly tractable in* Ω if

$$\lim_{(\varepsilon^{-1},d)\in\Omega\,;\,\varepsilon^{-1}+d\to\infty} \frac{\ln n(\varepsilon, S_d, \Lambda_d)}{\varepsilon^{-1}+d} = 0. \tag{1.6}$$

Weak tractability was first studied in [GnW08] and in [NoW08].

In the light of these generalized notions of tractability we refer to "classical" tractability and strong tractability defined by (1.2) and (1.3) as *polynomial tractability* and *strong polynomial tractability*.

Looking at these new definitions of tractability, one may ask the question if they are really necessary and useful. More to the point, one may ask in particular the following questions.

Question 1.1.1. Are there multivariate problems ...

(i) ... where restriction of the tractability domain helps to achieve tractability?

(ii) ... that are not polynomially tractable, but weakly tractable?

(iii) ... that are not polynomially tractable, but tractable with respect to another tractability function?

(iv) ... where it is more adequate to consider tractability functions not of product form?

In the next section we shall see that the answer to each of these four questions is affirmative and that these affirmative answers are demonstrated for certain multivariate linear tensor product problems.

In the following three sections we discuss the results from the articles collected in this thesis and explain their close relations. Results on generalized tractability for multivariate linear tensor product problems are considered in Section 1.2. Results on generalized tractability for linear functionals, and in particular for multivariate integration, are considered in Section 1.3.

1.2 Multivariate Linear Tensor Product Problems

In this section we consider multivariate problems defined as linear tensor product problems and study generalized tractability. Let us start by describing the setting in detail.

Let F_1 be a separable Hilbert space of real valued functions defined on a domain $D_1 \subseteq \mathbb{R}^m$, and let G_1 be an arbitrary separable Hilbert space. Let the solution operator $S_1 : F_1 \to G_1$ be a compact linear operator, and let $(\sigma_i)_{i\in\mathbb{N}}$ be the sequence of its singular values; we assume $\sigma_1 \geq \sigma_2 \geq \ldots$. In [GnW07, GnW09] we use the notation $\lambda_i = \sigma_i^2$ for $i = 1, 2, \ldots$. Without loss of generality, we assume that S_1 is not the zero operator, and normalize the problem by assuming that $\lambda_1 = 1$, implying $\|S_1\| = 1$.

For $d \geq 2$, let

$$F_d \;=\; F_1 \otimes \cdots \otimes F_1$$

be the complete d-fold tensor product Hilbert space of F_1 of real valued functions defined on $D_d = D_1 \times \cdots \times D_1 \subseteq \mathbb{R}^{d\,m}$. Similarly, let $G_d = G_1 \otimes \cdots \otimes G_1$, d times.

The linear solution operator S_d is defined as the tensor product operator

$$S_d \;=\; S_1 \otimes \cdots \otimes S_1 : F_d \to G_d.$$

We have $\|S_d\| = \|S_1\|^d = 1$ for all d. We call the multivariate problem $S = (S_d)_{d\in\mathbb{N}}$ a *linear tensor product problem*. In the papers [GnW07, GnW09] we analyze the problem S for the class of linear information $\Lambda^{\mathrm{all}} = (\Lambda_d^{\mathrm{all}})_{d\in\mathbb{N}}$. It is known that

$$n(\varepsilon, S_d, \Lambda_d^{\mathrm{all}}) = |\{(i_1, \ldots, i_d) \in \mathbb{N}^d \,|\, \lambda_{i_1} \ldots \lambda_{i_d} > \varepsilon^2\}|, \tag{1.7}$$

see e.g., [TWW88].

In [GnW07] we study *restricted tractability*, where the tractability domain Ω is of the form

$$\Omega^{\mathrm{res}} = [1,\infty) \times \{1,\dots,d^*\} \cup [1,\varepsilon_0^{-1}) \times \mathbb{N}, \quad d^* \in \mathbb{N} \cup \{0\} \text{ and } \varepsilon_0 \in (0,1], \tag{1.8}$$

where either $d^* > 0$ or $\varepsilon_0 < 1$. In [GnW09] we study *unrestricted tractability*, i.e., $\Omega = \Omega^{\mathrm{unr}}$, where

$$\Omega^{\mathrm{unr}} = [1,\infty) \times \mathbb{N}. \tag{1.9}$$

In both papers different rates of decay of the singular values of S_1 are considered. In each of these settings we provide necessary and sufficient conditions on the function T to guarantee that the linear tensor product problem S is (T,Ω)-tractable. Roughly speaking, these conditions are satisfied if certain limits of $\ln T(\varepsilon^{-1},d)$ divided by some functions of ε^{-1} and d, which depend on the decay rate of the singular values of S_1 and on λ_2, are strictly positive if $\varepsilon^{-1}+d$ approaches infinity. With the help of these limits we can also give good upper and lower bounds on the exponent of (T,Ω)-tractability. In the case where we know the decay rate of $(\lambda_j)_{j\in\mathbb{N}}$ explicitly, we can usually determine t^{tra} exactly.

The articles [GnW07, GnW09] consider many different settings; here we confine ourselves to highlight only a few results. We assume from now on that the minimal error ε_0 in (1.8) is strictly smaller than 1.

A first orientation is given by the following simple lemma, see [GnW07, Lemma 3.1].

Lemma 1.2.1. *Let T be an arbitrary tractability function, and let Ω contain some restricted tractability domain Ω^{res} as defined above. Let $S = (S_d)_{d\in\mathbb{N}}$ be a linear tensor product problem defined as above, and $\Lambda = (\Lambda_d)_{d\in\mathbb{N}}$ be an arbitrary class of information evaluations.*

- *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 $n(\varepsilon,S_d,\Lambda_d^{\mathrm{all}}) = 1$ for all $(\varepsilon^{-1},d) \in \Omega$ with $\varepsilon < 1$, and $t^{\mathrm{str}} = 0$.*

That is why we confine ourselves in the rest of this section to the interesting case where $1 = \lambda_1 > \lambda_2 > 0$.

Let us now consider polynomial tractability, i.e., $T(\varepsilon^{-1},d) = \varepsilon^{-1}d$. Then for the unrestricted domain Ω^{unr} we do not have $(T,\Omega^{\mathrm{unr}})$-tractability of S, even not in the case where $0 = \lambda_3 = \lambda_4 = \dots$ [Woź94b, Thm. 3.1]. For the restricted domain Ω^{res} however, S is $(T,\Omega^{\mathrm{res}})$-tractable if $\lambda_j = O(j^{-\beta})$ for some positive β and all $j \in \mathbb{N}$ [GnW07, Sect. 4.3]. In this case restriction of the tractability domain does indeed help to achieve polynomial tractability. Thus the answer to Question 1.1.1(i) is affirmative.

If $\lambda_j = \Omega(\exp(-\beta(\ln j)^\alpha))$ for some $\alpha \in (0,1)$, some positive β, and all $j \in \mathbb{N}$, then S is not $(T,\Omega^{\mathrm{res}})$-tractable [GnW07, Sect. 4.3]. In this situation (reasonable) restriction of the tractability domain does not help to achieve polynomial tractability. But even in this case we have weak tractability as long as $\lambda_j = o((\ln j)^{-2})$ for all $j \in \mathbb{N}$. This has

recently been proved by A. Papageorgiou and I. Petras [PaP09], improving the slightly weaker result in [GnW09, Thm. 6.1] and [NoW08, Thm. 5.5], where also the necessity of this condition for weak tractability is stated. The result from [PaP09] and also its slightly weaker predecessors show that the answer of Question 1.1.1(ii) is affirmative.

Let us now consider the tractability function

$$T(\varepsilon^{-1}, d) = \exp((1 + \ln \varepsilon^{-1})(1 + \ln d)). \tag{1.10}$$

If $\lambda_j = O(j^{-\beta})$ for some positive β and all $j \in \mathbb{N}$, then S is $(T, \Omega^{\mathrm{unr}})$-tractable and thus in particular $(T, \Omega^{\mathrm{res}})$-tractable [GnW09, Cor. 5.2]. This underlines again that the information complexity of problems which are not polynomially tractable does not necessarily depend exponentially on ε^{-1} or d in some direction. In other words, choosing a different tractability function than $\varepsilon^{-1}d$ can indeed help to achieve (T, Ω)-tractability. Thus also the answer to Question 1.1.1(iii) is affirmative.

It is instructive to compare the tractability function T in (1.10) with

$$T'(\varepsilon^{-1}, d) = \exp((1 + \ln \varepsilon^{-1})^{\mu} + (1 + \ln d)^{\nu}),$$

where μ, ν are positive. Let again $\lambda_j = O(j^{-\beta})$ for some positive β. Due to [Woź04] it is necessary and sufficient for S to be $(T', \Omega^{\mathrm{unr}})$-tractable that $(\mu - 1)(\nu - 1) \geq 1$ holds; see also the more general result [GnW09, Thm. 5.3]. So let us assume that $(\mu-1)(\nu-1) = 1$. From Theorem 5.3 and identity (32) in [GnW09] we obtain for the exponent of $(T', \Omega^{\mathrm{unr}})$-tractability

$$(t')^{\mathrm{tra}} = \frac{2}{\ln(\lambda_2^{-1})\mu^{1/\mu}\nu^{1/\nu}}.$$

If we furthermore assume that $\beta \geq \ln(\lambda_2^{-1})$, then the exponent of $(T, \Omega^{\mathrm{unr}})$-tractability is given by $t^{\mathrm{tra}} = 2/\ln(\lambda_2^{-1})$, see [GnW09, Cor. 5.2].

We always have $T(\varepsilon^{-1}, d)^{t^{\mathrm{tra}}} \leq T'(\varepsilon^{-1}, d)^{(t')^{\mathrm{tra}}}$, and for

$$\varepsilon^{-1} = \frac{1}{e} \exp\left((\nu - 1)^{1-1/\nu}(1 + \ln(d))^{\nu-1}\right)$$

we have $T(\varepsilon^{-1}, d)^{t^{\mathrm{tra}}} = T'(\varepsilon^{-1}, d)^{(t')^{\mathrm{tra}}}$. For fixed parameters ε^{-1} or d, the function $T(\varepsilon^{-1}, d)$ grows polynomially in the other variable, while $T'(\varepsilon^{-1}, d)$ grows sub-exponentially in the other variable. This shows that T describes the growth of $n(\varepsilon, S_d, \Lambda_d^{\mathrm{all}})$ more precisely than T', and underlines that it is not a good idea to require tractability functions to be of product form in the variables ε^{-1} and d. In particular, the comparison of T and T' shows that the answer to Question 1.1.1(iv) is affirmative.

1.2.1 Conclusion and Open Problems

The question of whether a multivariate linear tensor product problem is tractable (in the generalized sense) is essentially settled for the case where all variables are equally important and the admissible information is the class of linear information. Indeed,

the articles [GnW07, GnW09] and the very recent result [PaP09] leave only open some questions of detail.

As described in the introduction of [GnW07], it would be of interest to study which results on generalized tractability still hold if we consider instead of linear information the smaller class of standard information. Furthermore, it would be interesting to study the case where not all variables are equally important; this can be modelled by considering *weighted* tensor product spaces.

A first step to approach these modified problems has been made in the article [GnW08], where indeed standard information as well as weights are considered. But instead of studying general linear tensor product problems, we there restrict ourselves to linear functionals with focus on multivariate integration. The article [GnW08] will be discussed in Section 1.3.1.

1.3 Multivariate Numerical Integration

Multivariate integrals occur in various applications such as option pricing, statistical mechanics, the calculation of path integrals or the solution of integral or partial differential equations. As described in [LEL00], in most stochastic simulations the goal is to estimate the mathematical expectation of some kind of cost function. And even if this is not the ultimate goal, the estimation problem appears usually at an intermediate stage. The mathematical expectation that has to be estimated can generally be represented as an integral of some real-valued function f over the d-dimensional unit cube $[0,1]^d$, where d is the number of independent and identically distributed random variables over the interval $[0,1]$ that are used to generate randomness.

To tackle the multivariate integration problem one may use in dimension $d = 2$ or 3 product rules of one-dimensional quadrature rules. But for large d this approach is too costly, since the number of function calls grows exponentially with the dimension.

Which integration methods are known to break the curse of dimensionality (at least to some extent)? Apart from the probabilistic *Monte Carlo method* there are mainly two important deterministic approaches. On the one hand there are *quasi-Monte Carlo (QMC) methods*, including cubature rules based on *low-discrepancy sets* and on *lattice rules*, see, e.g., [Nie92, SlJ94]. On the other hand there are *sparse grid methods*, based on good one-dimensional quadrature rules and *Smolyak's construction*, see, e.g., [GeG98, NoR96, WaW95]. Many numerical examples demonstrate that in moderate dimension and for smooth integrands the sparse grid method is more efficient than QMC-methods based on low-discrepancy sets; for non-smooth integrands one observes the opposite result [GeG98]. Lattice rules seem to be especially advantageous if the integrands are periodic. The progress on the fast component-by-component construction of so-called *rank*-1 *lattice rules* in the last years by Cools, Joe, Kuo, Nuyens, Sloan, and their collaborators (see, e.g., [Kuo03, NuC06, SKJ02]) makes lattice rules a very competitive method. Nowadays, lattice rules are also used successfully for non-periodic functions.

Until recently it was commonly believed that quasi-Monte Carlo methods based on low-discrepancy sets cannot make use of additional smoothness of (non-periodic) inte-

grands (see, e.g., [GeG98]). This was proved wrong by J. Dick. In [Dic08a] he showed for integrands, whose mixed partial derivatives up to order s are square integrable, a convergence rate of the integration error of $O(n^{-s}(\log n)^{sd})$ by using special digital nets. For fixed dimension d this order of convergence is optimal up to logarithmic factors, since a lower bound for arbitrary cubatures is given by $\Omega(n^{-s}(\log n)^{(d-1)/2})$. Similar, but less general results for periodic functions were presented by J. Dick in the earlier paper [Dic07a]. In both papers the decay behavior of Walsh coefficients of smooth functions plays a central role.

In this section we investigate multivariate numerical integration on Hilbert spaces of d-variate functions. In the first article [GnW08] we present here, quasi-Monte Carlo rules are used to establish upper bounds on the worst case error of multivariate integration, in the second article [GLSS07] upper bounds are established with the help of sparse grid methods.

1.3.1 Generalized Tractability for Linear Functionals

In [GnW08] we start by studying a more general problem than multivariate numerical integration, namely the *approximation of continuous linear functionals*. In the setting we consider we have *reproducing kernel Hilbert spaces* H_1, H_2 of functions $f : D_1 \subseteq \mathbb{R} \to \mathbb{R}$, and F_1 is the orthogonal sum of H_1 and H_2 (and thus itself a reproducing kernel Hilbert space). We assume that the kernel of H_2 is *decomposable* (for details see [GnW08, Sect. 4]). The solution operator $I_1 : F_1 \to \mathbb{R}$ is here a continuous linear functional. For $\gamma > 0$ we define a new reproducing kernel Hilbert space $F_{1,\gamma}$ by endowing F_1 with the norm $\|\cdot\|_{F_{1,\gamma}}$ given by

$$\|f_1 + f_2\|^2_{F_{1,\gamma}} = \|f_1\|^2_{H_1} + \gamma^{-1}\|f_2\|^2_{H_2} \quad \text{for all } f_1 \in H_1,\ f_2 \in H_2.$$

Since the norms of F_1 and $F_{1,\gamma}$ are equivalent, the functional I_1 is also a continuous linear functional on $F_{1,\gamma}$.

We consider positive *weights* $\gamma = (\gamma_{d,j})_{d\in\mathbb{N};j=1,\ldots,d}$, which are non-increasing with respect to both indices. We define the tensor product Hilbert space $F_{d,\gamma}$ of d-variate functions given by $F_{d,\gamma} = F_{1,\gamma_{d,1}} \otimes \cdots \otimes F_{1,\gamma_{d,d}}$. The solution functional is $I_{d,\gamma} = I_d$, defined as the d-fold tensor product operator of I_1.

The weights $\gamma_{d,j}$ are useful to control the influence of the jth variable on the d-dimensional problem. The smaller the weight $\gamma_{d,j}$, the smaller is the importance of the jth coordinate. With the help of such weights one can model situations where some of the variables are more relevant than others; this concept was introduced in the paper [SlW98] and has turned out to be extremely fruitful in tractability studies.

What are the classes Λ_d of admissible information we study? Obviously, it is here completely uninteresting to consider linear information $\Lambda^{\mathrm{all}}_d = F^*_{d,\gamma}$ as in [GnW07, GnW09]: Since $I_{d,\gamma} \in \Lambda^{\mathrm{all}}_d$, we have $n(\varepsilon, I_{d,\gamma}, \Lambda^{\mathrm{all}}_d) \le 1$ for all $\varepsilon > 0$. Thus we study the more interesting case where only function evaluations are admissible, i.e., where $\Lambda_d = \Lambda^{\mathrm{std}}_d$.

In [GnW08, Thm. 1] we consider a domain of tractability Ω that contains $[1, \varepsilon_0^{-1}) \times \mathbb{N}$ for some $\varepsilon_0 \in (0, 1)$. Furthermore, we consider an arbitrary tractability function T. We state necessary conditions for (T, Ω)-tractability and strong (T, Ω)-tractability, as well

as for weak tractability in Ω of the multivariate problem $I_\gamma = (I_{d,\gamma})_{d\in\mathbb{N}}$. The necessary conditions require certain limits of the sums of the weights to be finite or even zero. So for weak tractability in Ω it is, e.g., necessary that the restriction of I_1 to H_1 is not the zero functional and

$$\lim_{d\to\infty} \frac{\sum_{j=1}^{d} \gamma_{d,j}}{d} = 0. \tag{1.11}$$

[GnW08, Thm. 1] extends [NoW01b, Thm. 3], where polynomial and strong polynomial tractability on the unrestricted tractability domain Ω^{unr} and positive decaying weights $(\gamma_j)_{j\in\mathbb{N}}$, independent of the dimension d, were considered. In fact, our proof of [GnW08, Thm. 1] follows the line of proof of [NoW01b, Thm. 3] and makes use of [NoW01b, Thm. 2].

In [GnW08, Sect. 6] we consider again reproducing kernel Hilbert spaces F_d of d-variate functions, but this time we do not require that the kernel of F_1 has a decomposable component, and F_d is not necessarily the d-fold tensor product of F_1 any more. The continuous linear functional I_d considered here is multivariate integration on a domain $D_d \subseteq \mathbb{R}^d$ against some probability density ρ_d. The main result of that section is Theorem 2, which provides sufficient conditions on the kernel K_d of F_d for weak tractability in Ω, (T,Ω)-tractability and strong (T,Ω)-tractability for multivariate integration. The sufficient conditions follow from upper bounds on $n(\varepsilon, I_d, \Lambda_d^{\mathrm{std}})$, which are obtained by averaging over the square of the worst case error of all quasi-Monte Carlo algorithms

$$Q_{n,d}f = \frac{1}{n}\sum_{i=1}^{n} f(z_i), \quad (z_1,\dots,z_n) \in D_d^n,$$

with respect to the (nd)-dimensional measure $\otimes_{i=1}^{n}\rho_d(z_i)\,dz_i$ on D_d^n. This technique can be found in [SlW98] and leads directly to

$$n(\varepsilon, I_d, \Lambda_d^{\mathrm{std}}) \le \lceil \eta_d \varepsilon^{-2} \rceil, \tag{1.12}$$

where

$$\eta_d = \left(\frac{C_d^{\mathrm{std}}}{C_d^{\mathrm{init}}}\right)^2 - 1,$$

and

$$C_d^{\mathrm{std}} = \left(\int_{D_d} \rho_d(x) K_d(x,x)\,dx\right)^{1/2}$$

and

$$C_d^{\mathrm{init}} = \left(\int_{D_d}\int_{D_d} \rho_d(x)\rho_d(y) K_d(x,y)\,dx\,dy\right)^{1/2}.$$

Thus [GnW08, Thm. 2] states, e.g., that

$$\lim_{d\to\infty} \frac{\ln \max\{1,\eta_d\}}{d} = 0 \tag{1.13}$$

implies weak tractability of I_d in Ω.

Our result [GnW08, Thm. 2] has been improved to some extent by L. Plaskota, G. W. Wasilkowski, and Y. Zhao in the very recent article [PWZ09]. There the authors define the quantity

$$C_d^{\mathrm{new}} = \int_{D_d} \rho_d(x) \sqrt{K_d(x,x)}\, dx$$

and show with the help of a new averaging technique that also

$$n(\varepsilon, I_d, \Lambda_d^{\mathrm{std}}) \le \lceil \tilde{\eta}_d \varepsilon^{-2} \rceil$$

holds, where

$$\tilde{\eta}_d = \left(\frac{C_d^{\mathrm{new}}}{C_d^{\mathrm{init}}} \right)^2 - 1.$$

Notice that always

$$C_d^{\mathrm{init}} \le C_d^{\mathrm{new}} \le C_d^{\mathrm{std}},$$

and that there are examples where C_d^{std} is infinite, but C_d^{new} is finite, see [PWZ09, Sect. 6]. Thus the result [PWZ09, Thm. 7] is applicable in some situations where [GnW08, Thm. 2] is not applicable any more (due to $C_d^{\mathrm{std}} = \infty$), and leads in general to a better upper bound on the information complexity.

In [GnW08, Sect. 7] we give three examples of multivariate integration problems, where the necessary conditions from Theorem 1 and the sufficient conditions from Theorem 2 indeed coincide. With regard to geometric discrepancy, which will be discussed in Section 1.4, the most interesting example is the Sobolev space for the bounded domain $D_d = [0,1]^d$. There $F_{1,\gamma}$ is the Sobolev space of absolutely continuous functions defined on $D_1 = [0,1]$, whose first derivative is square integrable. The inner product is given by

$$\langle f, g \rangle_{F_{1,\gamma}} = f(\tfrac{1}{2}) g(\tfrac{1}{2}) + \gamma^{-1} \int_0^1 f'(x) g'(x)\, dx. \tag{1.14}$$

The space $F_{d,\gamma}$ is given by $F_{d,\gamma} = \otimes_{i=1}^d F_{i,\gamma_{d,i}}$ for weights $\gamma = (\gamma_{d,j})_{d\in\mathbb{N};j=1,\ldots,d}$ as introduced above. As it can be seen in [GnW08, Thm. 3], in this case the necessary and sufficient conditions from Theorem 1 and 2 coincide. Consequently $I_\gamma = (I_{d,\gamma})_{d\in\mathbb{N}}$ is, e.g., weakly tractable in Ω if and only if condition (1.11) holds.

The results from [GnW08, Thm. 3] are directly applicable to geometric discrepancy, as we will explain now. For $u \subseteq \{1,\ldots,d\}$ and $x \in [0,1]^d$ let $(x_u, 1/2)$ denote the vector $\xi \in [0,1]^d$ with $\xi_j = x_j$ if $j \in u$, and $\xi_j = 1/2$ else. Let $J(x_u)$ be the axis-parallel box in $[0,1]^{|u|}$ with lower left and upper right corners given by x_u and a point from $\{0,1\}^{|u|}$, whose volume is minimal (this box is uniquely defined for almost all $x_u \in [0,1]^{|u|}$). Furthermore, denote by $1_{J(x_u)}$ the characteristic function of $J(x_u)$. As described in [NoW01b, Sect. 4] (see also [Hic98, NoW09]), we obtain for a quasi-Monte Carlo algorithm

$$Q_{n,d}(f) = \frac{1}{n} \sum_{i=1}^n f(z_i), \quad z_1, \ldots, z_n \in [0,1]^d,$$

the identity

$$e^{\mathrm{wor}}(Q_{n,d}) = \sup_{\|f\|_{F_{d,\gamma}} \le 1} |I_{d,\gamma}(f) - Q_{n,d}(f)| = d^c_{2,\gamma}(z_1, \dots, z_n), \tag{1.15}$$

where $d^c_{2,\gamma}(z_1, \dots, z_n)$ is the *weighted L_2-centered discrepancy* given by

$$d^c_{2,\gamma}(z_1, \dots, z_n) = \left(\sum_{\emptyset \neq u \subseteq \{1,\dots,d\}} \prod_{j \in u} \gamma_{d,\gamma} \int_{[0,1]^{|u|}} |d^c(z_1, \dots, z_n; (x_u, 1/2))|^2 \, dx_u \right)^{1/2}, \tag{1.16}$$

and

$$d^c(z_1, \dots, z_n; (x_u, 1/2)) = \mathrm{vol}(J(x_u)) - \frac{1}{n} \sum_{i=1}^{n} 1_{J(x_u)}((z_i)_u).$$

F. J. Hickernell introduced the L_2-centered discrepancy and proved (1.15) in the case where all weights are equal to 1 [Hic98]. Identity (1.15) shows that proving bounds on the geometrically motivated quantity $d^c_{2,\gamma}(z_1, \dots, z_n)$, which does not depend on functions from $F_{d,\gamma}$ any more, is equivalent to proving bounds on the (rather abstract) quantity $e^{\mathrm{wor}}(Q_{n,d})$. In general one can deduce in reproducing kernel Hilbert spaces identities of the form (1.15) in a rather canonical way, see, e.g., [Hic98, Sect. 1], [Mat99, Chap. 1] or [NoW09, Sect. 5, 6]. We will discuss discrepancy theory in more detail in Section 1.4.

1.3.2 Sparse Grid Methods on Multiwavelet Spaces

In the article [GLSS07] multivariate integration on certain Hilbert spaces $\mathcal{H}^d_{s,n}$ of functions on $[0,1]^d$ is considered. These spaces are spanned by *multiwavelets* $(\psi_\tau)_\tau$ and have a discrete norm

$$\|f\|^2_{d,s,n} = \sum_{\tau} 2^{|\tau| s} \langle f, \psi_\tau \rangle^2,$$

where $\langle \cdot, \cdot \rangle$ is the usual scalar product on $L_2([0,1]^d)$. In dimension $d = 1$ the multiwavelets consist of piecewise polynomials of degree up to $n-1$, in higher dimension they are tensor products of piecewise polynomials. If $s < n$, then the Sobolev space $H^s_{d,mix}$, which is the d-fold tensor product of the usual Sobolev space $H^s([0,1])$, is continuously embedded in $\mathcal{H}^d_{s,n}$; this can be proved with the help of Jackson-type inequalities. Furthermore, $\mathcal{H}^d_{s,n}$ also contains discontinuous functions.

The d-dimensional cubature rules proposed in [GLSS07] are based on one-dimensional composite quadrature rules chosen with respect to the space $\mathcal{H}^1_{s,n}$ of univariate functions and on *Smolyak's construction*, which results in a d-dimensional cubature rule on the tensor space $\mathcal{H}^d_{s,n} = \otimes^d_{i=1} \mathcal{H}^1_{s,n}$. The weights of such rules are not uniform like quasi-Monte Carlo weights and can actually take negative values. The general approach to build d-dimensional cubature formulas from one-dimensional quadratures via Smolyak's construction is well known and also referred to as a *sparse grid method*, see, e.g., [BuG04,

GeG98, NoR96, Smo63, WaW95]. Smolyak's method is not limited to multivariate integration, but can more universally be used to approximate solutions of tensor product problems, see, e.g., [BuG04, WaW95].

For $s > 1/2$ Theorem. 4.4 and Cor. 4.7 from [GLSS07] show that the worst case error of the proposed d-dimensional cubature rules is bounded by

$$O\left(\frac{(\ln N)^{(d-1)(s+1/2)}}{N^s}\right), \tag{1.17}$$

where N is the number of integration points used. This bound is smaller by a factor of order $(\ln N)^{-(d-1)/2}$ than the usual standard estimate for Smolyak's construction in the worst case setting, cf., e.g., [GeG98, NoR96, WaW95]. The proof of Theorem 4.4 exploits the locality of the multiwavelets spanning $\mathcal{H}^d_{s,n}$. It does not provide a reasonable estimate for the constant implicit in the O-notation in (1.17). To get an upper bound on the worst case error whose dependence on the dimension d is explicitly given, we adapted the proof of [WaW95, Lemma 2] and provide in Theorem 4.2 a bound which is worse than (1.17) by a factor of order $(\ln N)^{(d-1)/2}$, but exhibits in exchange concrete constants.

In [GLSS07, Thm. 4.9] the lower bound

$$\Omega\left(\frac{(\ln N)^{(d-1)/2}}{N^s}\right) \tag{1.18}$$

for the worst case error of arbitrary cubature rules is proved. This result extends an earlier result of S. Heinrich, F. J. Hickernell, and R.-X. Yue on Haar wavelet spaces; in fact it is not hard to see that their space $\mathcal{H}_{\mathrm{wav},s}$ coincide with the spaces $\mathcal{H}^d_{s,1}$ for base $b = 2$. Indeed our article was motivated by the article [HHY04]. In the following sense one may see our approach as an extension of [HHY04]: As already said, our spaces $\mathcal{H}^d_{s,n}$ include the Haar wavelet spaces considered in [HHY04]. For $s \geq 3/2$ these Haar wavelet spaces do not even contain the linear function $f(x) = x$, while our spaces $\mathcal{H}^d_{s,n}$ contain smooth functions as long as $n > s$. The cubature rules used in [HHY04] to establish an upper bound on the worst case error of multivariate integration on Haar wavelet spaces are so-called *scrambled nets* (see, e.g., [Mat99, Owe97]), which are cleverly randomized quasi-Monte Carlo rules. Scrambled nets are exact on finite Haar wavelet series up to a critical level, but not on finite multiwavelet series for $n > 1$, in contrast to our cubature rules.

Nevertheless, one has to say clearly that the scrambled nets considered in [HHY04] attain the optimal order of convergence for numerical integration on $\mathcal{H}^d_{s,1}$, while for our method we only prove almost optimal convergence (that is, optimal up to logarithmic factors) on $\mathcal{H}^d_{s,n}$ for arbitrary $n \in \mathbb{N}$. Furthermore, in [HHY04] not only the worst case error, but also the random case and average case errors were studied.

In [GLSS07, Sect. 5] we report on several numerical tests which allow us to compare our sparse grid method with other known methods.

1.3.3 Conclusion and Open Problems

In [GnW08] we studied generalized tractability of linear functionals on reproducing kernel Hilbert spaces. We presented necessary conditions for all kinds of tractability , and in the particular case of multivariate integration also sufficient conditions.

Obviously, there are many other interesting settings which can be studied. In the setting of standard information it would, e.g., be interesting to study generalized tractability on function spaces that are not reproducing kernel Hilbert spaces, to derive sufficient conditions for generalized tractability of general continuous linear functionals or even for other linear tensor product problems.

Notice that the upper bound (1.12) was proved by averaging over the square of the worst case error of all QMC algorithms. Thus we have not answered the question which explicit QMC algorithms (i.e., which deterministic sets of integration points) imply (1.12). The answer to this question depends of course heavily on the particular space F_d under consideration and is indeed not trivial, see [NoW09].

As indicated above, for integration points the property to induce a small QMC worst case error is often equivalent to the property of having a small corresponding geometric discrepancy measure. In Subsection 1.4.2 we will study the construction of integration points with small L_∞*-star discrepancy*. The L_∞-star discrepancy is related to the worst case error of multivariate integration on some L_1-Sobolev space, see Theorem 1.4.1, which is a setting different from multivariate integration on Hilbert spaces.

In [GLSS07] we presented for fixed d asymptotically almost optimal algorithms for multivariate integration on multiwavelet spaces. Looking at [Dic07a, Dic08a, GLSS07, HHY04] it seems to be promising to make further investigations to exploit good properties of certain function spaces and their bases to find good cubature rules.

1.4 Geometric Discrepancy

Geometric discrepancy theory studies the uniformity of distribution of finite point sets.

The notion of discrepancy was introduced by Bergström [Ber36] and van der Corput [Cor35a, Cor35b], but already before contributions to the theory of uniform distribution had been made, including the fundamental work of Weyl [Wey16].

Although the focus of this section is mainly on the so-called star discrepancy and the extreme discrepancy, we want to present here briefly a rather general framework to define discrepancy measures. Quantitative measures for the deviation of the distribution of a finite (randomly drawn) subset of a probability space (the "empirical distribution") from the given probability distribution can be established in the following way:

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathcal{A}$ a subset of the σ-algebra $\Sigma = \Sigma(\Omega)$. For a finite subset T of Ω and a measurable set $A \in \mathcal{A}$ we define the *discrepancy function* $\mathrm{disc}(T, A, \mathbb{P})$ by

$$\mathrm{disc}(T, A, \mathbb{P}) = \left| \mathbb{P}(A) - \frac{|T \cap A|}{|T|} \right|$$

and the L_∞-*discrepancy* $\operatorname{disc}_\infty(T, \mathcal{A}, \mathbb{P})$ of T with respect to $\mathcal{A}$ and $\mathbb{P}$ by

$$\operatorname{disc}_\infty(T, \mathcal{A}, \mathbb{P}) = \sup_{A \in \mathcal{A}} \operatorname{disc}(T, A, \mathbb{P}).$$

Let $\Sigma(\mathcal{A})$ be a σ-algebra on $\mathcal{A}$ and $\mu : \Sigma(\mathcal{A}) \mapsto [0, +\infty)$ be a finite positive measure. If the function $\operatorname{disc}(T, \cdot, \mathbb{P})$ is measurable with respect to $\Sigma(\mathcal{A})$, we may define the L_p-*discrepancy* of T with respect to $\mathcal{A}$, $\mathbb{P}$, and μ by

$$\operatorname{disc}_p(T, \mathcal{A}, \mathbb{P}) = \left(\int_{\mathcal{A}} \operatorname{disc}(T, A, \mathbb{P})^p \, \mu(\mathrm{d}A) \right)^{1/p}.$$

It seems natural to normalize the measure μ such that it is a probability measure, since then the L_p-discrepancy can be seen as a kind of average value of the discrepancy function and converges (under suitable conditions on μ) to the L_∞-discrepancy in the limit $p \to \infty$. (Such a convergence behavior was, e.g., exploited in [HNWW01, Gne05]).

Here we want to confine ourselves to the case where Ω is the d-dimensional unit cube $[0,1]^d \subset \mathbb{R}^d$, Σ is the σ-algebra of Borel sets of $[0,1]^d$, and $\mathbb{P}$ is the d-dimensional Lebesgue measure λ_d restricted to $[0,1]^d$. The set systems we consider are the system $\mathcal{C}_d$ of all d-dimensional "corners"

$$\mathcal{C}_d := \{[0, x) \mid x \in [0,1]^d\}$$

and the system $\mathcal{R}_d$ of all d-dimensional "rectangles"

$$\mathcal{R}_d := \{[x, y) \mid x, y \in [0,1]^d,\ x \leq y\}$$

in $[0,1]^d$; here we denote for given vectors $x, y \in [0,1]^d$ the d-dimensional half-open box $[x_1, y_1) \times \cdots \times [x_d, y_d)$ by $[x, y)$, and the inequality $x \leq y$ is meant componentwise.

The resulting geometric discrepancies for sets $T \subset [0,1]^d$ are the *(L_∞-)star discrepancy* defined by

$$d^*(T, C) := \operatorname{disc}(T, C, \lambda_d) \quad \text{for } C \in \mathcal{C}_d$$

and

$$d^*_\infty(T) := \operatorname{disc}_\infty(T, \mathcal{C}_d, \lambda_d),$$

and the *(L_∞-)extreme discrepancy* defined by

$$d^e(T, R) := \operatorname{disc}(T, R, \lambda_d) \quad \text{for } R \in \mathcal{R}_d$$

and

$$d^e_\infty(T) := \operatorname{disc}_\infty(T, \mathcal{R}_d, \lambda_d).$$

The corresponding L_p-discrepancies for $1 \leq p < \infty$ can be established by

$$d^*_p(T) := \left(\int_{[0,1]^d} d^*_\infty(T, [0, x))^p \, \lambda_d(\mathrm{d}x) \right)^{1/p}$$

and, with $\widetilde{\mathcal{R}_d} := \{(x,y) \in [0,1]^{2d} \,|\, x \le y\}$,

$$d_p^e(T) := \left(\int_{\widetilde{R}_d} d_\infty^e(T,[x,y))^p \, 2^d \lambda_{2d}(\mathrm{d}x \otimes \mathrm{d}y) \right)^{1/p}.$$

Notice that here we identified boxes $[0,x) \in \mathcal{C}_d$ with their upper right corners x; via this identification we in principle endowed $\mathcal{C}_d$ with the σ-algebra of Borel sets and the probability measure λ_d on $[0,1]^d$. Furthermore, we considered the set $\widetilde{\mathcal{R}_d}$, endowed with the σ-algebra of Borel sets and the probability measure $\mu_d := 2^d \lambda_{2d}$. This time the map of identification $\Phi : \widetilde{\mathcal{R}_d} \to \mathcal{R}_d$, $(x,y) \mapsto [x,y)$ is surjective, but not injective, since $\Phi((x,x)) = \emptyset$ for all $x \in [0,1]^d$. But this is insignificant, as the set $\{(x,x) \in [0,1]^{2d}\}$ has μ_d-measure zero.

Further quantities of interest are for $1 \le p \le \infty$ the smallest possible L_p-star discrepancy of any n-point set in d dimensions

$$d_p^*(n,d) := \inf_{T \subset [0,1]^d,\, |T|=n} d_p^*(T)$$

and for given $\varepsilon > 0$

$$n_p^*(\varepsilon,d) := \min\{n \in \mathbb{N} \,|\, d_p^*(n,d) \le \varepsilon\},$$

the so-called *inverse of the L_p-star discrepancy*. The corresponding quantities $d_p^e(n,d)$ and $n_p^e(\varepsilon,d)$ for the extreme discrepancy are defined analogously.

The articles on discrepancy collected here focus on the L_∞-star discrepancy and discuss also related results for the L_∞-extreme discrepancy. The exception is the article [Gne05], where we additionally provide upper bounds for the average L_p-extreme and the average L_p-star discrepancy for $2 \le p < \infty$; here the average is taken over all n-point sets in $[0,1]^d$ with respect to the measure λ_{nd}.

The notion of discrepancy is intimately related to multivariate numerical integration. This relation is put in a quantitative form by inequalities of Koksma-Hlawka- or Zaremba-type. One example is identity (1.15), which can be seen as a sharp version of the inequality

$$|I_{d,\gamma}(f) - Q_{n,d}(f)| \le \|f\|_{F_{d,\gamma}} d_{2,\gamma}^c(z_1,\ldots,z_n) \quad \text{for all } f \in F_{d,\gamma}.$$

As a further example we state a sharp version of the classical Koksma-Hlawka inequality [Kok42, Hla61], which relates the star discrepancy to the worst-case error of quasi-Monte Carlo integration on certain function spaces. To this purpose, let us first define the normed function spaces we want to consider:

Let $H^{1,1}$ be the space of absolutely continuous functions f on $[0,1]$ whose derivatives f' are again integrable. A norm on $H^{1,1}$ is given by $\|f\|_{1,1} := |f(1)| + \|f'\|_{L_1([0,1])}$. The (algebraic) tensor product $\otimes_{i=1}^d H^{1,1}$ consists of linear combinations of functions f of product form $f(x) = f_1(x_1) \ldots f_d(x_d)$, $f_1,\ldots,f_d \in H^{1,1}$. The space $H^{1,d}$ is then defined as the closure of $\otimes_{i=1}^d H^{1,1}$ with respect to the norm

$$\|f\|_{1,d} := |f(\mathbf{1})| + \sum_{\emptyset \ne u \subseteq \{1,\ldots,d\}} \|f_u'\|_{L_1([0,1]^{|u|})},$$

where $\mathbf{1}$ denotes the vector $(1,\dots,1)$ and f_u' is defined by

$$f_u'(x_u) = \frac{\partial^{|u|}}{\prod_{k\in u}\partial x_k} f(x_u,\mathbf{1}), \quad \text{with} \quad (x_u,\mathbf{1})_k = \begin{cases} x_k & \text{if } k\in u, \\ 1 & \text{else.} \end{cases}$$

Then the following theorem holds:

Theorem 1.4.1. *Let $t^1,\dots,t^n \in [0,1)^d$, and let I_d be the integral operator and $Q_{d,n}$ be the quasi-Monte Carlo cubature defined by*

$$I_d(f) = \int_{[0,1]^d} f(t)\,\mathrm{d}t \quad \textit{and} \quad Q_{d,n}(f) = \frac{1}{n}\sum_{i=1}^{n} f(t^i).$$

Then

$$e^{\mathrm{wor}}(Q_{n,d}) = \sup_{f\in H^{1,d};\,\|f\|_{1,d}=1} |I_d(f) - Q_{d,n}(f)| = d_\infty^*(t^1,\dots,t^n). \tag{1.19}$$

In particular, we obtain

$$|I_d(f) - Q_{d,n}(f)| \le \|f\|_{1,d}\, d_\infty^*(t^1,\dots,t^n) \quad \textit{for all } f\in H^{1,d}. \tag{1.20}$$

The theorem above is a corollary of a more general theorem proved by F. J. Hickernell, I. H. Sloan, and G. W. Wasilkowski in [HSW04]. There in particular the so-called L_∞-*same-quadrant discrepancy*, which covers as special cases the star discrepancy as well as the L_∞-centered discrepancy, is related to the worst-case error of quasi-Monte Carlo approximation of multivariate integrals on *anchored* L_1-Sobolev spaces. In the special case of the star discrepancy the anchor is the point $\mathbf{1}$, in the case of the L_∞-centered discrepancy the anchor is $\frac{1}{2}\mathbf{1}$. A recent paper dealing with various L_2-discrepancies (including the L_2-same quadrant discrepancy) and their correspondence to multivariate integration on distinguished reproducing kernel Hilbert spaces is [NoW09]. Further results on the relation of different notions of discrepancy to multivariate integration can be found, e.g., in the original papers [Dic08b, Hic98, HoK96, JHK97, KlT97, Nie72b, Pas93, Pro88, SlW98, Woź91, Zar68] and the survey article [NoW01a], as well as in the monograph [Nie92].

Particularly with regard to Theorem 1.4.1 and the various other results connecting different notions of discrepancy with multivariate numerical integration, the following three questions seem to be very important.

Question 1.4.2. (i) Can we find good bounds for the smallest possible discrepancy of any n-point set for moderate n?

(ii) How can we construct point sets efficiently that satisfy such bounds?

(iii) How can we calculate the discrepancy of given point sets efficiently?

Here we want to discuss these questions for the (L_∞-)star discrepancy. So if we intent to approximate high-dimensional integrals of functions from $H^{1,d}$ by quasi-Monte Carlo

cubatures, and if we wish to minimize the corresponding worst-case error, then Theorem 1.4.1 tells us that we have to minimize the star discrepancy of the set of integration points we want to use. To this purpose it is certainly helpful to have upper bounds for the smallest star discrepancy that we can achieve with n points. In high dimensions cubatures whose number of integration points n are exponential in the dimension are not feasible. That is why we ask in question (i) for good bounds for the smallest possible discrepancy of sample sets of moderate size n. By "moderate" we mean that n grows not stronger in the dimension d than a polynomial of small degree.

Bounds for the smallest discrepancy achievable are certainly useful, but for quasi-Monte Carlo integration we need to have explicit integration points. Therefore question (ii) is essential.

In practice we may have some point sets that are reasonable candidates to use for quasi-Monte Carlo integration. This may be due to several reasons as, e.g., that in those points we can easily evaluate the functions we want to integrate or that those points are in some sense uniformly distributed. Therefore it would be desirable to be able to calculate the star discrepancy of a given set efficiently.

In fact question (iii) is directly related to question (ii) by the concentration of measure phenomenon, as, e.g., described in [Gne07a]:

Let us assume that we have a class of n-point sets endowed with some probability measure and the expected discrepancy of a random set is small enough for our needs. Under suitable conditions the measure of the discrepancy distribution is sharply concentrated around the expected discrepancy and a large deviation bound ensures that a randomly chosen set has a sufficiently small discrepancy with high probability. In this situation we may consider the following random algorithm, which is a *semi-construction* in the sense of E. Novak and H. Woźniakowski [NoW08]:

We choose a point set randomly and calculate its actual discrepancy. If it serves our needs, we accept the point set and stop; otherwise we make a new random choice. The large deviation bound guarantees that with very high probability we only have to perform a few random trials to receive an acceptable point set.

Apart from the practical problem of choosing the point set according to the law induced by the probability measure, we have to think of ways to calculate the discrepancy of a chosen set efficiently.

Although there are of course many other interesting facets of discrepancy theory related to numerical integration, as, e.g., weighted discrepancies (see (1.16) or [DSWW04, DSWW06, GnW08, HPS08, Joe06, NoW01a, NoW01b, NoW09, SiJ07, SlW98]) or higher order convergence (see, e.g., [Dic07a, Dic08a, DiP07, HHY04, Hic02, LaT94]), we confine ourselves in this section to the discussion of Question 1.4.2 for the unweighted star and extreme discrepancy.

1.4.1 Bounds for the Star Discrepancy with Explicit Constants

Low-Discrepancy Point Sets

Let us discuss Question 1.4.2(i) in detail. For fixed dimension d the asymptotically best upper bounds for $d^*_\infty(n,d)$ that have been proved so far are of the form

$$d^*_\infty(n,d) \le C_d \ln(n)^{d-1} n^{-1}\,, \quad n \ge 2\,, \tag{1.21}$$

see, e.g., [BeC87, Cha00, DrT97, Mat99, Nie92]. These bounds have been proved constructively, i.e., there are explicit constructions known that satisfy (1.21) for suitable constants C_d.

For $d = 1$ the set $T = \{1/2n, 3/2n, \ldots, (2n-1)/2n\}$ establishes (1.21) with $C_1 = 1/2$. For $d = 2$ the bound (1.21) can be derived from the results of Hardy and Littlewood [HaL22] and of Ostrowski [Ost22a, Ost22b]. For $d \ge 3$ the bound (1.21) was established by Halton, who showed in [Hal60] that the famous Halton-Hammersley points exhibit this asymptotic behavior. The Halton-Hammersley points can be seen as a generalization of the two-dimensional point sets obtained in a canonical way from the one-dimensional infinite sequence of van der Corput from [Cor35a, Cor35b]. (In general, if one has an infinite $(d-1)$-dimensional low-discrepancy sequence $(t^k)_{k\in\mathbb{N}}$, one canonically gets a d-dimensional low-discrepancy point set $\{p^1, \ldots, p^n\}$ for every n by putting $p^k = (k/n, t^k)$, see also [Mat99, Sect. 1.1, 2.1].)

The asymptotic bound (1.21) is widely believed to be sharp. That it is optimal up to logarithmic factors is clear from the trivial lower bound $1/2n$. A better lower bound was shown by K. Roth in [Rot54]:

$$d^*_\infty(n,d) \ge c_d \ln(n)^{\frac{d-1}{2}} n^{-1}\,, \quad n \ge 2\,. \tag{1.22}$$

In fact, Roth proved that the right hand side of (1.22) is a lower bound for the smallest possible L_2-star discrepancy $d^*_2(n,d)$, and this bound is best possible as was shown for $d = 2$ by Davenport [Dav56], and for $d \ge 3$ by Roth himself [Rot79, Rot80] and independently by Frolov [Fro80]. Although Roth's lower bound is sharp for the L_2-star discrepancy, one may suspect that it is not optimal for the L_∞-star discrepancy. That this is indeed not the case was proved by W. M. Schmidt in [Sch72]. He established in dimension $d = 2$ the lower bound

$$d^*_\infty(n,2) \ge c_2 \ln(n) n^{-1}\,, \quad n \ge 2\,, \tag{1.23}$$

and showed in this way that the upper bound (1.21) is optimal in dimension 2. In dimension $d \ge 3$ improvements of (1.22) were achieved by J. Beck [Bec89], and by D. Bilyk, M. T. Lacey, and A. Vagharshakyan [BiL08, BLV08]; but although those improvements are certainly deep mathematical results, their quantitative gain is rather modest. The remaining gap, baptized the "great open problem" by J. Beck and W. W. L. Chen in [BeC87], has still not been bridged so far.

Nonetheless, the solution of this intricate problem is not overly significant for numerical integration in high dimensions. In particular, bounds of the form (1.21) give us no helpful information for moderate values of n, since $\ln(n)^{d-1} n^{-1}$ is an increasing function for

$n \leq e^{d-1}$. This means that with respect to d we have to use at least exponentially many integration points to perceive at least some decay of the right hand side of inequality (1.21). It is instructive to compare the convergence rate $n^{-1}\ln(n)^{d-1}$ and the Monte Carlo convergence rate $n^{-1/2}$. For $d = 10$ we have $n^{-1}\ln(n)^{d-1} > n^{-1/2}$ for all $n \leq 1.295 \cdot 10^{34}$. Additionally, point configurations satisfying (1.21) will in general lead to constants C_d that depend critically on d. (Actually, it is known for some constructions that the constant C'_d in the representation

$$d^*_\infty(n,d) \leq \big(C'_d \ln(n)^{d-1} + o(\ln(n)^{d-1})\big)\, n^{-1}$$

of (1.21) tends to zero as d approaches infinity, see, e.g., [Ata04, Nie92, NiX96]. But seemingly no good bounds have been published so far for the implicit constant of the o-notation or, respectively, the constant C_d in (1.21).)

A bound more suitable for high-dimensional integration was established by S. Heinrich, E. Novak, G. W. Wasilkowski and H. Woźniakowski [HNWW01], who proved

$$d^*_\infty(n,d) \leq c d^{1/2} n^{-1/2} \quad \text{and} \quad n^*_\infty(d,\varepsilon) \leq \lceil c^2 d \varepsilon^{-2} \rceil\,, \tag{1.24}$$

where c does not depend on d, n or ε. Here the dependence of the inverse of the star discrepancy on d is optimal. This was also established in [HNWW01] by a lower bound for $n^*_\infty(d,\varepsilon)$, which was later improved by A. Hinrichs [Hin04] to

$$n^*_\infty(d,\varepsilon) \geq c_0 d \varepsilon^{-1} \quad \text{for } 0 < \varepsilon < \varepsilon_0, \tag{1.25}$$

where $c_0, \varepsilon_0 > 0$ are constants. The proof of (1.24) uses a large deviation bound of M. Talagrand for empirical processes [Tal94] and an upper bound of D. Haussler for covering numbers of Vapnik-Červonenkis classes [Hau95]. In particular, the proof is not constructive but probabilistic, and the proof approach does not provide an estimate for the value of c. (A. Hinrichs presented a more direct approach to prove (1.24) with $c = 10$ at the Dagstuhl Seminar 04401 "Algorithms and Complexity for Continuous Problems" in 2004.) As pointed out in [HNWW01], the bounds in (1.24) hold also for the extreme discrepancy and its inverse (probably with a larger constant c).

In the paper [HNWW01] the authors proved also two slightly weaker bounds with explicitly known constants: The first one is of the form

$$d^*_\infty(n,d) \leq k d^{1/2} n^{-1/2} \big(\ln(d) + \ln(n)\big)^{1/2}\,, \tag{1.26}$$

and its proof is again probabilistic and relies on Hoeffding's large deviation bound. (A similar probabilistic approach was already used by Beck in [Bec84] to prove upper bounds for other types of geometric discrepancies.) For the sake of explicit constants the proof technique has been adapted in subsequent papers on high-dimensional integration of certain function classes [HSW04, Mha04]. Inequality (1.26) leads to

$$n^*_\infty(d,\varepsilon) \leq O(d\varepsilon^{-2}(\ln(d) + \ln(\varepsilon^{-1}))) \tag{1.27}$$

where the implicit constant in the big-O-notation is known and independent of d and ε.

The second bound was shown in the following way: For $p \in [1,\infty)$ let $\mathrm{av}_p^*(n,d)$ denote the average L_p-star discrepancy, given by

$$\mathrm{av}_p^*(n,d) = \left(\int_{[0,1]^{nd}} d_p^*(T)^p \, \mathrm{d}T \right)^{1/p}, \quad T = (t^1, \ldots, t^n).$$

Heinrich et al. proved for even $p = 2kd$, $k \in \mathbb{N}$, and $n \geq 2(1+kd)^2$ an upper bound for the average L_p-star discrepancy of the form

$$\mathrm{av}_p^*(n,d) \leq 3^{2/3} 2^{5/2+d/p} p (p+2)^{-d/p} n^{-1/2}\,.$$

The detailed analysis is quite elaborate, since it uses a representation of $\mathrm{av}_p^*(n,d)$ as an alternating sum of weighted products of Stirling numbers of the first and second kind. The bound was used to derive upper bounds

$$n_\infty^*(\varepsilon,d) \leq C_k d^2 \varepsilon^{-2-1/k} \quad \text{for } \varepsilon \in (0,1/2) \text{ and } k \in \mathbb{N}, \tag{1.28}$$

where the constant C_k depends solely on k. The dependence of bound (1.28) on the dimension d is obviously worse than the dependence of bound (1.27). Nevertheless, in all comparisons reported in [HNWW01] for explicit values of d and ε, the approach leading to bound (1.28) gave better results than the one leading to (1.27). To improve the dependence of bound (1.28) on d, A. Hinrichs suggested to use *symmetrization*. This approach was sketched in [Nov04] and leads for $p = 2kd$, $k \in \mathbb{N}$, and $\varepsilon \in (0,1/2)$ to

$$\mathrm{av}_p^*(n,d) \leq 2^{1/2+d/p} p^{1/2} (p+2)^{-d/p} n^{-1/2} \quad \text{and} \quad n_\infty^*(\varepsilon,d) \leq C_k' d \varepsilon^{-2-1/k}.$$

(Actually there is an error in the calculations in [Nov04], therefore we stated the results of our own calculations presented in [Gne05, Remark 4 and 9]).

In the paper [Gne05] a symmetrization approach is used to prove the upper bound

$$\mathrm{av}_p^e(n,d) \leq 2^{1/2+3d/p} p^{1/2} (p+2)^{-d/p} (p+4)^{-d/p} n^{-1/2}$$

for the average L_p-extreme discrepancy $\mathrm{av}_p^e(n,d)$ for even integers $p \geq 4d$. The analysis is simpler as the one in [HNWW01], since it does not need Stirling numbers and uses rather simple combinatorial arguments. Similar as in [HNWW01, Nov04], from this bound upper bounds for the inverse of the L_∞-extreme discrepancy of the form

$$n_\infty^e(\varepsilon,d) \leq C_k'' d \varepsilon^{-2-1/k} \quad \text{for } \varepsilon \in (0,1/2) \text{ and } k \in \mathbb{N}$$

are derived, where the constant C_k'' depends only on k.

Nevertheless, since the bound (1.27) shows a better behavior then (1.28) one might be tempted to improve the former one such that it leads in particular to better results for explicit values of d and ε than the approach leading to (1.28).

So let us look at the proof approach for (1.26) and (1.27) from a conceptional point of view. The first step is to "discretize" the star discrepancy at the cost of an approximation error at most δ. This discretization is done by choosing a suitable finite set of test boxes anchored in zero whose upper right corners form a so-called δ-cover. We repeat here the definition from [Gne04, DGS05].

Definition 1.4.3. A finite subset Γ of $[0,1]^d$ is called a *δ-cover* of $\mathcal{C}_d$ (or of $[0,1]^d$) if for all $y \in [0,1]^d$ there exist $x, z \in \Gamma \setminus \{0\}$ such that $x \le y \le z$ and $\lambda_d([0,z)) - \lambda_d([0,x)) \le \delta$. Put $N(\mathcal{C}_d, \delta) := \min\{|\Gamma| \mid \Gamma \ \delta\text{-cover of } \mathcal{C}_d\}$.

Any δ-cover Γ of $\mathcal{C}_d$ satisfies the following approximation property:

Lemma 1.4.4. *Let Γ be a δ-cover of $\mathcal{C}_d$. For all finite subsets P of $[0,1]^d$ we have*

$$d^*_\infty(P) \le d^*_\Gamma(P) + \delta, \tag{1.29}$$

where

$$d^*_\Gamma(P) := \max_{x\in\Gamma} \left| \lambda_d([0,x)) - \frac{|P \cap [0,x)|}{|P|} \right|$$

can be seen as a discretized version of the star discrepancy.

A proof of Lemma 1.4.4 is, e.g., contained in [DGS05].

The essence of the approach is now that a random set P has star discrepancy at most ε if $d^*_\Gamma(P)$ is at most $\delta := \varepsilon/2$. Now the event $\{d^*_\Gamma(P) > \delta\}$ is the union of finitely many events, namely those that a test box $[0,x)$, $x \in \Gamma$, has discrepancy $d^*(P,[0,x))$ larger than δ. The probability of such a single event can easily be controlled by Hoeffding's large deviation bound, and the probability of the union of those events by a simple union bound. To keep the loss caused by the union bound as small as possible, one should choose the δ-cover as small as possible. If one wants to receive a bound for the star discrepancy with explicit constants, then one needs bounds for the size of the δ-covers used (or, more generally, for $N(d,\delta)$) with explicit constants. These tasks seem to be interesting problems on their own, apart from the application to the star discrepancy.

In [HNWW01] a δ-cover in form of an equidistant grid with cardinality $O(d^d\delta^{-d})$ was used. Constructions of smaller δ-covers were first reported in [Gne04] and are discussed in full detail in [DGS05]. There a simple construction of a δ-cover in form of a non-equidistant grid Γ with cardinality $O(\ln(d)^d\delta^{-d})$ is given. This construction has the advantage that the resulting set of test boxes $[0,x)$, $x \in \Gamma$, is still stable under intersections, a fact that is essential for the paper [DoG08], which will be discussed in Sect. 1.4.2.

In [DGS05] also the asymptotic results

$$d^{1/2}\mathrm{e}^{-d}\delta^{-d} + O_d(|\ln(\delta)|^{d-1}) \le N(\mathcal{C}_d,\delta) \le \sqrt{\frac{2}{\pi d}}\,\mathrm{e}^d\delta^{-d} + O_d(\delta^{-d+1}) \tag{1.30}$$

are provided, where the implicit constants in the O_d-notation may depend on d, but not on δ. The upper bound on $N(\mathcal{C}_d,\delta)$ in (1.30) is again proved constructively. A drawback is that—written down with explicit constant—this asymptotically best upper bound from [DGS05] has the form

$$N(\mathcal{C}_d,\delta) \le \sqrt{\frac{2}{\pi d}}\,\mathrm{e}^d \left(\delta^{-1} + \frac{d+1}{4}\right)^d \tag{1.31}$$

and thus shows unfortunately a super-exponential dependence on d, see [DGS05, Thm. 2.7]. Nevertheless, with these results on δ-covers (1.26) can be improved to

$$d^*_\infty(n,d) \le k' d^{1/2} n^{-1/2} \ln(n)^{1/2}\,, \tag{1.32}$$

where k' is smaller than k [DGS05, Thm. 3.2]. (Essentially we have $k \simeq 2\sqrt{2}$ and $k' \simeq \sqrt{2}$). The corresponding estimate for the inverse of the star discrepancy gives for concrete values of d and ε results better than the one stated in [HNWW01]; in particular they are better than the ones resulting from the average L_p-star discrepancy analysis in [HNWW01].

Furthermore, in [DGS05, Remark 3.5] a simple observation is stated, which leads immediately to improved constants in certain bounds derived in [HSW04] and [Mha04].

In [Gne08a] the notion of δ-covers was related to the concept of bracketing entropy, which is well known in the theory of empirical processes. We state here the definition for the set system $\mathcal{C}_d$ of anchored axis-parallel boxes (a general definition can, e.g., be found in [Gne08a, Sect. 1]):

Definition 1.4.5. A closed axis-parallel box $[x, z] \subset [0,1]^d$ is a *δ-bracket* of $\mathcal{C}_d$ if $x \leq z$ and $\lambda_d([0,z)) - \lambda_d([0,x)) \leq \delta$. A *$\delta$-bracketing cover* of $\mathcal{C}_d$ is a set of δ-brackets whose union is $[0,1]^d$. By $N_{[]}(\mathcal{C}_d, \delta)$ we denote the *bracketing number* of $\mathcal{C}_d$, i.e., the smallest number of δ-brackets whose union is $[0,1]^d$. The quantity $\ln N_{[]}(\mathcal{C}_d, \delta)$ is called the *bracketing entropy*.

It is easy to see that

$$N(\mathcal{C}_d, \delta) \leq 2N_{[]}(\mathcal{C}_d, \delta) \leq N(\mathcal{C}_d, \delta)(N(\mathcal{C}_d, \delta) + 1) \tag{1.33}$$

holds. In [Gne08a] it is shown that

$$\delta^{-d}(1 - O_d(\delta)) \leq N_{[]}(\mathcal{C}_d, \delta) \leq 2^{d-1}(2\pi d)^{-1/2}\mathrm{e}^d(\delta^{-1} + 1)^d, \tag{1.34}$$

see [Gne08a, Thm. 1.5 and 1.15]. The construction that leads to the upper bound in (1.34) implies also

$$N_{[]}(\mathcal{C}_d, \delta) \leq (2\pi d)^{-1/2}\mathrm{e}^d\delta^{-d} + O_d(\delta^{-d+1}) \tag{1.35}$$

(see [Gne08a, Remark 1.16]) and

$$N(\mathcal{C}_d, \delta) \leq 2^d(2\pi d)^{-1/2}\mathrm{e}^d(\delta^{-1} + 1)^d, \tag{1.36}$$

a bound which does not exhibit a super-exponential dependence on d, in contrast to (1.31).

From the estimate (1.36) one obtains the discrepancy bound

$$d^*_\infty(n, d) \leq k'' d^{1/2} n^{-1/2} \ln(n/d)^{1/2} \tag{1.37}$$

for $d \lesssim n$, which improves upon (1.32) (here we have again $k'' \simeq \sqrt{2}$, see [Gne08a, Thm. 2.1]). Since the inverse of the star discrepancy depends linearly on the dimension d, the practically most relevant choice of n seems to be n proportional to d. Note that in this case (1.37) behaves asymptotically as the bound (1.24). In fact, if (1.24) holds with $c = 10$ (as claimed by A. Hinrichs), then the bound [Gne08a, (22)], a version of (1.37), is better than (1.24) for all $n \leq 1.5 \cdot \mathrm{e}^{95} d$. Actually, we may use the upper bound in (1.34) to reprove (1.24) without using Haussler's result on covering numbers of Vapnik-Červonenkis classes—a version of Talagrand's large deviation bound for empirical

processes holds under the condition that the δ-bracketing number of the set system under consideration is bounded from above by $(C\delta^{-1})^d$ for some constant C not depending on δ or d, see [Tal94, Thm. 1.1].

In [Gne08a] furthermore the notions of δ-covers and δ-bracketing covers of the set system $\mathcal{R}_d$ are considered and related, and bounds similar to (1.34) are derived [Gne08a, Thm. 1.10, Lemma 1.18]. The corresponding constructive upper bound leads to a bound for the extreme discrepancy similar to (1.37) (roughly d has to be substituted by $2d$), which improves on the bounds on the extreme discrepancy from [Gne05] and [Mha04], see [Gne08a, Thm. 2.2].

The author believes that the lower bound in (1.34) is sharp for all dimensions d. In the case $d = 2$ this is shown in [Gne08b], where a new construction method for minimal bracketing covers of $\mathcal{C}_2$ is introduced implying

$$N_{[]}(\mathcal{C}_2, \delta) \leq \delta^{-2} + o(\delta^{-2}). \tag{1.38}$$

Hence, in the case $d = 2$ the correct coefficient of the most significant term δ^{-d} in the expansion of $N_{[]}(\mathcal{C}_d, \delta)$ in powers of δ^{-1} is indeed 1. In [Gne08b] it is conjectured that in general

$$N_{[]}(\mathcal{C}_d, \delta) \leq \delta^{-d} + o_d(\delta^{-d}) \tag{1.39}$$

holds. For comparison also other bracketing covers of $\mathcal{C}_2$, namely the covers used by Thiémard in [Thi01a] and the covers presented in [DGS05], are studied, and the coefficient of δ^{-2} in the expansion of their cardinality is calculated explicitly—for those bracketing covers it is strictly larger than 1.

Infinite dimensional infinite sequences

So far we have discussed the existence of point sets that satisfy reasonably good discrepancy bounds. But those point sets might vary largely for different values of n and d. In practice it is desirable to have integration points at hand that can be extended in the number of points, and preferably also in the dimension d. This allows to achieve higher approximation accuracy while still be able to reuse earlier calculations. For this reason, e.g., lattice rules extensible in the parameters n and d have been studied, see [CKN06, HHLL00, HiN03, Nie03].

In [Dic07b] the probabilistic bounds stated in the previous subsection were extended by J. Dick to infinite sequences of infinite dimensional points. For an infinite sequence $\mathcal{P}$ of points in $[0,1)^{\mathbb{N}}$, let us denote by $\mathcal{P}_d$ the sequence of the projections of the points of $\mathcal{P}$ onto their first d components, and by $\mathcal{P}_{n,d}$ the first n points of $\mathcal{P}_d$. Then in [Dic07b] the following results were shown:

There exists an unknown constant C such that for every strictly increasing sequence $(n_m)_{m\in\mathbb{N}}$ in $\mathbb{N}$ there is an infinite sequence $\mathcal{P}$ in $[0,1)^{\mathbb{N}}$ satisfying, for all $m, d \in \mathbb{N}$,

$$d^*_\infty(\mathcal{P}_{n_m,d}) \leq C\sqrt{d/n_m}\sqrt{\ln(m+1)}.$$

Furthermore, there exists an explicitly given constant C' such that for every strictly increasing sequence $(n_m)_{m\in\mathbb{N}}$ in $\mathbb{N}$ there is an infinite sequence $\mathcal{P}$ satisfying, for all $m, d \in$

$\mathbb{N}$,

$$d^*_\infty(\mathcal{P}_{n_m,d}) \le C' \sqrt{\left(m + d + d \ln\left(1 + \frac{d\sqrt{n_m}}{m+d}\right)\right)/n_m}. \tag{1.40}$$

Thus, the results in [Dic07b] are both extensible in the dimension and in the number of points, which is particularly useful. A disadvantage of (1.40) is nevertheless that in the case where, e.g., $n_m = m$ for all m it is not better than the trivial bound $d^*_\infty(\mathcal{P}_{m,d}) \le 1$.

In the paper [DGKP08] another result for infinite sequences $\mathcal{P}$ in $[0,1)^{\mathbb{N}}$ is presented: There exists an explicitly given constant C'' such that for every strictly increasing sequence $(n_m)_{m\in\mathbb{N}}$ in $\mathbb{N}$ there is an infinite sequence $\mathcal{P}$ satisfying, for all $m, d \in \mathbb{N}$,

$$d^*_\infty(\mathcal{P}_{n_m,d}) \le C'' \sqrt{d \ln\left(1 + \frac{n_m}{d}\right)/n_m}. \tag{1.41}$$

This bound is clearly an improvement of (1.40), which in particular is still useful in the case $n_m = m$ for all m. Moreover, it establishes the existence of infinite sequences $\mathcal{P}$ in $[0,1)^{\mathbb{N}}$ having the following property: To guarantee $d^*_\infty(\mathcal{P}_{n,d}) \le \varepsilon$ for a given ε, we only have to take $n \ge c_\varepsilon d$, where c_ε is a constant depending only on ε, see [DGKP08, Cor. 2.4]. Note that this result cannot be deduced directly from the results in [Dic07b]. As mentioned above, it is known from [HNWW01, Hin04] that we have to take at least $n \ge c'_\varepsilon d$ if ε is sufficiently small. (Here c'_ε depends again only on ε.) In this sense [DGKP08, Cor. 2.4] shows that the statement "the inverse of the star discrepancy depends linearly on the dimension" (which is the title of the paper [HNWW01]) extends to the projections of infinite sequences in $[0,1)^{\mathbb{N}}$. To make this more precise, the notion of the *inverse of the star discrepancy of an infinite sequence* $\mathcal{P}$ is introduced in [DGKP08], given by

$$N^*_{\mathcal{P}}(\varepsilon, d) := \min\{n \,:\, \forall m \ge n : d^*_\infty(\mathcal{P}_{m,d}) \le \varepsilon\}.$$

Then Corollary 2.4 of [DGKP08] states that there exist sequences $\mathcal{P}$ such that

$$N^*_{\mathcal{P}}(\varepsilon, d) \le O(d\varepsilon^{-2} \ln(1 + \varepsilon^{-1})). \tag{1.42}$$

In fact even more holds: If we endow the set $[0,1)^{\mathbb{N}}$ with the canonical probability measure $\lambda_{\mathbb{N}} = \otimes_{i=1}^{\infty} \lambda_1$ and allow the implicit constant in the big-O-notation to depend on the particular sequence $\mathcal{P}$, then inequality (1.42) holds almost surely for a random sequence $\mathcal{P}$, see again [DGKP08, Cor. 2.4]. In [DGKP08, Thm. 2.3] bounds of the form (1.41) and (1.42) with explicitly given constants and estimates for the measure of the sets of sequences satisfying such bounds are provided.

1.4.2 Construction of Small Discrepancy Samples

Let us now discuss Question 1.4.2(ii) in more detail. Here we face the following situation, which may at first seem a bit surprising: On the one hand there are several construction methods known that provide point sets satisfying (1.21), and these constructions can be done quite efficiently. So one can construct, e.g., Halton-Hammersley points of size n in

dimension d with at most $O(dn\ln(n))$ elementary operations. On the other hand it seems to be hard to construct point sets efficiently that satisfy bounds like (1.24) or (1.32), although random sets should do this with high probability. That it is not trivial to find such constructions was underlined by S. Heinrich, who pointed out in [Hei03] that even the following easier problems are unsolved.

Problem 1.4.6. (i) For each $\varepsilon > 0$ and $d \in \mathbb{N}$, give a construction of a point set $\{t^1, \dots, t^n\} \subset [0,1]^d$ with $n \le c_\varepsilon d^{\kappa_\varepsilon}$ and $d^*_\infty(t^1, \dots, t^n) \le \varepsilon$, where c_ε and κ_ε are positive constants which may depend on ε, but not on d.

(ii) For each $n, d \in \mathbb{N}$, give a construction of a point set $\{t^1, \dots, t^n\} \subset [0,1]^d$ with $d^*_\infty(t^1, \dots, t^n) \le cd^\kappa n^{-\alpha}$, where c, κ and α are positive constants not depending on n or d.

(Although not stated explicitly in [Hei03], it is clear that these constructions are required to be efficiently executable, preferably in polynomial time in d, and ε^{-1} or n, respectively, see also [NoW08, Open Problem 6].) As stressed by Heinrich, it remains in particular a challenging question whether any of the various known constructions satisfies estimates like in Problem 1.4.6 or even the bounds (1.24) or (1.32).

There had been attempts from computer scientists to construct small low-discrepancy samples [EGLNV92, CRS00], but the size of those samples with guaranteed discrepancy at most ε in dimension d is not polynomial in d and ε^{-1}. The size of the best construction is polynomial in ε^{-1} and $(d/\ln(\varepsilon^{-1}))^{\ln(\varepsilon^{-1})}$ [CRS00]. Formally, those constructions solve Problem 1.4.6(i) (but not Problem 1.4.6(ii)), however, the authors did not state clearly how costly they actually are (in terms of memory or CPU-time). Obviously, the size of the samples is a lower bound of the costs, which are therefore not polynomial in d and ε^{-1}. For a more detailed discussion see [DGS05, Sect. 4.3].

An advantage of the probabilistic approach to prove upper bounds for the star discrepancy by using δ-covers and Hoeffding's inequality over the approach based on Haussler's result and the large deviation bound of Talagrand (as used in [HNWW01] and outlined in Section 1.4.1) is that it is known how to *derandomize* it to construct deterministically small samples satisfying bounds like (1.32).

Such a derandomized algorithm is presented in [DGS05]. There the underlying probabilistic experiment is to place n points uniformly at random on a properly chosen grid. We discretize the star discrepancy with the help of a δ-cover and consider only finitely many anchored test boxes (namely exactly those whose upper right corners lie in the δ-cover). We employ Hoeffding's inequality to show that the discrepancy of a random set in each single test box is small (i.e., near by the expected value) with high probability. A union bound shows that the maximum of the discrepancies in all finitely many test boxes is still small with high probability and, due to Lemma 1.4.4, differs at most by δ from the actual star discrepancy.

The basic idea of the derandomization now is to choose point after point such that the respective conditional probabilities for the event that we fail to satisfy the desired discrepancy bound (1.32) decrease. The actual algorithm does not compute the conditional probabilities exactly, which would be too costly, but calculates more economically

suitable upper bounds. Such suitable bounds appeared in the hyperbolic cosine algorithm of J. Spencer [Spe87] and were defined rigorously as so-called *pessimistic estimators* by P. Raghavan [Rag88]. The algorithm in [DGS05] relies on the efficiently computable pessimistic estimators described in [SrS96]. It constructs an n-point set T in dimension d satisfying (1.32) in a running time bounded by $O(C^d n^{d+2} \ln(d)^d / \ln(n)^{d-1})$, where C is a suitable constant.

An improved construction is presented in [DoG08]. There the probabilistic experiment to be derandomized is the following: We decompose $[0,1)^d$ into cells of a non-equidistant grid. Our test boxes are those boxes anchored in 0 that are unions of grid cells, and their upper right corners form a δ-cover. We randomly round the volume of each cell multiplied by n (which is the expected fraction of n uniformly distributed random points that lies inside this cell) to a non-negative integer (which is the actual number of points we will finally place inside this cell). This is done in such a way that the sum over all these integers is n (the number of output points). Since the randomized roundings have to satisfy this hard constraint, they are not independent random variables, but still negatively correlated, and thus large deviation inequalities of Chernoff-Hoeffding type still hold [PaS97].

By derandomizing this probabilistic experiment with techniques from [Doe06], we get a deterministic algorithm that constructs an n-point set T in dimension d satisfying again the bound (1.32), and its running time is bounded by $O(d \ln(dn)(\sigma n)^d)$, where $\sigma = \sigma(d) = \Theta\left(\ln(d)^2 / d \ln\ln(d)\right)$ tends to zero as d approaches infinity. Apart from the improved running time, the algorithm is technically less demanding and thus easier to implement than the algorithm presented in [DGS05].

Another approach is presented in [DGKP08]. There a component-by-component (CBC) construction of n-point sets via derandomization is proposed. In particular, via this approach given point sets can be extended in the dimension. Here the underlying probabilistic experiment is as follows: Given an n-point set $\mathcal{P}_{d'} = \{p^1, \ldots, p^n\}$ in dimension d', we choose new components $x^1, \ldots, x^n$ randomly from an appropriate one-dimensional grid and receive the new n-point set $\mathcal{P}_{d'+1} = \{(p^1, x^1), \ldots, (p^n, x^n)\}$. We may repeat this procedure until we obtain an n-point set in the desired dimension d. This probabilistic experiment can be derandomized with the classical method of Raghavan [Rag88]. If we start the CBC-construction in dimension one, the deterministic output set $\mathcal{P}_d$ of size n in dimension d satisfies the bound

$$d^*_\infty(\mathcal{P}_d) \leq O(d^{3/2} n^{-1/2} \ln(1 + n/d)^{1/2}). \tag{1.43}$$

and the running time of the algorithm is bounded by $O(c^d n^{(d+3)/2} (d \ln(1 + n/d))^{-(d+1)/2})$, c a suitable constant independent of n and d. Certainly the bound (1.43) is weaker than (1.32), but the bound on the running time of the CBC algorithm is a reasonable improvement upon the running time guarantees of the two derandomized algorithms mentioned before. The CBC-algorithm has the additional nice feature that it can calculate the exact discrepancy of the output set without essentially more effort.

In [DGW09] some more implementation details of the CBC-algorithm are provided and several numerical tests are performed. In particular, the experiments indicate that the

discrepancies of the output sets of the CBC-algorithm behave in practice much better than predicted by the theoretical bound (1.43). They depend rather linear on the dimension d than proportional to $d^{3/2}$. The numerical experiments reveal that the discrepancies of the output sets, which are subsets of certain full d-dimensional grids, are almost exactly equal to the discrepancies of the full grids (for reasons explained in [DGW09] we want to call the latter discrepancies "grid gaps"). For output sets of size n the corresponding full grid has size larger than $n^{s/2}/s!$. We may interpret this result in a positive way: The CBC-algorithm provides a sparse sample from a complete d-dimensional grid, which exhibits essentially the same discrepancy as the full grid.

To overcome the lower bound on the discrepancy given by the "grid gap", we also consider a randomized CBC-variant: After receiving an output set $\mathcal{P}_d$, we randomize its points locally to receive a new output set $\mathcal{P}_d^*$. For the randomized set $\mathcal{P}_d^*$ the theoretical discrepancy bound (1.43) still holds, and in all the numerical tests in dimension $d = 10$ its discrepancy was always much smaller as the corresponding grid gap (which, as already said, is a lower bound for $d_\infty^*(\mathcal{P}_d)$). (To be more precise, an estimator for $d_\infty^*(\mathcal{P}_d^*)$, which majorizes $d_\infty^*(\mathcal{P}_d^*)$ with certainty at least 95%, is always much smaller than the corresponding grid gap. We use this estimator, since calculating the actual discrepancy of $\mathcal{P}_d^*$ is a much harder problem than calculating the discrepancy of $\mathcal{P}_d$. We shall discuss the problem of calculating the star discrepancy of given sets in detail in Section 1.4.3.)

As already mentioned, we can use the CBC method to extend an existing point set to one in higher dimension. This is a promising idea, because the classical low-discrepancy constructions like Halton-Hammersley, Faure or Niederreiter-Xing points [Hal60, Fau82, NiX96] behave very well in small dimensions. This advantage is used by the related idea of *hybrid-Monte Carlo sequences*, which are sometimes also called *mixed sequences*—there a low-discrepancy sequence is extended in the dimension by choosing the additional coordinates randomly, see, e.g., [Spa95, ÖTB06, Gne09]. In many applications it has been observed that hybrid-Monte Carlo sequences perform better than pure Monte Carlo and pure quasi-Monte Carlo sequences, especially in difficult problems.

In our numerical tests we compare in dimension $d = 10$ pure Halton-Hammersley point sets with pure CBC-constructions and d'-dimensional Halton-Hammersley points, $d' < d$, that are extended in the dimension by the CBC-algorithm. Our results indicate that extended point sets behave better for moderate values of n than pure low-discrepancy point sets or than pure CBC-points.

As already mentioned above, the CBC-approach has relations to hybrid-Monte Carlo sequences, which are often called *mixed sequences*.

For a mixed s-dimensional sequence m, whose elements are, technically speaking, vectors obtained by concatenating d'-dimensional vectors from a low-discrepancy sequence q with $(d-d')$-dimensional random vectors, probabilistic upper bounds for its star discrepancy have been provided. If m_n and q_n denote the sets of the first n points of the sequences m and q respectively, then G. Ökten, B. Tuffin, and V. Burago showed in [ÖTB06] that

$$\mathbb{P}(d_\infty^*(m_n) - d_\infty^*(q_n) < \varepsilon) \geq 1 - 2\exp\left(-\frac{\varepsilon^2 n}{2}\right) \quad \text{for } n \text{ sufficiently large.} \tag{1.44}$$

The authors did not study how large n actually has to be and if and how this actually depends on the parameters d and ε. In the note [Gne09] a lower bound for n is derived, which significantly depends on s and ε. Furthermore, the probabilistic bound

$$\mathbb{P}(d_\infty^*(m_n) - d_\infty^*(q_n) < \varepsilon) > 1 - 2N(\mathcal{C}_d, \varepsilon/2) \exp\left(-\frac{\varepsilon^2 n}{2}\right) \tag{1.45}$$

is proved, which holds without any restriction on n. In this sense it improves the bound (1.44) and is more helpful in practice, especially for small samples sizes n. As we know from (1.30), for small ε the quantity $N(\mathcal{C}_d, \varepsilon/2)$ grows exponentially in d. As pointed out in [Gne09, Remark 3.4] a factor depending exponentially on d has to appear in front of $\exp(-\varepsilon^2 n/2)$ in the bound (1.45) if we want to hold it for all $n \in \mathbb{N}$.

In the literature one can find other bounds similar to (1.44) without any restrictions on n. Examples are [Ökt96, Corollary 1] and [Roş07, Theorem 8] (the latter result is actually not about the star discrepancy, but about a generalization of the extreme discrepancy). As discussed in [ÖkG09] and [GnR09], these two results are unfortunately incorrect.

1.4.3 Calculation of the Star Discrepancy

Let us now discuss Question 1.4.2(iii). In some applications it is of interest to measure the quality of certain point sets by calculating their star discrepancy, e.g., to test whether successive pseudo random numbers are statistically independent [Nie92], or whether sample sets are suitable for multivariate numerical integration of particular classes of integrands, cf. Theorem 1.4.1. Apart from that, it is particularly interesting with respect to Question 1.4.2(ii) that the fast calculation or approximation of the star discrepancy would allow practicable semi-constructions of low-discrepancy samples of moderate size as described below Question 1.4.2.

Let us recall the underlying simple idea: Let $\mathcal{X}$ be a set of point configurations endowed with some probability measure. Let us assume that the mean value of the star discrepancy taken over all point configurations in $\mathcal{X}$ is small enough for our purpose, and that there exists a large deviation bound ensuring that the star discrepancy is sharply concentrated around its mean. Then we can choose a point set from $\mathcal{X}$ randomly and calculate its actual star discrepancy. If it is below or at least sufficiently close to the mean, we accept the set, otherwise we choose another set randomly from $\mathcal{X}$. The large deviation bound guarantees us that with high probability we have to make only a small number of random choices to receive a low-discrepancy set that serves our needs.

As discussed in Section 1.4.2 derandomized algorithms to construct such samples deterministically are known [DoG08, DGKP08, DGS05], but all of those have a large running time. Therefore efficient semi-constructions would be of practical value to avoid the costly derandomization procedures. The critical step of the semi-construction described above is the efficient calculation or approximation of the star discrepancy of random sets.

It is, e.g., known that the L_2-star discrepancy of a given n-point set in dimension d can be calculated via Warnock's formula [War72] with $O(dn^2)$ arithmetic operations and similar formulas hold for weighted versions of the L_2-star discrepancy. S. Heinrich and

K. Frank developed an asymptotically even faster algorithm for the L_2-star discrepancy using $O(n(\log n)^{d-1})$ operations for fixed d [FrH96, Hei96]. (Due to the exponent of the log-term, the algorithm is only practicable in low dimensions.) But, as pointed out by J. Matoušek in [Mat99, Sect. 2.4], no similarly efficient algorithms are known for the L_p-star discrepancy if $p \neq 2$.

So what methods are known to calculate or approximate the star discrepancy of a given set $\mathcal{P}$? At the first glance an exact calculation seems to be difficult since the star discrepancy is defined as the supremum over infinitely many test boxes and we can of course not calculate infinitely many quantities. But it is not hard to see that we can in fact calculate the discrepancy of $\mathcal{P}$ exactly by considering only finitely many test boxes. So if $\mathcal{P} = \{p^1, \ldots, p^n\} \subset [0,1)^d$, let us define

$$\Gamma_j(\mathcal{P}) = \{p_j^i \mid \ i \in \{1, ..., n\}\} \quad \text{and} \quad \bar{\Gamma}_j(\mathcal{P}) = \Gamma_j(\mathcal{P}) \cup \{1\},$$

and let us put

$$\Gamma(\mathcal{P}) = \Gamma_1(\mathcal{P}) \times \cdots \times \Gamma_d(\mathcal{P}) \quad \text{and} \quad \bar{\Gamma} = \bar{\Gamma}_1(\mathcal{P}) \times \cdots \times \bar{\Gamma}_d(\mathcal{P}).$$

Then one can show

$$d_\infty^*(\mathcal{P}) = \max\left\{ \max_{y \in \bar{\Gamma}(\mathcal{P})} \left(\lambda_d([0,y)) - \frac{1}{n}|\mathcal{P} \cap [0,y)| \right) \ , \ \max_{y \in \Gamma(\mathcal{P})} \left(\frac{1}{n}|\mathcal{P} \cap [0,y]| - \lambda_d([0,y)) \right) \right\}, \tag{1.46}$$

see, e.g., [GSW09]. Thus we need to consider at most $O(n^d)$ test boxes to compute $d_\infty^*(\mathcal{P})$. For a random n-point set $\mathcal{P}$ we have almost surely $|\Gamma(\mathcal{P})| = n^d$, resulting in $\Omega(n^d)$ test boxes that we have to take into account to calculate (1.46). This underlines that (1.46) is in general impractical if n and d are large. There are some more sophisticated methods known to calculate the star discrepancy, which are especially helpful in low dimensions. If we have in the one-dimensional case $p^1 \leq p^2 \leq \cdots \leq p^n$, then (1.46) simplifies to

$$\begin{aligned} d_\infty^*(\mathcal{P}) &= \max_{i=1}^{n} \max\left\{ \left| \frac{i}{n} - p^i \right|, \left| \frac{i-1}{n} - p^i \right| \right\} \\ &= \frac{1}{2n} + \max_{i=1}^{n} \left| p^i - \frac{2i-1}{2n} \right|, \end{aligned}$$

a result due to H. Niederreiter, see [Nie72a, Nie72b].

In dimension $d = 2$ a reduction of the number of steps to calculate (1.46) was achieved by L. DeClerk [DeC86]. In [BuZ93] her formula was slightly extended and simplified by P. Bundschuh and Y. Zhu. If we assume $p_1^1 \leq p_1^2 \leq \cdots \leq p_1^n$ and rearrange for each $i \in \{1, \ldots, n\}$ the numbers $0, p_2^1, \ldots, p_2^i, 1$ in increasing order and rewrite them as $0 = \xi_{i,0} \leq \xi_{i,1} \leq \cdots \leq \xi_{i,i} \leq \xi_{i,i+1} = 1$, then [BuZ93, Thm. 1] states that

$$d_\infty^*(\mathcal{P}) = \max_{i=0}^{n} \max_{k=0}^{i} \max\left\{ \left| \frac{k}{n} - p_1^i \xi_{i,k} \right|, \left| \frac{k}{n} - p_1^{i+1} \xi_{i,k+1} \right| \right\}.$$

Bundschuh and Zhu provided also a corresponding formula for the three-dimensional case. The method can be generalized to arbitrary dimension d and requires $O(n^d/d!)$ elementary operations. This method was, e.g., used in [WiF97] to calculate the exact discrepancy of particular point sets, so-called rank-1 lattice rules (cf. [SlJ94]), up to size $n = 236$ in dimension $d = 5$ and to $n = 92$ in dimension $d = 6$. But as pointed out by P. Winker and K.-T. Fang, for this method instances like, e.g., sets of size $n \geq 2000$ in $d = 6$ are completely infeasible.

Another method to calculate the star discrepancy in time $O(n^{1+d/2})$ was proposed by D. P. Dobkin, D. Eppstein, and D. P. Mitchell in [DEM96]. It uses sophisticated, but complicated data structures. Asymptotically slightly slower variants of the algorithm have only been implemented in dimension $d = 2$.

The discussion shows that all known methods that calculate the star discrepancy exactly depend exponentially on the dimension d and are infeasible for large values of n and d. What is known about approximation algorithms?

An approach that approximates the star discrepancy of a given set $\mathcal{P}$ up to a user-specified error δ was presented by E. Thiémard [Thi00, Thi01a]. It is in principle based on the generation of distinguished δ-bracketing covers (which were not named this way in [Thi00, Thi01a]), the relation (1.33) and Lemma 1.4.4. For the best of his δ-bracketing covers $\mathcal{B}_\delta^d$, constructed in [Thi01a], Thiémard proved the bound

$$|\mathcal{B}_\delta^d| \leq e^d(\delta^{-1}\ln(\delta^{-1}) + 1)^d,$$

which is weaker than the bounds (1.34) and (1.35) from [Gne08a]. Concrete comparisons of Thiémard's bracketing covers with other constructions in dimension $d = 2$ can be found in [Gne08b], where optimal two-dimensional bracketing covers are provided. The lower bound in (1.34) proved in [Gne08a] immediately implies a lower bound for the running time of Thiémard's algorithm, regardless how cleverly the δ-bracketing cover is chosen. That is because the dominating factor in the running time is the construction of the δ-bracketing cover $|\mathcal{B}_\delta^d|$, which is of order $\Theta(d|\mathcal{B}_\delta^d|)$. Thus (1.34) shows that the running time of the algorithm is exponential in d. (Nevertheless smaller δ-bracketing covers, which may, e.g., be generated by extending the ideas from [Gne08b] to arbitrary dimension d, would widen the range of applicability of Thiémard's algorithm.)

Since in high dimension no efficient algorithm for the exact calculation or tight approximation of the star discrepancy is known, some authors tried to tackle the large scale enumeration problem (1.46) by using optimization heuristics. In [WiF97] P. Winker and K.-T. Fang used threshold accepting [DuS90], a refined randomized local search algorithm based on a similar idea as the well-known simulated annealing algorithm [KGV83], to find lower bounds for the star discrepancy. The algorithm performed well in numerical tests on rank-1 lattice rules, but in general no approximation quality can be guaranteed. A modified version of this algorithm was proposed in [Gne07b], and tests in [Gne07b] and [Win07] indicate that this version improves on the algorithm from [WiF97], especially in higher dimension $20 \leq d \leq 40$.

In [Thi01b] E. Thiémard gave an integer linear programming formulation for the problem and used techniques as cutting plane generation and branch and bound to tackle it.

With the resulting algorithm Thiémard performed non-trivial star discrepancy comparisons between low-discrepancy sequences. Numerical tests in [Thi01a, Thi01b] show that the algorithm from [Thi01b] can handle instances that are infeasible for the algorithm from [Thi01a], which indicates strongly that the former algorithm is in practice superior to the latter one.

The key observation to approach a highly non-linear expression as (1.46) via linear programming is that one can reduce it to at most $2n$ sub-problems of the type "optimal volume subintervals with k points". These sub-problems are the problems of finding the largest boxes $[0, y)$, $y \in \overline{\Gamma}(\mathcal{P})$, containing k points, $k \in \{0, 1, \ldots, n-1\}$, and the smallest boxes $[0, y]$, $y \in \Gamma(\mathcal{P})$, containing ℓ points for $\ell \in \{1, \ldots, n\}$. Thiémard conjectured these sub-problems to be NP-hard.

In [GSW09] the conjecture of Thiémard is proved rigorously by establishing the NP-hardness of the optimal volume subinterval problems. Recall that NP-hardness of an optimization problem U is proved by verifying that deciding the so-called threshold language of U is an NP-hard decision problem (see, e.g., [Hro03, Sect. 2.3.3]). Thus actually the NP-completeness of decision problems corresponding to the optimization problems mentioned above is verified. The verification is done by reduction of the problem Dominating Set to the maximal volume subinterval problems and of the problem Balanced Subgraph to the minimal subinterval problems, respectively; the graph theoretical decision problems Dominating Set and Balanced Subgraph are known to be NP-hard, see [GaJ79, Joh87].

With the help of these NP-hardness results for the optimal volume subinterval problems it is shown that the problem of calculating the star discrepancy itself is NP-hard.

Furthermore, some errors occurring in [Thi01b] are listed in [GSW09]. Since those errors may lead to incorrect solutions of Thiémard's algorithm for certain instances, it is explained how to avoid their undesired consequences.

1.4.4 Conclusion and Open Problems

In the previous subsections we discussed Question 1.4.2 and described in particular how the articles [Gne05, DGS05, Gne08a, Gne08b, DoG08, DGKP08, DGW09, Gne09, GSW09] contributed to achieve at least partial solutions.

The discussion shows that good bounds for the star and the extreme discrepancy with explicitly known constants are available, see (1.37) and [Gne08a, Thm. 2.2]. Similar bounds hold also for the star discrepancy of point sets that are extensible in the number of points and in the dimension (1.41), and the statement that the inverse of the star discrepancy depends linearly on the dimension d [HNWW01] can be extended to this situation: As shown in [DGKP08] the inverse of the star discrepancy of infinite sequences in $[0, 1)^{\mathbb{N}}$ depends almost surely linearly on the dimension d, see (1.42).

Can we find even better bounds than (1.37) or (1.24)? A lower bound for the star discrepancy that follows directly from (1.25) is of the form $d^*_\infty(n, d) \geq \max\{\varepsilon_0, c_0 d n^{-1}\}$, c_0 and ε_0 suitable constants [Hin04, Thm. 1], so improvements of (1.37) or (1.24) might be possible. Instead of answering this intriguing question, let us state the conjecture of

H. Woźniakowski (see [Hei03], or [NoW08, Open Problem 7]): If there exist constants $C, \alpha > 0$ and a polynomial p such that

$$d^*_\infty(n,d) \le C\, p(d) n^{-\alpha} \quad \text{for all } d, n \in \mathbb{N}, \tag{1.47}$$

then necessarily $\alpha \le 1/2$. If this conjecture holds, then it is not possible to derive a bound polynomial in d and n that is better than (1.24).

Concerning the construction of point sets satisfying bounds like (1.24) or (1.32) some reasonable progress has been made, see [DGS05, DoG08, DGKP08, DGW09]. Nevertheless, the algorithms from [DGS05, DoG08, DGKP08] exhibit running times exponential with respect to the dimension d, a fact prohibiting their use in really high dimensions.

This is maybe not too surprising, since even the seemingly easier problem of calculating the star discrepancy of an arbitrary point set (or approximating it up to a user-specified error) can so far be solved only in exponential time in d. And indeed the problem of calculating the star discrepancy is now known to be NP-hard [GSW09].

The discussed derandomized algorithms can be used in low and modestly high dimension d. With the actual version of the fastest derandomized algorithm, presented in [DGKP08] and implemented in [DGW09], point sets of size up to 1000 points in dimension $d = 8$ or 500 points in dimension $d = 10$ can be constructed in tolerable time.

It would be of interest to make further progress in designing algorithms that construct low-discrepancy point sets of small size or that approximate the star discrepancy of arbitrary n-point sets (which would allow "semi-constructions" as described above). It would be interesting to know if one is able to strengthen the NP-hardness result from [GSW09]: there the point sets whose discrepancy is hard to compute are rather artificial and they exhibit in fact a large discrepancy. That means the NP-hardness result does not imply that the approximation of the discrepancy of certain classes of low-discrepancy point sets cannot be done efficiently.

Bibliography

[Ata04] E. Atanassov, On the discrepancy of the Halton sequences, Math. Balkanica (N.S.) 18 (2004), 15–32.

[Bec84] J. Beck, Some upper bounds in the theory of irregularities of distribution, Acta Arith. 43 (1984), 115–130.

[Bec89] J. Beck, A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution, Composito Math. 72 (1989), 269–339.

[BeC87] J. Beck, W. W. L. Chen, Irregularities of Distribution, Cambridge University Press, Cambridge, 1987.

[Bel57] R. E. Bellman, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

[Ber36] V. Bergström, Einige Bemerkungen zur Theorie der diophantischen Approximationen, Fysiog. Sälsk. Lund. Förh. 6 (1936), 1–19.

[BiL08] D. Bilyk, M. T. Lacey, On the small ball inequality in three dimensions, Duke Math. J. 143 (2008), 81–115.

[BLV08] D. Bilyk, M. T. Lacey, A. Vagharshakyan, On the small ball inequality in all dimensions, J. Funct. Anal. 254 (2008), 2470–2502.

[BSS89] L. Blum, M. Shub, 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 (1989), 1–46.

[BuZ93] P. Bundschuh, Y. Zhu, A method for the exact calculation of the discrepancy of low-dimensional finite point sets (I), Abh. Math. Sem. Univ. Hamburg 63 (1993), 115–133.

[BuG04] H. J. Bungartz, M. Griebel, Sparse grids, Acta Numerica 13 (2004), 147–269.

[CRS00] S. Chari, P. Rohatgi, A. Srinivasan, Improved algorithms via approximation of probability distributions, J. Comput. System Sci. 61 (2000), 81–107.

[Cha00] B. Chazelle, The Discrepancy Method, Cambridge University Press, Cambridge, 2000.

[CKN06] R. Cools, F. Y. Kuo, D. Nuyens, Constructing embedded lattice rules for multivariate integration, SIAM J. Sci. Comput. 28 (2006), 2162–2188.

[Cor35a] J. G. van der Corput, Verteilungsfunktionen I, Nederl. Akad. Wetensch. Proc. Ser. B 38 (1935), 813–821.

[Cor35b] J. G. van der Corput, Verteilungsfunktionen II, Nederl. Akad. Wetensch. Proc. Ser. B 38 (1935), 1058–1066.

[Dav56] H. Davenport, Note on irregularities of distribution, Mathematika 3 (1956), 131–135.

[DeC86] L. De Clerk, A method for the exact calculation of the star-discrepancy of plane sets applied to the sequences of Hammersley, Mh. Math. 101 (1986), 261–278.

[Dic07a] J. Dick, Explicit constructions of quasi-Monte Carlo rules for the numerical integration of high-dimensional periodic functions, SIAM J. Numer. Anal. 45 (2007), 2141–2176.

[Dic07b] J. Dick, A note on the existence of sequences with small star discrepancy, J. Complexity 23 (2007), 649–652.

[Dic08a] J. Dick, Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order, SIAM J. Numer. Anal. 46 (2008), 1510–1553.

[Dic08b] J. Dick, Koksma-Hlawka type inequalities of fractional order, Ann. Mat. Pura Appl. 187 (2008), 385–403.

[DiP07] J. Dick, F. Pillichshammer, Strong tractability of multivariate integration of arbitrary high order using digitally shifted polynomial lattice rules, J. Complexity 23 (2007), 436–453.

[DSWW04] J. Dick, I. H. Sloan, X. Wang, H. Woźniakowski, Liberating the weights, J. Complexity 20 (2004), 593–623.

[DSWW06] J. Dick, I. H. Sloan, X. Wang, H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights, Numer. Math. 103 (2006), 63–97.

[DEM96] D. P. Dobkin, D. Eppstein, D. P. Mitchell, Computing the discrepancy with applications to supersampling patterns, ACM Trans. on Graphics 15 (1996), 354–376.

[Doe06] B. Doerr, Generating randomized roundings with cardinality constraints and derandomizations, in: B. Durand, W. Thomas (Eds.), Proceedings of the 23rd Annual Symposium on Theoretical Aspects of Computer Science, 571–583, LNiCS 3884, Springer, Berlin-Heidelberg, 2006.

[DoG08] B. Doerr, M. Gnewuch, Construction of low-discrepancy point sets of small size by bracketing covers and dependent randomized rounding, in: A. Keller, S. Heinrich, H. Niederreiter (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2006, 299–312, Springer, Berlin Heidelberg, 2008.

[DGKP08] B. Doerr, M. Gnewuch, P. Kritzer, F. Pillichshammer, Component by component construction of low-discrepancy sets of small size, Monte Carlo Methods Appl. 14 (2008), 129–149.

[DGS05] B. Doerr, M. Gnewuch, A. Srivastav, Bounds and constructions for the star-discrepancy via δ-covers, J. Complexity 21 (2005), 691–709.

[DGW09] B. Doerr, M. Gnewuch, M. Wahlström, Implementation of a component-by-component algorithm to generate small low-discrepancy samples, to appear in the proceedings of Monte Carlo and Quasi-Monte Carlo Methods 2008.

[DrT97] M. Drmota, R. F. Tichy, Sequences, Discrepancies and Applications, Lecture Notes in Mathematics 1651, Springer, Berlin Heidelberg, 1997.

[DuS90] G. Dueck, T. Scheuer, Threshold accepting: A general purpose algorithm appearing superior to simulated annealing, J. Comp. Phys. 90 (1990), 161–175.

[EGLNV92] G. Even, O. Goldreich, M. Luby, N. Nisan, B. Veličković, Approximations of general independent distributions, in: Proceedings of the 24th ACM Symposium on Theory of Computing (STOC), 10–16, 1992.

[Fau82] H. Faure, Discrépance de suites associées à un système de numération (en dimension s), Acta Arith. 41 (1982), 337–351.

[FrH96] K. Frank, S. Heinrich, Computing discrepancies of Smolyak quadrature rules, J. Complexity 12 (1996), 287–314.

[Fro80] K. K. Frolov, An upper estimate for the discrepancy in the L_p-metric, $2 \leq p \leq \infty$, Soviet. Math. Dokl. 21 (1980), 840–842.

[GaJ79] M. R. Garey, D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman and Co., San Francisco, 1979.

[GeG98] T. Gerstner, M. Griebel, Numerical integration using sparse grids, Numer. Algorithms 18 (1998), 209–232.

[Gne04] M. Gnewuch (Joint with B. Doerr), Geometric discrepancies and delta-covers, Oberwolfach Reports 1 (2004), 687-690.

[Gne05] M. Gnewuch, Bounds for the average L^p-extreme and the L^∞-extreme discrepancy, Electron. J. Combin. 12 (2005), Research Paper 54.

[Gne07a] M. Gnewuch, Construction of low-discrepancy point sets of small size, Proc. Appl. Math. Mech. 7 (2007), Special Issue: Sixth International Congress on Industrial Applied Mathematics and GAMM Annual Meeting, Zürich 2007, 1022605-1022606.

[Gne07b] M. Gnewuch, A randomized algorithm to compute the star-discrepancy of a given n-point set, manuscript, 2007.

[Gne08a] M. Gnewuch, Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy, J. Complexity 24 (2008), 154–172.

[Gne08b] M. Gnewuch, Construction of Minimal Bracketing Covers for Rectangles, Electron. J. Combin. 15(1) (2008), Research Paper 95.

[Gne09] M. Gnewuch, On Probabilistic results for the discrepancy of a hybrid-Monte Carlo sequence, J. Complexity 2009. (http://dx.doi.org/10.1016/j.jco.2009.02.009)

[GLSS07] M. Gnewuch, R. Lindloh, R. Schneider, A. Srivastav, Cubature formulas for function spaces with moderate smoothness, J. Complexity 23 (2007), 828–850.

[GnR09] M. Gnewuch, A. V. Roşca, On G-discrepancy and mixed Monte Carlo and quasi-Monte Carlo sequences, submitted.

[GSW09] M. Gnewuch, A. Srivastav, C. Winzen, Finding optimal volume subintervals with k points and calculating the star discrepancy are NP-hard problems, J. Complexity 25 (2009), 115–127.

[GnW07] M. Gnewuch, H. Woźniakowski, Generalized tractability for multivariate problems, part I: Linear tensor product problems and linear information, J. Complexity 23 (2007), 262-295.

[GnW08] M. Gnewuch, H. Woźniakowski, Generalized tractability for linear functionals, in: A. Keller, S. Heinrich, H. Niederreiter (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2006, 359-381, Springer, Berlin Heidelberg, 2008.

[GnW09] M. Gnewuch, H. Woźniakowski, Generalized tractability for multivariate problems, part II: Linear tensor product problems, linear information, and unrestricted tractability, Found. Comput. Math. 2009. (DOI: 10.1007/s10208-009-9044-6)

[Hal60] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, Numer. Math. 2 (1960), 84–90.

[HaL22] G. H. Hardy, J. E. Littlewood, Some problems of diophantine approximation: the lattice points of a right-angled triangle. Part I, Proc. London Math. Soc. 20 (1922), 212–249.

[Hau95] D. Haussler, Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension, J. Combin. Theory A 69 (1995), 279–302.

[Hei96] S. Heinrich, Efficient algorithms for computing the L_2 discrepancy, Math. Comp. 65 (1996), 1621–1633.

[Hei03] S. Heinrich, Some open problems concerning the star-discrepancy, J. Complexity 19 (2003), 416–419.

[HHY04] S. Heinrich, F. J. Hickernell, R.-X. Yue, Optimal quadrature for Haar wavelet spaces, Math. Comp. 73 (2004), 259–277.

[HNWW01] S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, Acta Arith. 96 (2001), 279–302.

[Hic98] F. J. Hickernell, A generalized discrepancy and quadrature error bound, Math. Comp. 67 (1998), 299–322.

[Hic02] F. J. Hickernell, Obtaining $O(N^{-2+\varepsilon})$ convergence for lattice quadrature rules, in: K. T. Fang, F. J. Hickernell, H. Niederreiter (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2000, 274–289, Springer, Berlin, 2002.

[HHLL00] F. J. Hickernell, H. S. Hong, P. L'Écuyer, C. Lemieux, Extensible lattice sequences for quasi-Monte Carlo quadrature, SIAM J. Sci. Comput. 22 (2000), 1117–1138.

[HiN03] F. J. Hickernell, H. Niederreiter, The existence of good extensible rank-1 lattices, J. Complexity 19 (2003), 286–300.

[HSW04] F. J. Hickernell, I. H. Sloan, G. W. Wasilkowski, On tractability of weighted integration over bounded and unbounded regions in $\mathbb{R}^s$, Math. Comp. 73 (2004), 1885–1901.

[Hin04] A. Hinrichs, Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy, J. Complexity 20 (2004), 477-483.

[HPS08] A. Hinrichs, F. Pillichshammer, W. Ch. Schmid, Tractability properties of the weighted star discrepancy, J. Complexity 23 (2008), 134–143.

[Hla61] E. Hlawka, Funktionen von beschränkter Variation in der Theorie der Gleichverteilung, Ann. Mat. Pura Appl. 54 (1961), 325–333.

[HoK96] J. Hoogland, R. Kleis, Discrepancy-based error estimates for quasi-Monte Carlo, I: General formalism, Comp. Phys. Comm. 98 (1996), 111–127.

[Hro03] J. Hromković, Algorithmics for Hard Problems, 2nd edition, Springer, Berlin Heidelberg, 2003.

[JHK97] F. James, J. Hoogland, R. Kleis, Multidimensional sampling for simulation and integration: measures, discrepancies, and quasi-random numbers, Comp. Phys. Comm. 99 (1997), 180–220.

[Joe06] S. Joe, Construction of good rank-1 lattice rules based on the weighted star discrepancy, in: H. Niederreiter, D. Talay (Eds.), Monte Carlo and quasi-Monte Carlo methods 2004, 181–196, Springer, Berlin, 2006.

[Joh87] D. S. Johnson, The NP-completeness column: an ongoing guide, J. Algorithms 8 (1987), 438–448.

[KGV83] S. Kirkpatrick, C. Gelatt, M. Vecchi, Optimization by simulated annealing, Science 20 (1983), 671–680.

[KlT97] B. Klinger, R. F. Tichy, Polynomial discrepancy of sequences, J. Comput. Appl. Math. 84 (1997), 107–117.

[Kok42] J. F. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1, Mathematica B (Zutphen) 11 (1942/3), 7–11.

[Kuo03] 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 (2003), 301–320.

[LaT94] G. Larcher, C. Traunfellner, On the numerical integration of Walsh series by number-theoretic methods, Math. Comput. 63 (1994), 277–291.

[LEL00] P. L'Ecuyer, C. Lemieux, Variance reduction via lattice rules, Management Science 46 (2000), 1214–1235.

[Mat99] J. Matoušek, Geometric Discrepancy, Springer, Berlin, 1999.

[Mha04] H. N. Mhaskar, On the tractability of multivariate integration and approximation by neural networks, J. Complexity 20 (2004), 561–590.

[Nie72a] H. Niederreiter, Discrepancy and convex programming, Ann. Mat. Pura Appl. 93 (1972), 89–97.

[Nie72b] H. Niederreiter, Methods for estimating discrepancy, in: S. K. Zaremba (Ed.), Applications of Number Theory to Numerical Analysis, 203–236, Academic Press, New York, 1972.

[Nie92] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, 1992.

[Nie03] H. Niederreiter, The existence of good extensible polynomial lattice rules, Monatsh. Math. 139 (2003), 295–307.

[NiX96] H. Niederreiter, C. Xing, Low-discrepancy sequences and global function fields with many rational places, Finite Fields Appl. 2 (1996), 561–590.

[Nov95] E. Novak, The real number model in numerical analysis, J. Complexity 11 (1995), 57–73.

[Nov04] E. Novak (Joint work with A. Hinrichs), New bounds for the star discrepancy, Oberwolfach Reports 1 (2004), 696-699.

[NoR96] E. Novak, K. Ritter, High-dimensional integration of smooth functions over cubes, Numer. Math. 75 (1996), 79–97.

[NoW01a] E. Novak, H. Woźniakowski, When are integration and discrepancy tractable?, Foundations of computational mathematics (Oxford, 1999), 211–266, London Math. Soc. Lecture Note Ser., 284, Cambridge Univ. Press, Cambridge, 2001.

[NoW01b] E. Novak, H. Woźniakowski, Intractability results for integration and discrepancy, J. Complexity 17 (2001), 388–441.

[NoW08] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, Volume I: Linear Information, European Mathematical Society, Freiburg, 2008.

[NoW09] E. Novak, H. Woźniakowski, L_2 discrepancy and multivariate integration, in: W. W. L. Chen, H. T. Gowers, H. Halberstam, W. M. Schmidt, R. C. Vaughan (Eds.), Analytic Number Theory: Essays in Honour of Klaus Roth, 359–388, Cambridge University Press 2009.

[NuC06] D. Nuyens, R. Cools, Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, Math. Comput. 75 (2006), 903–920.

[Ökt96] G. Ökten, A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications, Monte Carlo Methods Appl. 2 (1996), 255–270.

[ÖkG09] G. Ökten, M. Gnewuch, Correction of a proof in "A Probabilistic Result on the Discrepancy of a Hybrid-Monte Carlo Sequence and Applications", to appear in Monte Carlo Methods Appl..

[ÖTB06] G. Ökten, B. Tuffin, V. Burago, A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance, J. Complexity 22 (2006), 435–458.

[Ost22a] A. Ostrowski, Bemerkungen zur Theorie der Diophantischen Approximationen I, Abh. Math. Semin. Univ. Hamburg 1 (1922), 77–98.

[Ost22b] A. Ostrowski, Bemerkungen zur Theorie der Diophantischen Approximationen II, Abh. Math. Semin. Univ. Hamburg 1 (1922), 250–251.

[Owe97] A. B. Owen, Monte Carlo variance of scrambled net quadrature, SIAM J. Numer. Anal. 34 (1997), 1884–1910.

[PaS97] A. Panconesi, A. Srinivasan, Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds, SIAM J. Comput. 26 (1997), 350–368.

[PaP09] A. Papageorgiou, I. Petras, On the tractability of linear tensor product problems in the worst case, J. Complexity (2009). (doi:10.1016/j.jco.2009.05.002)

[Pas93] S. H. Paskov, Average case complexity of multivariate integration for smooth functions, J. Complexity 9 (1993), 291–312.

[Pla96] L. Plaskota, Noisy Information and Computational Complexity, Cambridge University Press, Cambridge, 1996.

[PWZ09] L. Plaskota, G. W. Wasilkowski, Y. Zhao, New averaging technique for approximating weighted integrals, J. Complexity 25 (2009), 268–291.

[Pro88] P. D. Proinov, Discrepancy and integration of continuous functions, J. Approx. Theory 52 (1988), 121–131.

[Rag88] P. Raghavan, Probabilistic construction of deterministic algorithms: approximating packing integer programs, J. Comput. System Sci. 37 (1988), 130–143.

[Roş07] A. V. Roşca, A mixed Monte Carlo and quasi-Monte Carlo sequence for multidimensional integral estimation, Acta Univ. Apulensis Math. Inform. 14 (2007), 141–160.

[Rot54] K. F. Roth, On irregularities of distribution, Mathematika 1 (1954), 73–79.

[Rot79] K. F. Roth, On irregularities of distribution III, Acta Arith. 35 (1979), 373–384.

[Rot80] K. F. Roth, On irregularities of distribution IV, Acta Arith. 37 (1980), 67–75.

[Sch72] W. M. Schmidt, On irregularities of distribution VII, Acta Arith. 21 (1972), 45–50.

[SiJ07] V. Sinescu, S. Joe, Good lattice rules based on the general weighted star discrepancy, Math. Comp. 76 (2007), 989–1004.

[SlJ94] I. H. Sloan, S. Joe, Lattice Methods for Multiple Integration, Oxford University Press, New York, 1994.

[SKJ02] I. H. Sloan, F. Y. Kuo, S. Joe, On the step-by-step construction of quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, Math. Comput. 71 (2002), 1609–1640.

[SlW98] I. H. Sloan, H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? J. Complexity 14 (1998), 1–33.

[Smo63] S. A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet Math. Dokl. 4 (1963), 240–243.

[Spa95] J. Spanier, Quasi-Monte Carlo methods for particle transport problems, in: H. Niederreiter, P. J.-S. Shiue (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, 121–148, Lecture Notes in Statistics 106, Springer, Berlin, 1995.

[Spe87] J. Spencer, Ten Lectures on the Probabilistic Method, SIAM, Philadelphia, 1987.

[SrS96] A. Srivastav, P. Stangier, Algorithmic Chernoff-Hoeffding inequalities in integer programming, Random Structures Algorithms 8 (1996), 27–58.

[Tal94] M. Talagrand, Sharper bounds for Gaussian and empirical processes, Ann. Probab. 22 (1994), 28–76.

[Thi00] E. Thiémard, Computing bounds for the star discrepancy, Computing 65 (2000), 169–186.

[Thi01a] E. Thiémard, An algorithm to compute bounds for the star discrepancy, J. Complexity 17 (2001), 850–880.

[Thi01b] E. Thiémard, Optimal volume subintervals with k points and star discrepancy via integer programming, Math. Meth. Oper. Res. 54 (2001), 21–45.

[Tra99] J. F. Traub, A continuous model of computation, Physics Today, May 1999, 39–44.

[TWW88] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, New York, 1988.

[TrW98] J. F. Traub, A. G. Werschulz, Complexity and Information, Cambridge University Press, Cambridge, 1998.

[War72] T. T. Warnock, Computational investigations of low-discrepancy point sets, in: S. K. Zaremba (Ed.), Applications of number theory to numerical analysis, 319–343, Academic Press, New York, 1972.

[WaW95] G. W. Wasilkowski, H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, J. Complexity 11 (1995), 1–56.

[Weg05] I. Wegener, Complexity theory. Exploring the limits of efficient algorithms, Springer-Verlag, Berlin, 2005.

[Wey16] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, Math. Ann. 77 (1916), 313–352.

[WiF97] P. Winker, K.-T. Fang, Application of threshold-accepting to the evaluation of the discrepancy of a set of points, SIAM J. Numer. Anal. 34 (1997), 2028–2042.

[Win07] C. Winzen, Approximative Berechnung der Sterndiskrepanz, Diploma Thesis, Department of Computer Science, Christian-Albrechts-Universität Kiel, Kiel, 2007.

[Woź91] H. Woźniakowski, Average case complexity of multivariate integration, Bull. Amer. Math. Soc. (N.S.) 24 (1991), 185–191.

[Woź94a] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, J. Complexity 10 (1994), 96–128.

[Woź94b] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, J. of Computing and Information 4 (1994), 1–19.

[Woź04] H. Woźniakowski, On generalized tractability for multivariate problems, in: M. Drinov (Ed.), Approximation Theory; a Volume Dedicated to Borislav Bojanov, 312–324, Acad. Publ. House, Sofia, 2004.

[Zar68] S. K. Zaremba, Some applications of multidimensional integration by parts, Ann. Pol. Math. 21 (1968), 85–96.

Part 2

Appendix