Entropy, Randomization, Derandomization, and Discrepancy

Michael Gnewuch

Abstract The star discrepancy is a measure of how uniformly distributed a finite point set is in the d-dimensional unit cube. It is related to high-dimensional numerical integration of certain function classes as expressed by the Koksma-Hlawka inequality. A sharp version of this inequality states that the worst-case error of approximating the integral of functions from the unit ball of some Sobolev space by an equal-weight cubature is exactly the star discrepancy of the set of sample points. In many applications, as, e.g., in physics, quantum chemistry or finance, it is essential to approximate high-dimensional integrals. Thus with regard to the Koksma-Hlawka inequality the following three questions are very important:
(i) What are good bounds with explicitly given dependence on the dimension d 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?
We want to discuss these questions and survey and explain some approaches to tackle them relying on metric entropy, randomization, and derandomization.

1 Introduction

Geometric discrepancy theory studies the uniformity of distribution of finite point sets. There are many different notions of discrepancies to measure quantitatively different aspects of "uniformity", see, e.g., [5, 16, 25, 58, 62, 68].

Michael Gnewuch

Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Christian-Albrechts-Platz 4, 24098 Kiel, Germany, e-mail: mig@informatik.uni-kiel.de

1.1 The Star Discrepancy

A particularly relevant measure is the star discrepancy, which is defined in the following way: Let $P \subset [0,1]^d$ be an n-point set. (We always want to understand an "n-point set" as a "multi-set": It consists of n points, but these points are not necessarily pairwise different.) For $x = (x_1, \ldots, x_d) \in [0,1]^d$ the *local discrepancy* $\Delta(x,P)$ of P in the axis-parallel box anchored at zero $[0,x) := [0,x_1) \times \cdots \times [0,x_d)$ (which we, likewise, simply want to call *test box*) is given by

$$\Delta(x,P) := \lambda_d([0,x)) - \frac{1}{n}|P \cap [0,x)|;$$

here λ_d denotes the d-dimensional Lebesgue measure and $|A|$ denotes the cardinality of a multi-set A. The *star discrepancy* of P is defined as

$$d_\infty^*(P) := \sup_{x \in [0,1]^d} |\Delta(x,P)|.$$

Further quantities of interest are the *smallest possible star discrepancy of any n-point set in* $[0,1]^d$

$$d_\infty^*(n,d) = \inf_{P \subset [0,1]^d; |P|=n} d_\infty^*(P),$$

and, for $\varepsilon \in (0,1)$, the *inverse of the star discrepancy*

$$n_\infty^*(\varepsilon,d) = \min\{n \in \mathbb{N} \,|\, d_\infty^*(n,d) \leq \varepsilon\}.$$

Although we mainly focus on the star discrepancy, we will also mention from time to time the L_p-star discrepancy of P for $1 \leq p < \infty$, which is defined by

$$d_p^*(P) := \left(\int_{[0,1]^d} |\Delta(x,P)|^p \, dx \right)^{1/p}.$$

1.2 Relation to Numerical Integration

Discrepancy notions are related to multivariate numerical integration. Such relations are put in a quantitative form by inequalities of Koksma-Hlawka- or Zaremba-type. Here we want to state a sharp version of the classical Koksma-Hlawka inequality [56, 51], which relates the star discrepancy to the worst-case error of quasi-Monte Carlo integration on certain function spaces. For other relations of discrepancy notions to numerical integration we refer the reader to the original papers [94, 93, 46, 82, 48, 47, 15, 67, 33] or the survey in [68, Chap. 9].

To state a sharp version of the Koksma-Hlawka inequality, 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 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 $f(x) = f_1(x_1)\cdots f_d(x_d)$, $f_1,\ldots,f_d \in H^{1,1}$. The space $H^{1,d}$ is then defined as the closure of $\otimes_{i=1}^d H^{1,1}$ with respect to the norm

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

where $\mathbf{1}$ denotes the vector $(1,\ldots,1)$ and f'_u is defined by

$$f'_u(x_u) = \frac{\partial^{|u|}}{\prod_{k\in u} \partial x_k} f(x_u, \mathbf{1}), \tag{2}$$

with $(x_u, \mathbf{1})_k = x_k$ if $k \in u$, and $(x_u, \mathbf{1})_k = 1$ otherwise. Then the following theorem holds:

Theorem 1. *Let $t^{(1)}, \ldots, t^{(n)} \in [0,1)^d$, and let I_d be the integration functional and $Q_{d,n}$ be the quasi-Monte Carlo cubature defined by*

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

Then the worst-case error $e^{\mathrm{wor}}(Q_{n,d})$ of $Q_{n,d}$ satisfies

$$e^{\mathrm{wor}}(Q_{n,d}) := \sup_{f\in H^{1,d};\, \|f\|_{1,d}=1} |I_d(f) - Q_{d,n}(f)| = d^*_\infty(t^{(1)},\ldots,t^{(n)}). \tag{3}$$

In particular, we obtain for all $f \in H^{1,d}$

$$|I_d(f) - Q_{d,n}(f)| \leq \|f\|_{1,d}\, d^*_\infty(t^{(1)},\ldots,t^{(n)}). \tag{4}$$

Theorem 1 is a corollary of a more general theorem proved by Hickernell, Sloan, and Wasilkowski in [47]. There the so-called L_∞-*same-quadrant discrepancy*, which covers the star discrepancy as a special case, 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 $\mathbf{1}$.

Particularly with regard to Theorem 1 the following three questions are very important.

Questions:

 (i) *What are good bounds with explicitly given dependence on the dimension d 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?*

Let us discuss the relevance of these questions for the star discrepancy. If we intend to approximate high-dimensional integrals of functions from $H^{1,d}$ by a quasi-Monte Carlo cubature $Q_{n,d}$, and if we wish to minimize the corresponding worst-case error $e^{\mathrm{wor}}(Q_{n,d})$, then Theorem 1 tells us that we have to minimize the star discrepancy of the set of integration points we want to use. For 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 does not grow faster than a polynomial of small degree in the dimension d.

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*:

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 randomized algorithm, which is a *semi-construction* in the sense of Novak and Woźniakowski [66]:

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 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.

In this bookchapter our main goal is to study the bracketing entropy of axis-parallel boxes anchored at zero and use the results, in particular upper bounds for the bracketing number and explicit constructions of bracketing covers of small size, to tackle question (i), (ii), and (iii).

Before we do so, we want to survey known bounds for the smallest possible star discrepancy, the problem of constructing small low-discrepancy samples, and known algorithms to calculate or approximate the star discrepancy of given point sets.

1.3 Known Bounds for the Star Discrepancy

We may distinguish two kinds of bounds for the smallest possible star discrepancy $d_\infty^*(n,d)$: *Asymptotic bounds* which describe the behavior of $d_\infty^*(n,d)$ well in the *asymptotic range*, i.e., for fixed dimension d and a large number of points n (which usually has to be exponential in d, see the discussion in Sect. 1.3.1), and *pre-asymptotic bounds* which describe its behavior well in the *pre-asymptotic range*, i.e., for moderate values of n (which depend at most polynomially on d).

Usually asymptotic bounds do not reveal the explicit dependence of $d_\infty^*(n,d)$ on d, while pre-asymptotic bounds exhibit the dependence of $d_\infty^*(n,d)$ on both parameters n and d. (Thus an alternative terminology might be "dimension-insensitive bounds" and "dimension-sensitive bounds".)

1.3.1 Asymptotic Bounds

For fixed dimension d the asymptotically best upper bounds for $d_\infty^*(n,d)$ that have been proved so far are of the form

$$d_\infty^*(n,d) \le C_d \ln(n)^{d-1} n^{-1}, \quad n \ge 2, \tag{5}$$

see, e.g., the original papers [40, 28, 65] or the monographs [5, 16, 25, 58, 62]. These bounds have been proved constructively, i.e., there are explicit constructions known that satisfy (5) for suitable constants C_d.

For $d = 1$ the set $T = \{1/2n, 3/2n, \ldots, (2n-1)/2n\}$ establishes (5) with $C_1 = 1/2$. For $d = 2$ the bound (5) can be derived from the results of Hardy and Littlewood [41] and of Ostrowski [72, 73] (the essential ideas can already be found in Lerch's paper [57]). For $d \ge 3$ the bound (5) was established by Halton, who showed in [40] that the Hammersley points exhibit this asymptotic behavior. The 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 [11, 12]. (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 [58, Sect. 1.1, 2.1].)

Looking at the asymptotic bound (5) it is natural to ask whether it is sharp or not. That it is optimal up to logarithmic factors is clear from the trivial lower bound $1/2n$. A better lower bound was shown by Roth in [76]:

$$d_\infty^*(n,d) \ge c_d \ln(n)^{\frac{d-1}{2}} n^{-1}, \quad n \ge 2. \tag{6}$$

In fact, Roth proved that the right hand side of (6) 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 [13], and for $d \ge 3$ by Roth himself [77, 78] and independently by Frolov [30]. Although Roth's lower bound is sharp for the L_2-star discrepancy, it is not optimal for the L_∞-star discrepancy. This was shown by Schmidt in

[79]. He established in dimension $d = 2$ the lower bound

$$d_\infty^*(n,2) \geq c_2 \ln(n) n^{-1}, \quad n \geq 2, \qquad (7)$$

and proved in this way that the upper bound (5) is optimal in dimension 2. In dimension $d \geq 3$ improvements of (6) were achieved by Beck [4], and later by Bilyk, Lacey, and Vagharshakyan [6, 7]; but although those improvements are deep mathematical results, their quantitative gain is rather modest. The remaining gap, baptized the "great open problem" by Beck and Chen in [5], 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 (5) give us no helpful information for moderate values of n, since $\ln(n)^{d-1} n^{-1}$ is an increasing function in n as long as $n \leq e^{d-1}$. This means that with respect to d we have to use at least exponentially many integration points to perceive any rate of decay of the right hand side of inequality (5). Additionally it is instructive to compare the convergence rate $n^{-1} \ln(n)^{d-1}$ and the Monte Carlo convergence rate $n^{-1/2}$: For example, in dimension $d = 3$ we have $n^{-1} \ln(n)^{d-1} < n^{-1/2}$ if $n \geq 5504$, but for $d = 10$ we already have $n^{-1} \ln(n)^{d-1} > n^{-1/2}$ for all $n \leq 1.295 \cdot 10^{34}$. Furthermore, point configurations satisfying (5) may 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

$$d_\infty^*(n,d) \leq \left(C_d' \ln(n)^{d-1} + o_d(\ln(n)^{d-1}) \right) n^{-1}$$

of (5) tends to zero as d approaches infinity, see, e.g., [2, 62, 65]. Here the o-notation with index d should emphasize that the implicit constant may depend on d; so far no good bounds for the implicit constant or, respectively, the constant C_d in (5), have been published.)

1.3.2 Pre-Asymptotic Bounds

A bound more suitable for high-dimensional integration was established by Heinrich, Novak, Wasilkowski and Woźniakowski [45], who proved

$$d_\infty^*(n,d) \leq c d^{1/2} n^{-1/2} \quad \text{and} \quad n_\infty^*(d,\varepsilon) \leq \lceil c^2 d \varepsilon^{-2} \rceil, \qquad (8)$$

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 [45] by a lower bound for $n_\infty^*(d,\varepsilon)$, which was later improved by Hinrichs [49] to

$$n_\infty^*(d,\varepsilon) \geq c_0 d \varepsilon^{-1} \quad \text{for } 0 < \varepsilon < \varepsilon_0, \qquad (9)$$

where $c_0, \varepsilon_0 > 0$ are suitable constants. The proof of (8) uses a large deviation bound of Talagrand for empirical processes [86] and an upper bound of Haussler for covering numbers of Vapnik-Červonenkis classes [42]. In particular, the proof is not constructive but probabilistic, and the proof approach does not provide an estimate

for the value of c. (Hinrichs presented a more direct approach to prove (8) with $c \leq 10$ at the Dagstuhl Seminar 04401 "Algorithms and Complexity for Continuous Problems" in 2004, but this result has not been published. Shortly after the submission of this book chapter Aistleitner gave a proof of (8) with $c \leq 10$ [1]. Since it relies on bracketing entropy and the bracketing covers we present in Section 2, we added a discussion of his approach in Section 3.2.1.)

In the paper [45] the authors proved also two slightly weaker bounds with explicitly known constants: The first one relies on upper bounds for the *average L_p-star discrepancy* for even p, the fact that the L_p-star discrepancy converges to the star discrepancy as p tends to infinity, and combinatorial arguments. For a detailed description of the approach, improvements, and closely related results we refer to [45, 34, 85].

Here we are more interested in the second bound from [45] with explicitly known small constants, which is of the form

$$d_\infty^*(n,d) \leq k d^{1/2} n^{-1/2} \big(\ln(d) + \ln(n) \big)^{1/2}, \tag{10}$$

and leads to

$$n_\infty^*(d,\varepsilon) \leq O(d\varepsilon^{-2}(\ln(d) + \ln(\varepsilon^{-1}))) \tag{11}$$

where essentially $k \approx 2\sqrt{2}$ and the implicit constant in the big-O-notation is known and independent of d and ε. The proof of (10) is probabilistic and relies on Hoeffding's large deviation bound. (A similar probabilistic approach was already used by Beck in [3] to prove upper bounds for other discrepancies.) From a conceptional point of view it uses *bracketing covers* (although in [45] the authors do not call them that way). As we will see later in Section 3.3, the probabilistic proof approach can actually be derandomized to construct point sets deterministically that satisfy the discrepancy bound (10).

1.4 Construction of Small Discrepancy Samples

On the one hand there are several construction methods known that provide point sets satisfying (5), and these constructions can be done quite efficiently. So one can construct, e.g., 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 (8) or (10), although random sets should do this with high probability. That it is not trivial to find such constructions was underlined by Heinrich, who pointed out in [44] that even the following easier problems are unsolved.

Problems:

(i) *For each $\varepsilon > 0$ and $d \in \mathbb{N}$, give a construction of a point set $\{t^{(1)}, \ldots, t^{(n)}\} \subset [0,1]^d$ with $n \leq c_\varepsilon d^{\kappa_\varepsilon}$ and $d_\infty^*(t^{(1)}, \ldots, t^{(n)}) \leq \varepsilon$, 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 $\{t^{(1)}, \ldots, t^{(n)}\} \subset [0,1]^d$ with $d_\infty^*(t^{(1)}, \ldots, t^{(n)}) \leq c d^\kappa n^{-\alpha}$, where c, κ and α are positive constants not depending on n or d.*

Although not stated explicitly in [44], these constructions are required to be efficiently executable, preferably in polynomial time in d, and ε^{-1} or n, respectively, see also [66, Open Problem 6]. If our ultimate goal is numerical integration, we may view the construction of low-discrepancy points as a precomputation. Since we can use the resulting integration points for the (efficient) evaluation of various integrands, we may still accept a little bit higher costs for the construction itself.

As stressed by Heinrich, it remains in particular a challenging question whether any of the various known classical constructions satisfies estimates like in problem (i) and (ii) or even the bound (8) or (10).

There had been attempts from computer scientists to construct small low-discrepancy samples [27, 9], 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 $(d/\ln(\varepsilon^{-1}))^{\ln(\varepsilon^{-1})}$ [9]. Formally, those constructions solve problem (i) (but not problem (ii)). Obviously, the size of the samples is a lower bound for the costs of the construction, which are therefore not polynomial in d and ε^{-1}.

We will discuss alternative constructions, based on bracketing covers and derandomization in Section 3.3.

1.5 Calculating the Star Discrepancy

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 [62], or whether sample sets are suitable for multivariate numerical integration of particular classes of integrands, cf. Theorem 1. Apart from that, it is particularly interesting with respect to question (ii) that the fast calculation or approximation of the star discrepancy would allow practicable semi-constructions of low-discrepancy samples of moderate size.

It is known that the L_2-star discrepancy of a given n-point set in dimension d can be calculated via Warnock's formula [91] with $O(dn^2)$ arithmetic operations and similar formulas hold for weighted versions of the L_2-star discrepancy. Heinrich and Frank developed an asymptotically faster algorithm for the L_2-star discrepancy using only $O(n \log(n)^{d-1})$ operations for fixed d [29, 43]. (Due to the exponent of the log-term, the algorithm is only practicable in low dimensions.)

What methods are known to calculate or approximate the star discrepancy of a given set 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. But for calculating the discrepancy of P exactly it suffices to consider only finitely many test boxes. So if $P = \{p^{(1)}, \ldots, p^{(n)}\} \subset [0,1)^d$, let us define

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

and let us put

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

Then it is not hard to verify that

$$d_\infty^*(P) =$$
$$\max \left\{ \max_{y \in \bar{\Gamma}(P)} \left(\lambda_d([0, y)) - \frac{|P \cap [0, y)|}{n} \right) , \ \max_{y \in \Gamma(P)} \left(\frac{|P \cap [0, y]|}{n} - \lambda_d([0, y)) \right) \right\}, \tag{12}$$

for a proof see, e.g., [38]. Thus we need to consider at most $O(n^d)$ test boxes to compute $d_\infty^*(P)$. For a random n-point set P we have almost surely $|\Gamma(P)| = n^d$, resulting in $\Omega(n^d)$ test boxes that we have to take into account to calculate (12). This underlines that (12) 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 (12) simplifies to

$$d_\infty^*(P) = \frac{1}{2n} + \max_{1 \leq i \leq n} \left| p^{(i)} - \frac{2i-1}{2n} \right|,$$

a result due to Niederreiter, see [60, 61].

In dimension $d = 2$ a reduction of the number of steps to calculate (12) was achieved by de Clerck [10]. In [8] her formula was slightly extended and simplified by Bundschuh and 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 [8, Thm. 1] states that

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

Bundschuh and Zhu provided also a corresponding formula for the three-dimensional case. The method can be generalized to arbitrary dimension d and requires roughly $O(n^d/d!)$ elementary operations. This method was, e.g., used in [92] to calculate the exact discrepancy of particular point sets, so-called (shifted) rank-1 lattice rules (cf. [81]), up to size $n = 236$ in dimension $d = 5$ and to $n = 92$ in dimension $d = 6$. But as pointed out by Winker and Fang [92], 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 Dobkin, Eppstein, and Mitchell in [17]. It uses sophisticated, but complicated data structures, and the authors implemented only asymptotically slightly slower variants of the algorithm 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.

Indeed, the problem of calculating the star discrepancy is *NP-hard*, as was proved in [38]. We will briefly outline the main proof ideas below in this section. In [32] Giannopoulos, Knauer, Wahlström, and Werner proved a result on the *parametrized complexity* of the problem of calculating the star discrepancy, namely they showed that it is $W[1]$-*hard* with respect to the parameter d. It follows from [32] that the general problem cannot be solved in time $O(n^{o(d)})$ unless the *exponential time hypothesis* is false, which is widely regarded as extremely unlikely.

Notice that the complexity results above are about the exact calculation of the discrepancy of arbitrary point sets; they do not directly address the complexity of approximating the discrepancy. So what is known about approximation algorithms?

Since in high dimension no efficient algorithm for the exact calculation of the star discrepancy is known, some authors tried to tackle the large scale enumeration problem (12) by using optimization heuristics. In [92] Winker and Fang used *threshold accepting* [26], a refined randomized local search algorithm based on a similar idea as the well-known simulated annealing algorithm [55], to find lower bounds for the star discrepancy. The algorithm performed well in numerical tests on (shifted) rank-1 lattice rules.

In [89] 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.

The key observation to approach the non-linear expression (12) 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 $[0,y)$, $y \in \overline{\Gamma}(P)$, containing k points, $k \in \{0,1,\ldots,n-1\}$, and the smallest boxes $[0,y]$, $y \in \Gamma(P)$, containing ℓ points for $\ell \in \{1,\ldots,n\}$. Thiémard conjectured these sub-problems to be NP-hard.

The conjecture of Thiémard is proved rigorously in [38] 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., [53, 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 BALANCED SUBGRAPH to the minimal volume subinterval problems, respectively; the graph theoretical decision problems DOMINATING SET and BALANCED SUBGRAPH are known to be NP-hard, see [31, 54]. 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 minor errors occurring in [89] are listed in [38]. Since those errors may lead to incorrect solutions of Thiémard's algorithm for certain instances, it is explained how to avoid their undesired consequences.)

A *genetic algorithm* to approximate the star discrepancy was recently proposed by Shah [80].

In the recent paper [39] a new randomized algorithm to approximate the star discrepancy based on threshold accepting was presented. Comprehensive numerical tests indicate that it improves on the algorithms from [92, 89, 80], especially in higher dimension $20 \leq d \leq 50$.

All the approximation algorithms we have mentioned so far have shown their usefulness in practice, but unfortunately none of them provides an approximation guarantee.

An approach that approximates the star discrepancy of a given set P up to a user-specified error δ was presented by Thiémard [87, 88]. It is in principle based on the generation of small bracketing covers (which were not named this way in [87, 88]).

2 Bracketing Entropy

In this section we want to study the bracketing entropy of axis-parallel boxes anchored at zero. We start by introducing the necessary notion.

2.1 Basic Definitions

Definition 1. Let $x, y \in [0,1]^d$ with $x_i \leq y_i$ for $i = 1, \ldots, d$. We assign a *weight* $W([x,y])$ to the closed box $[x,y] := [x_1, y_1] \times \cdots \times [x_d, y_d]$ by

$$W([x,y]) = \lambda_d([0,y]) - \lambda_d([0,x]).$$

Let $\delta > 0$. The box $[x,y]$ is a δ-*bracket* if $W([x,y]) \leq \delta$. A set $\mathscr{B}$ of δ-brackets whose union covers $[0,1]^d$ is a δ-*bracketing cover* of $[0,1]^d$. The *bracketing number* $N_{[]}(d, \delta)$ denotes the smallest cardinality of any δ-bracketing cover of $[0,1]^d$. Its logarithm $\ln(N_{[]}(d, \delta))$ is the *bracketing entropy* (or *entropy with bracketing*).

The notion of bracketing entropy is well established in empirical process theory, see, e.g., [86, 90]. In some places it will be more convenient for us to use the related notion of δ-covers from [21] instead of the notion of bracketing covers.

Definition 2. Let $\delta > 0$. A finite set Γ is a δ-*cover* of $[0,1]^d$ if for all $y \in [0,1]^d$ there exist $x, z \in \Gamma \cup \{0\}$ such that $[x,z]$ is a δ-bracket and $y \in [x,z]$. Let $N(d, \delta)$ denote the smallest cardinality of any δ-cover of $[0,1]^d$.

If, on the one hand, we have a δ-bracketing cover $\mathscr{B}$, then it is easy to see that

$$\Gamma_{\mathscr{B}} := \{x \in [0,1]^d \setminus \{0\} \mid \exists y \in [0,1]^d : [x,y] \in \mathscr{B} \text{ or } [y,x] \in \mathscr{B}\} \qquad (13)$$

is a δ-cover. If, on the other hand, Γ is a δ-cover, then

$$\mathscr{B}_\Gamma := \{[x,y] \,|\, x,y \in \Gamma \cup \{0\}, [x,y] \text{ is a } \delta\text{-bracket}, x \neq y\}$$

is a δ-bracketing cover. Therefore we have

$$N(d,\delta)+1 \leq 2N_{[]}(d,\delta) \leq (N(d,\delta)+1)N(d,\delta). \tag{14}$$

(The second inequality is obviously a weak one, and it would be nice to have a tighter bound.) The bracketing number and the quantity $N(d,\delta)$ are related to the *covering* and the *L_1-packing number*, see, e.g., [21, Rem. 2.10].

2.2 Construction of Bracketing Covers

How large is the bracketing entropy and how does a small δ-bracketing cover look like? To get some idea, we have a look at some examples of δ-bracketing covers.

2.2.1 Cells of an Equidistant Grid

To prove (10), Heinrich et al. used in [45] a δ-cover in form of an equidistant grid $E_\delta = \{0,1/m,2/m,\ldots,1\}^d$ with $m = \lceil d/\delta \rceil$. The grid cells, i.e., all closed boxes of the form $[x,x^+]$, where $x_i \in \{0,1/m,\ldots,1-1/m\}$ and $x_i^+ = x_i + 1/m$ for $i \in \{1,\ldots,d\}$, form a δ-bracketing cover $\mathscr{E}_\delta$. Indeed, the grid cell with the largest weight is $[(1-1/m)\mathbf{1},\mathbf{1}]$ with

$$W([(1-1/m)\mathbf{1},\mathbf{1}]) = 1-(1-1/m)^d \leq d/m \leq \delta.$$

The cardinality of the δ-bracketing cover $\mathscr{E}_\delta$ is clearly

$$|\mathscr{E}_\delta| = m^d \leq (d\delta^{-1}+1)^d. \tag{15}$$

Although the weight of the grid cell $[(1-1/m)\mathbf{1},\mathbf{1}]$ is nearly δ, the weights of most of the other grid cells are reasonably smaller than δ. For example, the weight of the cell $[0,(1/m)\mathbf{1}]$ is $(1/m)^d \leq (\delta/d)^d$, which is for $d \geq 2$ much smaller than δ.

2.2.2 Cells of a Non-Equidistant Grid

We generate a smaller δ-bracketing cover by using a *non-equidistant grid* Γ_δ of the form

$$\Gamma_\delta = \{\gamma_0,\ldots,\gamma_{\kappa(\delta,d)}\}^d, \tag{16}$$

where $\gamma_0,\gamma_1,\ldots,\gamma_{\kappa(\delta,d)}$ is a decreasing sequence in $(0,1]$. We calculate this sequence recursively in the following way (cf. Figure 1):

We set $\gamma_0 := 1$ and choose $\gamma_1 \in (0,1)$ such that $y^{(0)} := \gamma_1\mathbf{1}$ and $z^{(0)} := \mathbf{1}$ satisfy $W([y^{(0)},z^{(0)}]) = \delta$. Obviously, $\gamma_1 = (1-\delta)^{1/d}$. Let γ_i be calculated. If $\gamma_i > \delta$, we

compute the real number $\gamma_{i+1} \in (0, \gamma_i)$ that ensures that $y^{(i)} := (\gamma_{i+1}, \gamma_1, ..., \gamma_1)$ and $z^{(i)} := (\gamma_i, 1, ..., 1)$ satisfy $W([y^{(i)}, z^{(i)}]) = \delta$. If $\gamma_i \leq \delta$, then we put $\kappa(\delta, d) := i$ and stop. From the geometrical setting it is easy to see that $\gamma_0, \gamma_1, ...$ is a decreasing sequence with $\gamma_i - \gamma_{i+1} \leq \gamma_{i+1} - \gamma_{i+2}$. Therefore $\kappa(\delta, d)$ is finite.
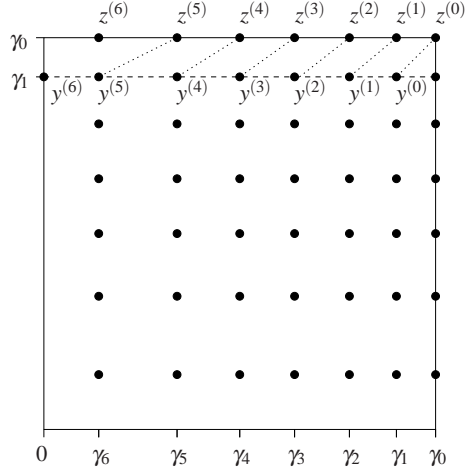
Fig. 1 Construction of the non-equidistant grid Γ_δ for $d = 2$ and $\delta = 0.2$. Here, $\kappa(\delta, d) = 6$.

The following result was proved in [21, Thm. 2.3].

Theorem 2. *Let $d \geq 2$, and let $0 < \delta < 1$. Let $\Gamma_\delta = \{\gamma_0, \gamma_1, \ldots, \gamma_{\kappa(\delta, d)}\}^d$ be as in (16). Then Γ_δ is a δ-cover of $[0, 1]^d$, and consequently*

$$N(d, \delta) \leq |\Gamma_\delta| \leq (\kappa(\delta, d) + 1)^d, \tag{17}$$

where

$$\kappa(\delta, d) = \left\lceil \frac{d}{d-1} \frac{\ln(1 - (1 - \delta)^{1/d}) - \ln(\delta)}{\ln(1 - \delta)} \right\rceil. \tag{18}$$

The inequality $\kappa(\delta, d) \leq \left\lceil \frac{d}{d-1} \frac{\ln(d)}{\delta} \right\rceil$ holds, and the quotient of the left and the right hand side of this inequality converges to 1 as δ approaches 0.

From the δ-cover Γ_δ we obtain a δ-bracketing cover $\mathcal{G}_\delta$ by taking the grid cells of the form $[y, y^+]$, where $y_i = \gamma_j$ for some $j = j(i) \in \{1, \ldots, \kappa(\delta, d)\}$ and $y_i^+ = \gamma_{j-1}$ for all $i \in \{1, \ldots, d\}$, and the d brackets of the form $[0, z]$ with z having $d - 1$ coordinates equal to 1 and one coordinate equal to $\gamma_{\kappa(\delta, d)}$. Thus

$$|\mathcal{G}_\delta| = \kappa(\delta, d)^d + d = \left(\frac{d}{d+1} \right)^d \ln(d)^d \delta^{-d} + O_d\left(\delta^{-d+1} \right); \tag{19}$$

the last identity follows from

$$\kappa(\delta,d) = \frac{d}{d-1}\ln(d)\delta^{-1} + O_d(1) \quad \text{as } \delta \text{ approaches } 0,$$

see [36, Sect. 2]. Note that $(\frac{d}{d-1})^d$ is bounded above by 4 and converges to e as d tends to infinity.

2.2.3 A Layer Construction

By construction the brackets $[y^{(i)}, z^{(i)}]$, $i = 0, 1, \ldots, \kappa(\delta,d)-1$, satisfy $W([y^{(i)}, z^{(i)}]) = \delta$, but it can be shown that the weights of the brackets $[v,w]$ in $\mathscr{G}_\delta$, with $w_i < 1$ for more than one index $i \in \{1, \ldots, d\}$, are strictly smaller than δ. It seems obvious that a suitable δ-bracketing cover consisting almost exclusively of brackets with weights exactly δ should exhibit a smaller cardinality than $\mathscr{G}_\delta$. We outline here a construction $\mathscr{Z}_\delta$ which satisfies this specification. To simplify the representation, we confine ourselves to the case $d = 2$ and refer to [35] for a generalization of the construction to arbitrary dimension d. Let δ be given. The essential idea is the following:

We define $a_i = a_i(\delta) := (1 - i\delta)^{1/2}$ for $i = 0, \ldots, \zeta = \zeta(\delta) := \lceil \delta^{-1} \rceil - 1$, and $a_{\zeta+1} := 0$. We decompose $[0,1]^2$ into layers

$$L^{(i)}(\delta) := [0, a_i\mathbf{1}] \setminus [0, a_{i+1}\mathbf{1}), \quad i = 0, \ldots, \zeta,$$

and cover each layer separately with δ-brackets. To cover $L^{(0)}(\delta)$, we can simply use the δ-brackets $[y^{(i)}, z^{(i)}]$, $i = 0, 1, \ldots, \kappa(\delta,2)-1$, from our previous construction and the δ-brackets we obtain after permuting the first and second coordinates of $y^{(i)}$ and $z^{(i)}$, respectively. To cover the remaining layers, we observe that the brackets $[a_{i+1}\mathbf{1}, a_i\mathbf{1}]$, $i = 1, \ldots, \zeta-1$, all have weight δ, and we can cover the layers $L^{(i)}(\delta)$, $i = 1, \ldots, \zeta-1$, by a straightforward modification of the procedure we used to cover $L^{(0)}(\delta)$.

The final layer $L^{(\zeta)}(\delta) = [0, a_\zeta\mathbf{1}]$ is trivially covered by the δ-bracket $[0, a_\zeta\mathbf{1}]$ itself. Figure 2 shows the resulting bracketing cover $\mathscr{Z}_\delta$ for $\delta = 0.075$.

As shown in [36, Prop. 4.1], the two-dimensional δ-bracketing cover $\mathscr{Z}_\delta$ satisfies

$$|\mathscr{Z}_\delta| = 2\ln(2)\delta^{-2} + O(\delta^{-1}). \tag{20}$$

Notice that the coefficient $2\ln(2) \approx 1.3863$ in front of δ^{-2} is smaller than the corresponding coefficient $(2\ln(2))^2 \approx 1.9218$ in (19).

2.2.4 An Essentially Optimal Construction

The layer construction was generated in a way to guarantee that all δ-brackets have weight exactly δ (except of maybe those which intersect with the coordinate axes). To minimize the number of brackets needed to cover $[0,1]^2$, or, more generally, $[0,1]^d$, it seems to be a good idea to find brackets with weight δ that exhibit max-
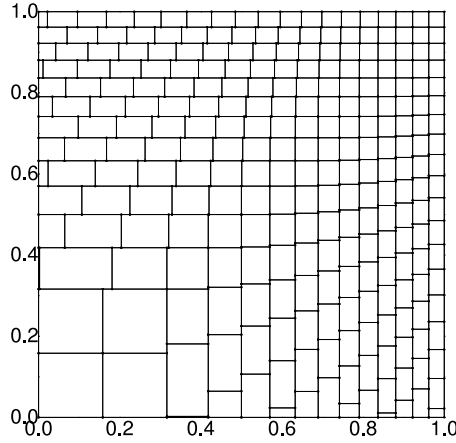
Fig. 2 The layer construction $\mathscr{Z}_\delta$ for $\delta = 0.075$.

imum volume. The following lemma [35, Lemma 1.1] shows how such δ-brackets look like.

Lemma 1. *Let $d \geq 2$, $\delta \in (0,1]$, and let $z \in [0,1]^d$ with $\lambda_d([0,z]) = z_1 \cdots z_d \geq \delta$. Put*

$$x = x(z,\delta) := \left(1 - \frac{\delta}{z_1 \cdots z_d}\right)^{1/d} z. \tag{21}$$

Then $[x,z]$ is the uniquely determined δ-bracket having maximum volume of all δ-brackets containing z. Its volume is

$$\lambda_d([x,z]) = \left(1 - \left(1 - \frac{\delta}{z_1 \cdots z_d}\right)^{1/d}\right)^d \cdot z_1 \cdots z_d.$$

A positive aspect of the previous construction $\mathscr{Z}_\delta$ is that (essentially) all its brackets have largest possible weight δ and overlap only on sets of Lebesgue measure zero. But if we look at the brackets in $\mathscr{Z}_\delta$ which are close to the first or the second coordinate axis and away from the main diagonal, then these boxes do certainly not satisfy the "maximum area criterion" stated in Lemma 1. The idea of the next construction is to generate a bracketing cover $\mathscr{R}_\delta$ similarly as in the previous section, but to "re-orientate" the brackets from time to time in the course of the algorithm to enlarge the area which is covered by a single bracket. Of course this procedure should not lead to too much overlap of the generated brackets. Let us explain the underlying geometrical idea of the construction:

Like all the constructions we have discussed so far, our new bracketing cover should be symmetric with respect to both coordinate axes. Thus we only have to state explicitly how to cover the subset

$$H := \{(x,y) \in [0,1]^2 \,|\, x \leq y\}$$

of $[0,1]^2$. For a certain number $p = p(\delta)$ we subdivide H into sectors

$$T^{(h)} := \left\{ (x,y) \in H \setminus \{(0,0)\} \,\middle|\, \frac{h-1}{2^p} \leq \frac{x}{y} \leq \frac{h}{2^p} \right\} \cup \{(0,0)\}, \quad h = 1,\ldots,2^p.$$

We start with $T^{(2^p)}$ and cover this subset of $[0,1]^2$ in the same way as we covered it in the construction $\mathcal{Z}_\delta$, i.e., we decompose $T^{(2^p)}$ into horizontal stripes $[(0,a_{i+1}),(a_i,a_i)] \cap T^{(2^p)}$, $i = 0,1,\ldots,\zeta = \lceil \delta^{-1} \rceil - 1$, and cover each stripe separately with δ-brackets whose weights are (except of maybe one bracket per stripe) exactly δ. Notice that the δ-brackets of $\mathcal{Z}_\delta$ that cover the main diagonal of $[0,1]^2$ are volume optimal due to Lemma 1. Hence, if we choose p sufficiently large, the sector $T^{(2^p)}$ will be thin and all the δ-brackets we use to cover it will have nearly maximum volume.

If $p = 0$, then $H = T^{(2^p)}$ and our new construction $\mathcal{R}_\delta$ will actually be equal to $\mathcal{Z}_\delta$. If $p > 0$, then we have additional sectors $T^{(1)},\ldots,T^{(2^p-1)}$. Again, for a given $i \in \{1,\ldots,2^p - 1\}$ we decompose $T^{(i)}$ into horizontal stripes, but the vertical heights of the stripes increases as i decreases. We essentially choose the heights of each stripe in a way that the bracket on the right hand side of the stripe having this heights and weight exactly δ exhibits maximum volume. Thus, if the sector $T^{(i)}$ is sufficiently thin, again essentially all δ-brackets that cover it will have nearly maximum volume. Therefore we should choose $p = p(\delta)$ large enough. On the other hand, we usually will have overlapping δ-brackets at the common boundary of two sectors. To minimize the number of brackets needed to cover H (and thus $[0,1]^2$), we should try to avoid too much overlap of brackets and consequently not choose p too large. Since $T^{(2^p)}$ has the most horizontal stripes of all sectors $T^{(i)}$, namely $\lceil \delta^{-1} \rceil$, a choice satisfying $2^{p(\delta)} = o(\delta^{-1})$ ensures that the overlap has no impact on the coefficient in front the most significant term δ^{-2} in the expansion of $|\mathcal{R}_\delta|$ in terms of δ^{-1}. Figure 3 and 4 show bracketing covers $\mathcal{R}_\delta$ based on this idea constructed in [36] for $\delta = 0.075$ and $\delta = 0.03$. The parameter p was chosen to be

$$p = p(\delta) = \left\lfloor \frac{\ln(\delta^{-1})}{1.7} \right\rfloor.$$

The figures show the overlapping of brackets at the common boundaries of different sectors. Note in particular that the 16 squares near the origin in Figure 4 are not individual δ-brackets with weight δ – these squares just occur since larger brackets intersect near the origin.
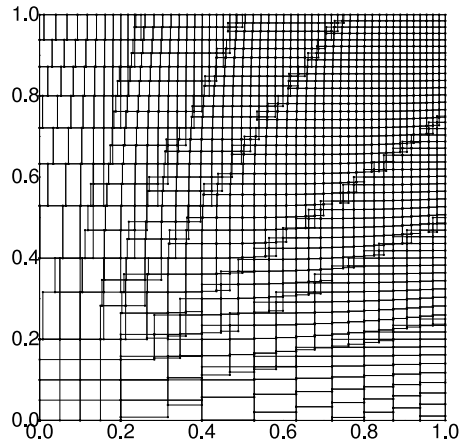
For all technical details of the δ-bracketing cover $\mathcal{R}_\delta$ of $[0,1]^2$ we refer to [36]. As shown there in Proposition 5.1, its size is of order

$$|\mathcal{R}_\delta| = \delta^{-2} + o(\delta^{-2}) \tag{22}$$
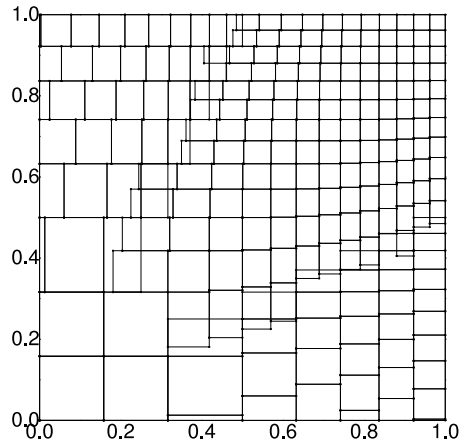
Fig. 3 The essentially optimal construction $\mathscr{R}_\delta$ for $\delta = 0.075$.

Fig. 4 The essentially optimal construction $\mathscr{R}_\delta$ for $\delta = 0.03$.

as long as $p = p(\delta)$ is a decreasing function on $(0,1)$ with $\lim_{\delta \to 0} p(\delta) = \infty$ and $2^p = o(\delta^{-1})$ as δ tends to zero.

The construction $\mathscr{R}_\delta$ is (essentially) optimal, as will be shown by a lower bound in the next section.

2.3 Bounds for the Bracketing Number

Here we state bounds for the bracketing number for arbitrary dimension d.

Theorem 3. *Let d be a positive integer and $0 < \delta \leq 1$. Then we have the following two upper bounds on the bracketing number:*

$$N_{[]}(d,\delta) \leq \frac{d^d}{d!}\delta^{-d} + O_d(\delta^{-d+1}) \tag{23}$$

and

$$N_{[]}(d,\delta) \leq 2^{d-1}\frac{d^d}{d!}\left(\delta^{-1} + 1\right)^d. \tag{24}$$

Both bounds were proved constructively in [35] by a δ-bracketing cover which can be seen as d-dimensional generalization of the two-dimensional construction $\mathscr{Z}_\delta$ from Section 2.2.3. In the same paper the following lower bound for the bracketing number was shown, see [35, Thm. 1.5].

Theorem 4. *For $d \geq 2$ and $0 < \delta \leq 1$ there exist a constant c_d which may depend on d, but not on δ, with*

$$N_{[]}(d,\delta) \geq \delta^{-d}(1 - c_d\delta). \tag{25}$$

The proof of Theorem 4 is based on the fact that the bracketing number $N_{[]}(d,\delta)$ is bounded from below by the average of $[\lambda_d(B_\delta(x))]^{-1}$ over all $x \in [0,1]^d$, where $B_\delta(x)$ is a δ-bracket containing x with maximum volume.

The lower bound shows that the upper bound $N_{[]}(2,\delta) \leq \delta^{-2} + o(\delta^{-2})$, resulting from the bound (22) on the cardinality of $\mathscr{R}_\delta$ from Section 2.2.4, is (essentially) optimal.

3 Application of Bracketing to Discrepancy

We want to discuss how the results about bracketing covers and bracketing entropy from the last section can be used to tackle the three questions from Section 1.2. We start with question (iii), where our results are most directly applicable.

3.1 Approximation of the Star Discrepancy

Bracketing covers can be used to approximate the star discrepancy by exploiting the following *approximation property*.

Lemma 2. *Let $\mathscr{B}$ be a bracketing cover of $[0,1]^d$, and let $\Gamma_{\mathscr{B}}$ as in (13). For finite subsets P of $[0,1]^d$ put*

$$d_\Gamma^*(P) := \max_{x \in \Gamma_{\mathscr{B}}} |\Delta(x,P)|. \tag{26}$$

Then we have

$$d_\Gamma^*(P) \leq d_\infty^*(P) \leq d_\Gamma^*(P) + \delta.$$

The proof is straightforward, but can also be found in, e.g., [21, Lemma 3.1].

The essential idea of Thiémard's algorithm from [87, 88] is to generate for a given point set P and a user-specified error δ a small δ-bracketing cover $\mathscr{B} = \mathscr{B}_\delta$ of $[0,1]^d$ and to approximate $d^*_\infty(P)$ by $\max_{x \in \Gamma_{\mathscr{B}}} |\Delta(x,P)|$.

The costs of generating $\mathscr{B}_\delta$ are of order $\Theta(d|\mathscr{B}_\delta|)$. If we count the number of points in $[0,x)$ for each $x \in \Gamma_{\mathscr{B}}$ in a naive way, this results in an overall running time of $\Theta(dn|\mathscr{B}_\delta|)$ for the whole algorithm. As Thiémard pointed out in [88], this orthogonal range counting can be done in moderate dimension d more effectively by employing data structures based on so-called range trees. This approach reduces the time $O(dn)$ per test box that is needed for the naive counting to $O(\log(n)^d)$. Since a range tree for n points can be generated in $O(C^d n \log(n)^d)$ time, $C > 1$ some constant, this results in an overall running time of

$$O((d + \log(n)^d)|\mathscr{B}_\delta| + C^d n \log(n)^d).$$

For the precise details of the implementation we refer to [88].

The upper bounds on the running time of the algorithm show that smaller δ-bracketing covers $\mathscr{B}_\delta$ will lead to shorter running times. But since the lower bound (25) implies

$$|\mathscr{B}_\delta| \geq \delta^{-d}(1 - c_d \delta),$$

even the time for generating a δ-bracketing cover $\mathscr{B}_\delta$ is bounded from below by $\Omega(d\delta^{-d})$, and this is obviously also a lower bound for the running time of the whole algorithm. This shows that the approach of Thiémard has practical limitations. Nevertheless, it is useful in moderate dimensions as was reported, e.g., in [23] or [70].

The smallest bracketing covers used by Thiémard are different from the constructions we presented in the previous section, see [88]. Figure 5 shows his construction $\mathscr{T}_\delta$ in dimension $d = 2$ for $\delta = 0.075$.
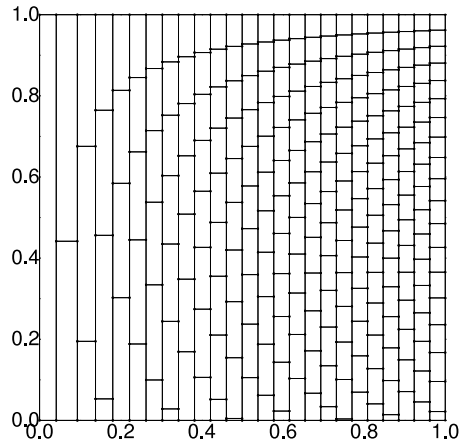
He proved the upper bound

$$|\mathscr{T}_\delta| \leq \binom{d+h}{d}, \quad \text{where} \quad h = \left\lceil \frac{d \ln(\delta)}{\ln(1-\delta)} \right\rceil.$$

This leads to

$$|\mathscr{T}_\delta| \leq e^d \left(\frac{\ln \delta^{-1}}{\delta} + 1 \right)^d,$$

a weaker bound than $|\mathscr{B}_\delta| \leq e^d \delta^{-d} + O_d(\delta^{-d+1})$ and $|\mathscr{B}_\delta| \leq 2^{d-1} e^d (\delta^{-1} + 1)^d$ which hold for the construction $\mathscr{B}_\delta$ that established Theorem 3.

For $d = 2$ the bound $|\mathscr{T}_\delta| = 2\ln(2)\delta^{-2} + O(\delta^{-1})$ was proved in [36], which shows that in two dimensions the quality of $\mathscr{T}_\delta$ is similar to the one of the layer construction $\mathscr{Z}_\delta$ that we presented in the section 2.2.3.

Fig. 5 Thiémard's construction $\mathcal{T}_\delta$ for $\delta = 0.075$.

3.2 Pre-Asymptotic Bounds via Randomization

Here we want discuss question (i) from Section 1.2. We distinguish between deterministic discrepancy bounds for n-point samples in $[0,1]^d$ and for d-dimensional projections of infinite sequences of points with infinitely many coordinates. Furthermore, we mention briefly probabilistic discrepancy bounds for hybrid-Monte Carlo sequences.

3.2.1 Point Sets in the d-Dimensional Unit Cube

Probabilistic pre-asymptotic bounds on the smallest possible star discrepancy of any n-point set in $[0,1]^d$ can be proved in three steps:

Probabilistic Proof Scheme:

1. We discretize the star discrepancy at the cost of an approximation error at most δ. More precisely, we use a δ-bracketing cover $\mathcal{B}$ and consider for a point set P instead of $d_\infty^*(P)$ its approximation $d_\Gamma^*(P)$ defined in (26), where $\Gamma = \Gamma_\mathcal{B}$ is as in (13).
2. We perform a random experiment that results in a random n-point set P in $[0,1]^d$ that fails to satisfy the events $\{|\Delta(x,P)| \leq \delta\}$, $x \in \Gamma_\mathcal{B}$, with small probability. If the random experiment is subject to the concentration of measure phenomenon, then these "failing probabilities" can be controlled with the help of large deviation bounds.
3. Since the event $\{d_\Gamma^*(P) > \delta\}$ is the union of the events $\{|\Delta(x,P)| > \delta\}$, $x \in \Gamma_\mathcal{B}$, a simple union bound shows that P satisfies $d_\Gamma^*(P) \leq \delta$ with positive probability if $\mathbb{P}\{|\Delta(x,P)| > \delta\} < |\Gamma_\mathcal{B}|^{-1}$ for all $x \in \Gamma_\mathcal{B}$.

Then for $\varepsilon = 2\delta$ there exists an n-point set P with $d_\infty^*(P) \le d_\Gamma^*(P) + \delta \le \varepsilon$. The aim is to choose ε as small as possible.

To keep the loss caused by the union bound small, the size of the δ-bracketing cover $\mathscr{B}$ (or the δ-cover $\Gamma_{\mathscr{B}}$, respectively) should be chosen as small as possible. To receive a bound for the star discrepancy with explicit constants, bounds with explicit constants are needed for the size of the δ-bracketing cover used.

The bound (10) from [45] was proved in this way: The δ-cover Γ was chosen to be the equidistant grid from Section 2.2.1 and the random experiment was to distribute n points uniformly and independently in $[0,1]^d$. The "failing probability" in each single test box was bounded above by *Hoeffding's large deviation bound* [52], which reads as follows:

Let $X_1,\dots,X_n$ be independent random variables with $a_i \le X_i \le b_i$ for all i. Then for all $\delta > 0$

$$\mathbb{P}\left\{\frac{1}{n}\left|\sum_{k=1}^{n}(X_i - \mathbb{E}(X_i))\right| \ge \delta\right\} \le 2\exp\left(-\frac{2\delta^2 n^2}{\sum_{k=1}^{n}(b_i - a_i)^2}\right).$$

Using the same probabilistic experiment and again Hoeffding's large deviation bound, but instead of the bracketing cover from Section 2.2.1 the one that implied the estimate (24), one obtains the improved discrepancy bound

$$d_\infty^*(n,d) \le k' d^{1/2} n^{-1/2} \ln\left(1 + \frac{n}{d}\right)^{1/2} \tag{27}$$

(here we have essentially $k' \approx \sqrt{2}$, see [35, 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 (27) behaves asymptotically as the bound (8). In fact, if (8) holds with $c = 10$ (as claimed by Hinrichs and recently published by Aistleitner), then the bound [35, (22)], a version of (27), is still better than (8) for all $n \le 1.5 \cdot e^{95} d$. Actually, we may use the upper bound in (24) to reprove (8) 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 [86, Thm. 1.1]. (As we discuss at the end of this subsection, Aistleitner's approach to prove (8) with a constant $c \le 10$ indeed uses the upper bound (24).)

For other discrepancy notions similar approaches, relying on uniformly and independently distributed random points, were used to prove pre-asymptotic bounds with explicitly given constants. This was done, e.g., for the *same-quadrant discrepancy* [47], discrepancies with respect to *ellipsoids*, *stripes*, and *spherical caps* in $\mathbb{R}^d$ [59], the *extreme discrepancy* [35], and the *weighted star discrepancy* [50].

One can modify the probabilistic experiment by using, e.g., the variance reduction technique *stratified sampling*. If, e.g., $n = v^d$, then one can subdivide $[0,1]^d$ into n subcubes of the same size and distribute in each subcube one point uniformly

at random (and independently from the other points). This experiment was used in [20, Thm. 4.3] (a preprint version of [21]) to derive

$$d_\infty^*(n,d) \leq k'' d n^{-\frac{1}{2}-\frac{1}{2d}} \ln(n)^{1/2}. \tag{28}$$

(Again, we have essentially $k'' \approx \sqrt{2}$. The proof used the δ-cover Γ_δ from (16).)

For the discrepancy of *tilted boxes* and of *balls* with respect to probability measures on $[0,1]^d$ which are absolutely continuous with respect to λ_d, a similar approach relying on a stratified sampling technique was used by Beck in [3] to prove asymptotic probabilistic upper bounds. But these bounds do not exhibit the dependence on the dimension; in particular, the involved constants are not explicitly known.

We will discuss a further random experiment in more detail in Section 3.3.

Let us finish this subsection with the discussion of the recent result of Aistleitner, who proved in [1] that the constant c in (8) is smaller than 10. As in the probabilistic proof scheme stated above, his approach starts by discretizing the star discrepancy at the cost of an approximation error $\delta = 2^{-K}$, where $K \approx -\log_2(d/n)/2$. The underlying probabilistic experiment is to distribute n points $p^{(1)}, \ldots, p^{(n)}$ uniformly and independently in $[0,1]^d$. An important observation is now that for measurable subsets A of $[0,1]^d$ the variance of the random variables $\xi_A^{(i)} := \lambda_d(A) - |\{p^{(i)}\} \cap A|$, $i = 1, \ldots, n$, depends strongly on the volume $\lambda_d(A)$ of A:

$$\text{Var}(\xi_A^{(i)}) = \lambda_d(A)(1 - \lambda_d(A)).$$

Now Hoeffding's large deviation bound gives good bounds for the failing probabilities $\mathbb{P}\{|\frac{1}{n}\sum_{i=1}^n \xi_A^{(i)}| > \delta_A\}$ for $\delta_A > 0$ if $\lambda_d(A) \approx 1/2$. But if $\lambda_d(A)$ is much smaller or larger than $1/2$, then Hoeffding's bound cannot exploit the fact that the variance of the random variable $\xi_A^{(i)}$ is small. A large deviation bound which can exploit this fact is *Bernstein's inequality* which reads as follows (see, e.g., [90]):

Let $X_1, \ldots, X_n$ be independent random variables with zero means and bounded ranges $|X_i| \leq M$ for all i. Then for all $t > 0$

$$\mathbb{P}\left\{\left|\sum_{k=1}^n X_i\right| \geq t\right\} \leq 2\exp\left(-\frac{t^2/2}{\sum_{k=1}^n \text{Var}(X_i) + Mt/3}\right).$$

Aistleitner uses Bernstein's inequality and the *dyadic chaining* technique, which can be seen as a "multi-cover" approach:

For all $k = 1, 2, \ldots, K$ consider a 2^{-k}-cover $\Gamma_{2^{-k}}$, and put $x^{(0)} := 0$. From the definition of a δ-cover it follows that for any $x^{(K)} \in \Gamma_{2^{-K}}$ one recursively finds points $x^{(k)} \in \Gamma_{2^{-k}}$, $k = K-1, \ldots, 1$, such that $x_j^{(K)} \geq x_j^{(K-1)} \geq \cdots \geq x_j^{(1)}$ for $j = 1, \ldots, d$, and

$$A_k = A_k\left(x^{(K)}\right) := \left[0, x^{(k)}\right) \setminus \left[0, x^{(k-1)}\right)$$

has volume at most $2^{-(k-1)}$. We have $\left[0, x^{(K)}\right) = \cup_{k=1}^{K} A_k$ and, if P denotes the set $\{p^{(1)}, \ldots, p^{(n)}\}$,

$$\left|\Delta\left(x^{(K)}, P\right)\right| \leq \sum_{k=1}^{K} \left|\lambda_d(A_k) - \frac{1}{n}|P \cap A_k|\right| = \sum_{k=1}^{K} \left|\frac{1}{n}\sum_{i=1}^{n} \xi_{A_k}^{(i)}\right|.$$

If for $k = 1, \ldots, K$ we define $\mathscr{A}_k := \left\{A_k\left(x^{(K)}\right) | x^{(K)} \in \Gamma_{2^{-K}}\right\}$, then $|\mathscr{A}_k| \leq |\Gamma_{2^{-k}}|$. Using a 2^{-k}-bracketing cover as constructed in [35], we obtain via (13) a 2^{-k}-cover $\Gamma_{2^{-k}}$ satisfying $|\Gamma_{2^{-k}}| \leq (2e)^d(2^k+1)^d$, see (24) and (14). Choosing a suitable sequence c_k, $k = 1, \ldots, K$, one essentially obtains with the help of a union bound, Bernstein's inequality, and the estimate (24)

$$\mathbb{P}\left(\bigcup_{A_k \in \mathscr{A}_k}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{A_k}^{(i)}\right| > c_k 2^{-K}\right\}\right) \leq \sum_{A_k \in \mathscr{A}_k}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{A_k}^{(i)}\right| > c_k 2^{-K}\right\} \leq 2^{-k}.$$

Recall that $|\mathscr{A}_k| \leq |\Gamma_{2^{-k}}| \leq O_d(2^{kd})$ and $\mathrm{Var}(\xi_{A_k}^{(i)}) \leq 2^{-(k-1)}$. In particular, $|\mathscr{A}_K|$ is of the size of the finest δ-cover $\Gamma_{2^{-K}}$, but, since the variance of all $\xi_{A_K}^{(i)}$ is small (namely at most $2^{-(K-1)}$), Bernstein's inequality ensures that we can choose a small c_K. If, on the other hand, $k = 1$, then it may happen that $\lambda_d(A_1) \approx 1/2$, so Bernstein's inequality gives us no advantage over Hoeffding's bound. But the size of $\mathscr{A}_1$ is relatively small, namely at most $O_d(2^d)$. In general, the larger k is, the more we can exploit the small variance of all $\xi_{A_k}^{(i)}$, but the larger is the size of $\mathscr{A}_k$. Aistleitner proved that this "trade off" ensures that one can choose $(c_k)_{k=1}^{K}$ such that $\sum_{k=1}^{K} c_k \leq 8.65$ holds. Thus the approximation property (see Lemma 2) leads to the estimate

$$\mathbb{P}\left\{d_\infty^*(P) > \left(1 + \sum_{k=1}^{K} c_k\right)2^{-K}\right\} \leq \mathbb{P}\left\{d_{\Gamma_{2^{-K}}}^*(P) > \sum_{k=1}^{K} c_k 2^{-K}\right\}$$

$$= \mathbb{P}\left(\bigcup_{x^{(K)} \in \Gamma_{2^{-K}}}\left\{\left|\Delta\left(x^{(K)}, P\right)\right| > \sum_{k=1}^{K} c_k 2^{-K}\right\}\right)$$

$$\leq \mathbb{P}\left(\bigcup_{x^{(K)} \in \Gamma_{2^{-K}}}\bigcup_{k=1}^{K}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{A_k}^{(i)}\right| > c_k 2^{-K}\right\}\right)$$

$$\leq \sum_{k=1}^{K}\mathbb{P}\left(\bigcup_{A_k \in \mathscr{A}_k}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{A_k}^{(i)}\right| > c_k 2^{-K}\right\}\right) < 1,$$

showing that there exists an n-point set P in $[0,1]^d$ that satisfies the estimate (8) with $c = 9.65$. (For the technical details we refer, of course, to [1].)

3.2.2 Infinite dimensional infinite sequences

So far we have discussed the existence of point sets that satisfy reasonably good discrepancy bounds. In practice it is desirable to have integration points 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 being able to reuse earlier calculations.

In [14] the probabilistic bounds stated in the previous subsection were extended by Dick to infinite sequences of infinite dimensional points. For an infinite sequence P of points in $[0,1)^{\mathbb{N}}$, let us denote by P_d the sequence of the projections of the points of P onto their first d components, and by $P_{n,d}$ the first n points of P_d. Then in [14] 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 P in $[0,1)^{\mathbb{N}}$ satisfying

$$d_\infty^*(P_{n_m,d}) \leq C\sqrt{d/n_m}\sqrt{\ln(m+1)} \quad \text{for all } m,d \in \mathbb{N}.$$

(We add here that with the help of Aistleitner's approach in [1] one can derive an upper bound for C.)

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 P satisfying

$$d_\infty^*(P_{n_m,d}) \leq C'\sqrt{\left(m+d+d\ln\left(1+\frac{d\sqrt{n_m}}{m+d}\right)\right)/n_m} \quad \text{for all } m,d \in \mathbb{N}. \quad (29)$$

The results from [14] show that there exist point sets that can be extended in the dimension and in the number of points while bounds similar to (10) or (27) remain valid.

A disadvantage of (29) is nevertheless that in the case where, e.g., $n_m = m$ for all m it is not better than the trivial bound $d_\infty^*(P_{m,d}) \leq 1$.

By using the bound (24), another result for infinite sequences P in $[0,1)^{\mathbb{N}}$ was presented in [19]: 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 P satisfying

$$d_\infty^*(P_{n_m,d}) \leq C''\sqrt{d\ln\left(1+\frac{n_m}{d}\right)/n_m} \quad \text{for all } m,d \in \mathbb{N}. \quad (30)$$

This bound is an improvement of (29), which in particular is still useful in the case $n_m = m$ for all m. Moreover, it establishes the existence of infinite sequences P in $[0,1)^{\mathbb{N}}$ having the following property: To guarantee $d_\infty^*(P_{n,d}) \leq \varepsilon$ for a given ε, we only have to take $n \geq c_\varepsilon d$, where c_ε is a constant depending only on ε, see [19, Cor. 2.4]. Note that this result cannot be deduced directly from the results in [14]. As mentioned above, it is known from [45, 49] that we have to take at least $n \geq c_\varepsilon' d$ if ε is sufficiently small. (Here c_ε' depends again only on ε.) In this sense [19, 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 [45]) 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 P* is introduced in [19], given by

$$N_P^*(\varepsilon,d) := \min\{n : \forall m \geq n : d_\infty^*(P_{m,d}) \leq \varepsilon\}.$$

Then Corollary 2.4 of [19] states that there exist sequences P such that

$$N_P^*(\varepsilon,d) \leq O(d\varepsilon^{-2}\ln(1+\varepsilon^{-1})) \quad \text{for all } d \in \mathbb{N}, \varepsilon \in (0,1]. \tag{31}$$

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 P, then inequality (31) holds almost surely for a random sequence P, see again [19, Cor. 2.4]. In [19, Thm. 2.3] bounds of the form (30) and (31) with explicitly given constants and estimates for the measure of the sets of sequences satisfying such bounds are provided.

3.2.3 Hybrid-Monte Carlo Sequences

A *hybrid-Monte Carlo sequence*, which is sometimes also called a *mixed sequence*, results from extending a low-discrepancy sequence in the dimension by choosing the additional coordinates randomly. In several 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, see, e.g., [83, 69, 71].

For a mixed d-dimensional sequences m, whose elements are, technically speaking, vectors obtained by concatenating the 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 Ökten, Tuffin, and Burago showed in [71] that

$$\mathbb{P}(d_\infty^*(m_n) - d_\infty^*(q_n) < \varepsilon) \geq 1 - 2\exp\left(-\frac{\varepsilon^2 n}{2}\right) \quad \text{for } n \text{ sufficiently large.} \tag{32}$$

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 [37] a lower bound for n is derived, which significantly depends on d and ε. Furthermore, with the help of the probabilistic proof scheme the probabilistic bound

$$\mathbb{P}(d_\infty^*(m_n) - d_\infty^*(q_n) < \varepsilon) > 1 - 2N(d,\varepsilon/2)\exp\left(-\frac{\varepsilon^2 n}{2}\right) \tag{33}$$

was established, which holds without any restriction on n. In this sense it improves the bound (32) and is more helpful in practice, especially for small samples sizes n.

As we know from (25) and (14), for small ε the quantity $N(d,\varepsilon/2)$ grows exponentially in d. As pointed out in [37, Remark 3.4] a factor depending exponentially on d has to appear in front of $\exp(-\varepsilon^2 n/2)$ in the bound (33) if we want it to hold for all $n \in \mathbb{N}$. Recall that we can use the bound (24) on the bracketing number to obtain an upper bound for $N(d,\varepsilon/2)$ with explicit constants.

Recently, there has been increasing interest in (deterministic) discrepancy bounds for (deterministic) mixed sequences, see, e.g., [63, 64].

3.3 Small Discrepancy Samples via Derandomization

Here we want to consider question (ii) from Section 1.2: How can we construct point sets that satisfy the probabilistic bounds stated in Section 3.2? How can we derandomize the probabilistic experiments to get deterministic point sets with low discrepancy? The probabilistic experiment of distributing n points uniformly at random in $[0,1]^d$ was derandomized in [21]. We illustrate the derandomization idea for a different probabilistic experiment used in [23], which leads to a simpler and faster algorithm.

3.3.1 Random Experiment

Let $k \in \mathbb{N}$ be given and let δ be the largest value that satisfies $k = \kappa(\delta,d)$, where $\kappa(\delta,d)$ is as in (18). Let $\Gamma = \Gamma_\delta$ be the non-equidistant grid from (16). Put $\gamma_{k+1} := 0$ and let $\mathscr{B} = \mathscr{B}_\delta$ the set of all (half-open) grid cells, i.e., all boxes $[y,y^+)$ with $y_i = \gamma_j$ for some $j = j(i) \in \{1,\ldots,k+1\}$ and $y_i^+ = \gamma_{j-1}$ for all $i \in d$. Then obviously $|\Gamma| = |\mathscr{B}|$.

Let $n \in \mathbb{N}$ be given. For $B \in \mathscr{B}$ let $x_B := n \cdot \lambda_d(B)$, i.e., x_B is the expected number of points inside B if we distribute n points independently at random in $[0,1]^d$.

Our aim is now to *round randomly* for each $B \in \mathscr{B}$ the real number x_B to an integer y_B such that the following two constraints are satisfied:

- *Weak constraint*: Each set Y with y_B points in B for all $B \in \mathscr{B}$ should have small discrepancy with high probability.
- *Hard constraint*: The equation $|Y| = \sum_{B \in \mathscr{B}} y_B = \sum_{B \in \mathscr{B}} x_B = n$ should hold.

We saw in Section 3.2 that in the previous random experiments the weak constraint can be satisfied for independent random points with the help of large deviation inequalities. But if our rounding procedure has to satisfy the hard constraint our random variables y_B, $B \in \mathscr{B}$, are clearly not independent any more.

Nevertheless, such a randomized rounding that satisfies the weak constraint with high probability and respects the hard constraint can be done. There are two approaches known, due to Srinivasan [84] and to Doerr [18]. We present here the randomized rounding procedure of Srinivasan:

Randomized Rounding Procedure:

- Initialize $y_B = x_B$ for all $B \in \mathscr{B}$.
- Repeat the following step until all y_B are integral:

  Pair Rounding Step: Choose y_B, $y_{B'}$ not integral.
  Choose $\sigma \in [0,1]$ minimal such that $y_B + \sigma$ or $y_{B'} - \sigma$ is integral.
  Choose $\tau \in [0,1]$ minimal such that $y_B - \tau$ or $y_{B'} + \tau$ is integral.
  Set
  $$(y_B, y_{B'}) := \begin{cases} (y_B + \sigma, y_{B'} - \sigma) & \text{with probability } \frac{\tau}{\sigma + \tau}, \\ (y_B - \tau, y_{B'} + \tau) & \text{with probability } \frac{\sigma}{\sigma + \tau}. \end{cases}$$

- Output: $(y_B)_{B \in \mathscr{B}}$.

The pair rounding step leaves $\sum_{B \in \mathscr{B}} y_B$ invariant. Hence we have always

$$\sum_{B \in \mathscr{B}} y_B = \sum_{B \in \mathscr{B}} x_B = n.$$

This shows particularly that if there is a variable y_B left which is not integral, there has to be another one $y_{B'}$, $B \neq B'$, which is not integral. Thus the algorithm terminates and the output set y_B, $B \in \mathscr{B}$, satisfies the hard constraint. Furthermore, the pair rounding step leaves $\mathbb{E}(y_B)$ invariant, hence $\mathbb{E}(y_B) = x_B$. Now let Y be a set with y_B points in B for all $B \in \mathscr{B}$. Then

$$\mathbb{E}(n\Delta(g,Y)) = \mathbb{E}\left(\sum_{B \in \mathscr{B}; B \subseteq [0,g)} (x_B - y_B) \right) = 0 \quad \text{for all } g \in \Gamma.$$

Furthermore, a concentration of measure result holds. The y_B, $B \in \mathscr{B}$, are not independent, but it can be shown that they satisfy certain negative correlation properties, cf. [84]. As shown by Panconesi and Srinivasan, Chernoff-Hoeffding-type bounds hold also in this situation [74]. This result and the earlier observations yield the following theorem, see [23].

Theorem 5. *The randomized rounding procedure generates in time $O(|\mathscr{B}|)$ randomized roundings y_B of x_B for all $B \in \mathscr{B}$ such that $\sum_{B \in \mathscr{B}} y_B = \sum_{B \in \mathscr{B}} x_B = n$ and*

$$\mathbb{P}\{|\Delta(g,Y)| > \lambda\} < 2\exp\left(-\frac{\lambda^2 n}{3} \right) \quad \text{for all } g \in \Gamma.$$

If we now choose $\lambda = \sqrt{3n^{-1}\ln(2|\Gamma|)}$ and $\delta \approx \sqrt{d/n}\sqrt{\ln\ln(d)}$, then the next theorem can be proved by following the three steps of the proof sheme in Section 3.2.1, see [23].

Theorem 6. *There exists a constant $C > 0$ such that*

$$\mathbb{P}\left\{ d_\infty^*(Y) \leq C\sqrt{d/n}\sqrt{\ln(\sigma n)} \right\} > 0, \tag{34}$$

where $\sigma = \sigma(d) < 1.03$ tends to zero if $d \to \infty$.

(Essentially we have $C \approx \sqrt{6}$.)

3.3.2 Derandomized Construction

Now we want to derandomize the random experiment, i.e., we want to construct an n-point set Y deterministically that satisfies the bound (34) in Theorem 6. More precisely, we want to compute a rounding $(y_B)_{B \in \mathscr{B}}$ of $(x_B)_{B \in \mathscr{B}}$ that satisfies

$$\sum_{B \in \mathscr{B}} y_B = \sum_{B \in \mathscr{B}} x_B = n \tag{35}$$

and

$$\left| \sum_{B \subseteq [0,g)} (x_B - y_B) \right| \leq \delta_g \cdot n \cdot \lambda_d([0,g)) \quad \text{for all } g \in \Gamma, \tag{36}$$

where the δ_gs are error tolerances fixed in the algorithm. If then Y is a set with y_B points in B for all $B \in \mathscr{B}$, we obtain $|Y| = n$ and

$$\left| \lambda_d([0,g)) - \frac{1}{n} |Y \cap [0,g)| \right| \leq \delta_g \cdot \lambda_d([0,g)) \quad \text{for all } g \in \Gamma.$$

To compute such a rounding we follow Raghavan [75] and define *pessimistic estimators* P_g^+, P_g^-, $g \in \Gamma$. For $B \in \mathscr{B}$ let $p_B = \{x_B\}$, where $\{x_B\}$ denotes the fractional part of x_B, and for $g \in \Gamma$ let $\mu_g := \sum_{B \subseteq [0,g)} \{x_B\}$. The pessimistic estimators are defined as

$$P_g^+ = (1 + \delta_g)^{-(1+\delta_g)\mu_g} \prod_{B \subseteq [0,g)} (1 + \delta_g p_B)$$

$$P_g^- = (1 + \delta_g)^{(1-\delta_g)\mu_g} \prod_{B \subseteq [0,g)} \left(1 + \left(\frac{1}{1+\delta_g} - 1 \right) p_B \right).$$

With the help of the pessimistic estimators we can see whether (36) is satisfied or not. This is easily seen by making the following observation: For $B \in \mathscr{B}$ let $q_B \in \{0,1\}$, and for $q \in \Gamma$ let Q_g^+, Q_g^- be the values of P_g^+ and P_g^-, respectively, calculated on values q_B instead of p_B (with μ_g unchanged). Then it is a simple observation that $Q_g^+ \geq 1$ if and only if $\sum_{B \subseteq [0,g)} q_B \geq (1 + \delta_g)\mu_g$, and $Q_g^- \geq 1$ if and only if $\sum_{B \subseteq [0,g)} q_B \leq (1 - \delta_g)\mu_g$.

By *updating* the pessimistic estimators for some adjustment $p_B \leftarrow x$, we shall mean the operation of replacing the factor $(1 + \delta_g p_B)$ in P_g^+ by $(1 + \delta_g x)$, and analogously for P_g^-, for each $g \in \Gamma$ such that $B \subseteq [0,g)$. (Again, μ_g stays unchanged.)

The derandomized rounding algorithm proceeds as follows.

Derandomized Rounding Procedure:

1. Initialize $p_B := \{x_B\}$ for all $B \in \mathscr{B}$.
2. Set the error tolerances δ_g such that for each $g \in \Gamma$ we have $P_g^+, P_g^- < 1/(2|\Gamma|)$. Let $U := \sum_{g \in \Gamma} (P_g^+ + P_g^-)$.
3. Let $\mathscr{J} = \{B \in \mathscr{B} \mid p_B \notin \{0,1\}\}$. While $|\mathscr{J}| \geq 2$:

   a. Pick $B, B' \in \mathscr{J}$.

    b. Let $(p_B^{(i)}, p_{B'}^{(i)})$, $i = 1, 2$, be the two possible outcomes of the pair-rounding step of the randomized rounding procedure with respect to the pair of variables $(p_B, p_{B'})$. Let U_i, $i = 1, 2$, be the sum of the pessimistic estimators U updated according to the corresponding outcome.

    c. Pick $i \in \{1, 2\}$ to minimize U_i. Let $p_B \leftarrow p_B^{(i)}$, $p_{B'} \leftarrow p_{B'}^{(i)}$ and update $\mathscr{J}$, the pessimistic estimators, and U.

4. Output: $y_B = \lfloor x_B \rfloor + p_B$, $B \in \mathscr{B}$.

Note that in step 2 we have $U < 1$. Furthermore, it was shown in [24, Sect. 3.1] that the minimum U_i of $\{U_1, U_2\}$ appearing in step 3.c satisfies $U_i \leq U$. After step 3 we have $\mathscr{J} = \emptyset$ and $p_B \in \{0, 1\}$ for every $B \in \mathscr{B}$. By our previous observation, $\sum_{B \subseteq [0,g)} p_B \geq (1 + \delta_g)\mu_g$ if and only if $P_g^+ \geq 1$, and analogously for the lower bound. Since $U < 1$ is maintained throughout the algorithm and since the pessimistic estimators are non-negative, this cannot occur. The process thus produces a rounding satisfying equation (36). Note that as in the randomized rounding, the value of $\sum_{B \in \mathscr{B}} p_B$ is kept constant throughout the process, thus (35) is satisfied.

Although the order in which variables are picked in step 3.a is not important for the theoretical bound, numerical tests indicate that it is preferable to use an order in which the tree formed by the pairings is a balanced binary tree (so that each value p_B is adjusted only $O(\log |\Gamma|)$ times), see [24] for details.

Using the bounds on the δ_gs derived by Raghavan [75] and choosing δ of order $\delta \approx \sqrt{d/n} \sqrt{\ln\ln(d)}$, the derandomized rounding algorithm leads to the following theorem, see [23].

Theorem 7. *There exists a deterministic algorithm which, on input n and d, computes in time $O(d \ln(dn)(\sigma n)^d)$ an n-point set $Y \subset [0,1]^d$ with discrepancy*

$$d_\infty^*(Y) \leq C\sqrt{d/n}\,\sqrt{\ln(\sigma n)};$$

here $C < 2.44$, and $\sigma = \sigma(d) < 1.03$ tends to zero if $d \to \infty$.

The output set Y has y_B points in each grid cell $B \in \mathscr{B}$. Although the exact placement of these points inside the boxes B does not affect the theoretical bound on $d_\infty^*(Y)$ from Theorem 7, numerical experiments indicate that it is a good idea to place these points independently, uniformly at random in B.

3.3.3 A Component-by-Component Derandomization

Another approach is presented in [19]. 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 random experiment is as follows: Given an n-point set $P_{d'} = \{p^{(1)}, \dots, p^{(n)}\}$ in dimension d', we choose new components $x^{(1)}, \dots, x^{(n)}$ randomly from some one-dimensional grid and receive the n-point set $P_{d'+1} = \{(p^{(1)}, x^{(1)}), \dots, (p^{(n)}, x^{(n)})\}$.

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 [75]. If we start the CBC-construction in dimension one, the deterministic output set P_d of size n in dimension d satisfies the bound

$$d_\infty^*(P_d) \le O(d^{3/2} n^{-1/2} \ln(1 + n/d)^{1/2}). \tag{37}$$

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 (37) is weaker than the bound in Theorem 7, but the bound on the running time of the CBC algorithm is a reasonable improvement upon the running time guarantee of the derandomized algorithm discussed 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 [22] 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 (37). 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 [22] we want to call the latter discrepancies "grid gaps"). For output sets of size n the corresponding full grid has size larger than $n^{d/2}/d!$. 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 P_d, we randomize its points locally to receive a new output set P_d^*. For the randomized set P_d^* the theoretical discrepancy bound (37) still holds, and in all the numerical tests in dimension $d = 10$ its discrepancy was always much smaller than the corresponding grid gap (which, as already said, is a lower bound for $d_\infty^*(P_d)$). (To be more precise, an estimator for $d_\infty^*(P_d^*)$, which majorizes $d_\infty^*(P_d^*)$ with certainty at least 95%, is always much smaller than the corresponding grid gap. We use this estimator, since calculating the actual discrepancy of P_d^* is a much harder problem than calculating the discrepancy of P_d.)

The star discrepancy of the output sets of both derandomized algorithms we presented here was compared in [23] to the star discrepancy of other low discrepancy point sets. These experiments took place in dimensions from 5 to 21 and indicate that the first derandomized algorithm leads to superior results if the dimension is relatively high and the number of points is rather small. (We use the phrase "indicate", since for dimension 10 or more, we are not able to calculate the exact discrepancy, but can only use upper and lower bounds on it.) For details see [23].

4 Conclusion and Open Problems

In the previous sections we discussed question (i), (ii), and (iii) and described in particular how approaches based on bracketing entropy, randomization, and derandomization lead to improvements on previously achieved results.

The discussion shows that good bounds for the star discrepancy with explicitly known constants are available. Similar bounds hold also for the star discrepancy of point sets that are extensible in the number of points and in the dimension, and the statement that the inverse of the star discrepancy depends linearly on the dimension d [45] can be extended to this situation: The inverse of the star discrepancy of infinite sequences in $[0,1)^{\mathbb{N}}$ depends almost surely linearly on the dimension d.

Can we find even better bounds than (27) or (8)? A lower bound for the star discrepancy that follows directly from (9) is of the form $d_{\infty}^{*}(n,d) \geq \min\{\varepsilon_0, c_0 d n^{-1}\}$, c_0 and ε_0 suitable constants [49, Thm. 1], and leaves some room for improvements of (27) or (8). Also the bound (28) shows that some trade-off between the dependence on the number of points and on the dimension is possible. But instead of agonizing over this intriguing question, let us state the *conjecture of Woźniakowski* (see [44], or [66, Open Problem 7]): *If there exist constants $C, \alpha > 0$ and a polynomial p such that*

$$d_{\infty}^{*}(n,d) \leq C\, p(d) n^{-\alpha} \quad \text{for all } d, n \in \mathbb{N}, \tag{38}$$

then necessarily $\alpha \leq 1/2$.

The construction of point sets satisfying bounds like (8) or (27) can be done with the help of derandomized algorithms [21, 19, 22, 23]. Unfortunately, these algorithms exhibit running times that are 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 only be solved in exponential time in d so far. And indeed the problem of calculating the star discrepancy is known to be *NP*-hard.

Nevertheless, the discussed derandomized algorithms can be used in low and modestly high dimension d.

In light of the discussion above, it would be of interest to make further progress in designing algorithms that construct low-discrepancy point sets of small size and algorihms that approximate the star discrepancy of arbitrary n-point sets (which would allow "semi-constructions" as described above). Furthermore, it would be interesting to learn more about the dependence of the star discrepancy of classical constructions on the dimension d and the complexity of approximating the star discrepancy of given point sets.

References

1. Aistleitner, C.: Covering numbers, dyadic chaining and discrepancy. J. Complexity 2011. (doi: 10.1016/j.jco.2011.03.001)
2. Atanassov, E.: On the discrepancy of the Halton sequences. Math. Balkanica (N.S.) **18**, 15–32 (2004)
3. Beck, J.: Some upper bounds in the theory of irregularities of distribution. Acta Arith. **43**, 115–130 (1984)
4. Beck, J.: A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. Composito Math. **72**, 269–339 (1989)
5. Beck, J., Chen, W.W.L.: Irregularities of Distribution. Cambridge University Press, Cambridge (1987)
6. Bilyk, D., Lacey, M.T.: On the small ball inequality in three dimensions. Duke Math. J. **143**, 81–115 (2008)
7. Bilyk, D., Lacey, M.T., Vagharshakyan, A.: On the small ball inequality in all dimensions. J. Funct. Anal. **254**, 2470–2502 (2008)
8. Bundschuh, P., Zhu, Y.C.: A method for exact calculation of the discrepancy of low-dimensional point sets I. Abh. Math. Sem. Univ. Hamburg **63**, 115–133 (1993)
9. Chari, S., Rohatgi, P., Srinivasan, A.: Improved algorithms via approximation of probability distributions. J. Comput. System Sci. **61**, 81–107 (2000)
10. de Clerck, L.: A method for exact calculation of the star-discrepancy of plane sets applied to the sequence of Hammersley. Monatsh. Math. **101**, 261–278 (1986)
11. van der Corput, J.G.: Verteilungsfunktionen I. Nederl. Akad. Wetensch. Proc. Ser. B **38**, 813–821 (1935)
12. van der Corput, J.G.: Verteilungsfunktionen II. Nederl. Akad. Wetensch. Proc. Ser. B **38**, 1058–1066 (1935)
13. Davenport, H.: Note on irregularities of distribution. Matematika **3**, 131–135 (1956)
14. Dick, J.: A note on the existence of sequences with small star discrepancy. J. Complexity **23**, 649–652 (2007)
15. Dick, J.: Koksma-Hlawka type inequalities of fractional order. Ann. Mat. Pura Appl. **187**, 385–403 (2008)
16. Dick, J., Pillichshammer, F.: Digital Nets and Sequences. Cambridge University Press, Cambridge (2010)
17. Dobkin, D.P., Eppstein, D., Mitchell, D.P.: Computing the discrepancy with applications to supersampling patterns. ACM Trans. Graph. **15**, 354–376 (1996)
18. Doerr, B.: Generating randomized roundings with cardinality constraints and derandomizations. In: B. Durand, W. Thomas (eds.) STACS 2006, 23rd Annual Symposium on Theoretical Aspects of Computer Science, *Lecture Notes in Computer Science*, vol. 3884, pp. 571–583. Springer (2006)
19. Doerr, B., Gnewuch, M., Kritzer, P., Pillichshammer, F.: Component-by-component construction of low-discrepancy point sets of small size. Monte Carlo Methods Appl. **14**, 129–149 (2008)
20. Doerr, B., Gnewuch, M., Srivastav, A.: Bounds and construction for the star discrepancy via δ-covers (2004). Berichtsreihe des Mathematischen Seminars der Christian-Albrechts-Universität zu Kiel, Report 04–13 (Preliminary preprint version of [21])
21. Doerr, B., Gnewuch, M., Srivastav, A.: Bounds and constructions for the star discrepancy via δ-covers. J. Complexity **21**, 691–709 (2005)
22. Doerr, B., Gnewuch, M., Wahlström, M.: Implementation of a component-by-component algorithm to generate low-discrepancy samples. In: P. L'Ecuyer, A.B. Owen (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2008, pp. 323–338. Springer, Berlin Heidelberg (2009)
23. Doerr, B., Gnewuch, M., Wahlström, M.: Algorithmic construction of low-discrepancy point sets via dependent randomized rounding. J. Complexity **26**, 490–507 (2010)
24. Doerr, B., Wahlström, M.: Randomized rounding in the presence of a cardinality constraint. In: I. Finocchi, J. Hershberger (eds.) Proceedings of the 10th Workshop on Algorithm Engineering and Experiments (ALENEX 2009), pp. 162–174. SIAM, New York, USA (2009)

25. Drmota, M., Tichy, R.F.: Sequences, Discrepancies, and Applications, *Lecture Notes in Math.*, vol. 1651. Springer, Berlin Heidelberg New York (1997)

26. Dueck, G., Scheuer, T.: Threshold accepting: A general purpose algorithm appearing superior to simulated annealing. J. Comp. Phys. **90** (1990)

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

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

29. Frank, K., Heinrich, S.: Computing discrepancies of Smolyak quadrature rules. J. Complexity **12**, 287–314 (1996)

30. Frolov, K.K.: An upper estimate for the discrepancy in the L_p-metric, $2 \le p \le \infty$. Soviet. Math. Dokl. **21**, 840–842 (1980)

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

32. Giannopoulos, P., Knauer, C., Wahlström, M., Werner, D.: Hardness of discrepancy computation and ε-net verification in high dimension. Preprint 2011. (Submitted for publication)

33. Gnewuch, M.: Weighted geometric discrepancies and numerical integration on reproducing kernel Hilbert spaces. J. Complexity 2011. (doi: 10.1016/j.jco.2011.02.003)

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

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

36. Gnewuch, M.: Construction of minimal bracketing covers for rectangles. Electron. J. Combin. **15**, Research Paper 95 (2008)

37. Gnewuch, M.: On probabilistic results for the discrepancy of a hybrid-Monte Carlo sequence. J. Complexity **23**, 312–317 (2009)

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

39. Gnewuch, M., Wahlström, M., Winzen, C.: A new randomized algorithm to approximate the star discrepancy based on threshold-accepting. Preprint Math. Archive 2011. (arXiv: 1103.2102v2 [cs.DS])

40. Halton, J.H.: On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals. Numer. Math. **2**, 84–90 (1960)

41. Hardy, G.H., Littlewood, J.E.: Some problems of diophantine approximation: the lattice points of a right-angled triangle. part i. Proc. London Math. Soc.. **20**, 212–249 (1922)

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

43. Heinrich, S.: Efficient algorithms for computing the L_2 discrepancy. Math. Comp. **65**, 1621–1633 (1996)

44. Heinrich, S.: Some open problems concerning the star-discrepancy. J. Complexity **19**, 416–419 (2003)

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

46. Hickernell, F.J.: A generalized discrepancy and quadrature error bound. Math. Comp. **67**, 299–322 (1998)

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

48. Hickernell, F.J., Wang, X.: The error bounds and tractability of quasi-monte carlo algorithms in infinite dimensions. Math. Comp. **71**, 1641–1661 (2001)

49. Hinrichs, A.: Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy. J. Complexity **20**, 477–483 (2004)

50. Hinrichs, A., Pillichshammer, F., Schmid, W.C.: Tractability properties of the weighted star discrepancy. J. Complexity **23**, 134–143 (2008)

51. Hlawka, E.: Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. Ann. Mat. Pura Appl. **54**, 325–333 (1961)
52. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. **58**, 13–30 (1963)
53. Hromković, J.: Algorithms for Hard Problems. Springer, Berlin Heidelberg (2003). 2nd edition
54. Johnson, D.S.: The NP-completeness column: an ongoing guide. J. Algorithms **8**, 438–448 (1987)
55. Kirckpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. Science **20**, 671–680 (1983)
56. Koksma, J.F.: Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1. Mathematica B (Zutphen) **11**, 7–11 (1942/3)
57. Lerch, M.: Question 1547. L'Intermédiaire des Mathématiciens **11**, 145–146 (1904)
58. Matoušek, J.: Geometric Discrepancy. Springer-Verlag, Berlin Heidelberg (1999). Revised second printing 2010.
59. Mhaskar, H.N.: On the tractability of multivariate integration and approximation by neural networks. J. Complexity **20**, 561–590 (2004)
60. Niederreiter, H.: Discrepancy and convex programming. Ann. Mat. Pura Appl. **93**, 89–97 (1972)
61. Niederreiter, H.: Methods for estimating discrepancy. In: S.K. Zaremba (ed.) Applications of number theory to numerical analysis, pp. 203–236. Academic Press, New York (1972)
62. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1992)
63. Niederreiter, H.: On the discrepancy of some hybrid sequences. Acta Arith. **138**, 373–398 (2009)
64. Niederreiter, H.: Further discrepancy bounds and an Erdös-Turán-Koksma inequality for hybrid sequences. Monatsh. Math. **161**, 193–222 (2010)
65. Niederreiter, H., Xing, C.: Low-discrepancy sequences and global function fields with many rational places. Finite Fields and their Applications **2**, 241–273 (1996)
66. Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems, Volume I: Linear Information. European Mathematical Society, Freiburg (2008)
67. Novak, E., Woźniakowski, H.: L_2 discrepancy and multivariate integration. In: W.W.L. Chen, W.T. Gowers, H. Halberstam, W.M. Schmidt, R.C. Vaughan (eds.) Essays in honour of Klaus Roth, pp. 359–388. Cambridge Univ. Press, Cambridge (2009)
68. Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems, Volume II: Standard Information for Functionals. European Mathematical Society, Freiburg (2010)
69. Ökten, G.: A probabilistic result on the discrepancy of a hybrid Monte Carlo sequence and applications. Monte Carlo Methods Appl. **2**, 255–270 (1996)
70. Ökten, G., Shah, M., Goncharov, Y.: Random and deterministic digit permutations of the Halton sequence (2009). Technical report
71. Ökten, G., Tuffin, B., Burago, V.: A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance. J. Complexity **22**, 435–458 (2006)
72. Ostrowski, A.: Bemerkungen zur Theorie der Diophantischen Approximationen i. Abh. Math. Semin. Univ. Hamburg **1**, 77–98 (1922)
73. Ostrowski, A.: Bemerkungen zur Theorie der Diophantischen Approximationen ii. Abh. Math. Semin. Univ. Hamburg **1**, 250–251 (1922)
74. Panconesi, A., Srinivasan, A.: Randomized distributed edge coloring via an extension of the Chernoff–Hoeffding bounds. SIAM J. Comput. **26**, 350–368 (1997)
75. Raghavan, P.: Probabilistic construction of deterministic algorithms: approximating packing integer programs. J. Comput. Syst. Sci. **37**, 130–143 (1988)
76. Roth, K.F.: On irregularities of distribution. Mathematika **1**, 73–79 (1954)
77. Roth, K.F.: On irregularities of distribution III. Acta Arith. **35**, 373–384 (1979)
78. Roth, K.F.: On irregularities of distribution IV. Acta Arith. **37**, 67–75 (1980)

79. Schmidt, W.M.: On irregularities of distribution VII. Acta Arith. **21**, 45–50 (1972)
80. Shah, M.: A genetic algorithm approach to estimate lower bounds of the star discrepancy. Monte Carlo Methods Appl. **16**, 379–398 (2010)
81. Sloan, I.H., Joe, S.: Lattice Methods for Multiple Integration. Oxford University Press, New York (1994)
82. Sloan, I.H., Woźniakowski, H.: When are quasi-monte carlo algorithms efficient for high dimensional integrals? J. Complexity **14**, 1–33 (1998)
83. Spanier, J.: 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, *Lecture Notes in Statistic*, vol. 106, pp. 121–148. Springer, Berlin (1995)
84. Srinivasan, A.: Distributions on level-sets with applications to approximation algorithms. In: Proceedings of FOCS'01, pp. 588–597 (2001)
85. Steinerberger, S.: The asymptotic behavior of the average L^p-discrepancies and a randomized discrepancy. Electron. J. Combin. **17(1)**, Research paper 106 (2010)
86. Talagrand, M.: Sharper bounds for Gaussian and empirical processes. Ann. Probab. **22**, 28–76 (1994)
87. Thiémard, E.: Computing bounds for the star discrepancy. Computing **65**, 169–186 (2000)
88. Thiémard, E.: An algorithm to compute bounds for the star discrepancy. J. Complexity **17**, 850–880 (2001)
89. Thiémard, E.: Optimal volume subintervals with k points and star discrepancy via integer programming. Math. Meth. Oper. Res. **54**, 21–45 (2001)
90. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer Series in Statistics. Springer, New York (1996)
91. Warnock, T.T.: Computational investigations of low-discrepancy point sets. In: S.K. Zaremba (ed.) Applications of number theory to numerical analysis, pp. 319–343. Academic Press, New York (1972)
92. Winker, P., Fang, K.T.: Application of threshold-accepting to the evaluation of the discrepancy of a set of points. SIAM J. Numer. Anal. **34**, 2028–2042 (1997)
93. Woźniakowski, H.: Average case complexity of multivariate integration. Bull. Amer. Math. Soc. (N. S.) **24**, 185–191 (1991)
94. Zaremba, S.K.: Some applications of multidimensional integration by parts. Ann. Polon. Math. **21**, 85–96 (1968)