

# Bounds and Constructions for the Star-Discrepancy via $\delta$ -Covers

Benjamin Doerr, Michael Gnewuch<sup>1</sup> and Anand Srivastav  
Mathematisches Seminar, Bereich II  
Christian-Albrechts-Universität zu Kiel  
Christian-Albrechts-Platz 4, 24098 Kiel, Germany  
Tel.: +49-431/880-{2776, 7451, 7252}, Fax: +49-431/880-1725  
E-mail: {bed, mig, asr}@numerik.uni-kiel.de

February 15, 2005

<sup>1</sup>Supported by the Deutsche Forschungsgemeinschaft under Grant SR7/10-1

## Abstract

For numerical integration in higher dimensions, bounds for the star-discrepancy with polynomial dependence on the dimension  $d$  are desirable. Furthermore, it is still a great challenge to give construction methods for low-discrepancy point sets.

In this paper we give upper bounds for the star-discrepancy and its inverse for subsets of the  $d$ -dimensional unit cube. They improve known results. In particular, we determine the usually only implicitly given constants. The bounds are based on the construction of nearly optimal  $\delta$ -covers of anchored boxes in the  $d$ -dimensional unit cube.

We give an explicit construction of low-discrepancy points with a derandomized algorithm. The running time of the algorithm, which is exponentially in  $d$ , is discussed in detail and comparisons with other methods are given.

**Keywords:** covering number, derandomization, low-discrepancy point sets, probabilistic methods, star-discrepancy

**Mathematics Subject Classification:** 11K38

# 1 Introduction

The  $L^\infty$ -*star-discrepancy* of points  $t_1, \dots, t_n$  in the  $d$ -dimensional unit cube  $[0, 1]^d$  is given by

$$d_\infty^*(t_1, \dots, t_n) = \sup_{x \in [0, 1]^d} \left| \text{vol}([0, x[) - \frac{1}{n} \sum_{k=1}^n 1_{[0, x[}(t_k) \right|,$$

where  $1_{[0, x[}$  is the characteristic function of the  $d$ -dimensional anchored half-open box  $[0, x[ = [0, x_1[ \times \dots \times [0, x_d[$ . The smallest possible discrepancy of any  $n$ -point configuration in  $[0, 1]^d$  is

$$d_\infty^*(n, d) = \inf_{t_1, \dots, t_n \in [0, 1]^d} d_\infty^*(t_1, \dots, t_n).$$

The *inverse of the star-discrepancy* is given by

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

The star-discrepancy is intimately related to the worst case error of multivariate integration of certain function classes (see, e.g., [2, 9, 11, 14]). A well-known result in this direction is the Koksma-Hlawka inequality

$$\left| \int_{[0, 1]^d} f(x) dx - \frac{1}{n} \sum_{i=1}^n f(t_i) \right| \leq d_\infty^*(t_1, \dots, t_n) V(f),$$

where  $V(f)$  denotes the so-called variation in the sense of Hardy and Krause, which depends only on  $f$ . This inequality illustrates that points with small discrepancy induce cubature formulas with small worst case errors. It also indicates that the number of discrepancy points corresponds to the number of function evaluations used by the class of cubature formulas under consideration, and the last quantity is typically directly related to the costs of those algorithms. So for multivariate integration it is of interest to find  $n$ -point configurations with small discrepancy and  $n$  not too large.

For fixed dimension  $d$  the best known upper bounds for  $d_\infty^*(n, d)$  are of the form

$$d_\infty^*(n, d) \leq C_d \ln(n)^{d-1} n^{-1}, \quad n \geq 2. \quad (1)$$

If we are seeking for good bounds for large  $d$  and moderate values of  $n$ , then those bounds do not give us any helpful information, since  $\ln(n)^{d-1} n^{-1}$  is

an increasing function for  $n \leq e^{d-1}$ . Apart from that, point configurations satisfying (1) will in general lead to constants  $C_d$  that depend critically on  $d$ . If we take, e.g., the famous Halton-Hammersley points, then  $C_d$  grows superexponentially in  $d$  (see, e.g., [14]).

A bound that seems to be more suitable for multivariate integration was established by Heinrich, Novak, Wasilkowski and Woźniakowski [8], who proved

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

where  $c$  does not depend on  $d$ ,  $n$  or  $\varepsilon$ . Here the dependence on  $d$  is optimal. This was also established in [8] by a lower bound for  $n_{\infty}^*(d, \varepsilon)$ , which has recently been improved by Hinrichs [10] to  $n_{\infty}^*(d, \varepsilon) \geq c_0d\varepsilon^{-1}$  for  $0 < \varepsilon < \varepsilon_0$ , where  $c_0, \varepsilon_0 > 0$  are constants. The proof of (2) is not constructive but probabilistic and the constant  $c$  is unfortunately not known.

In the same paper the authors proved a slightly weaker bound with an explicitly known small constant  $k$ :

$$d_{\infty}^*(n, d) \leq kd^{1/2}n^{-1/2}(\ln(d) + \ln(n))^{1/2}. \quad (3)$$

The proof is again probabilistic and uses Hoeffding's inequality. For the sake of explicit constants the proof technique has been adapted in subsequent papers on high-dimensional integration of certain function classes  $\mathcal{F}$  [9, 12]. As pointed out by Mhaskar [12], the key idea is to find finite *one-sided*  $(\mu, \delta)$ -covers for the class of functions  $\mathcal{F}$  under consideration. Since here we are only interested in the situation where  $\mathcal{F}$  is the set of characteristic functions of anchored half-open boxes in the unit cube, we shall use the shorter term  *$\delta$ -cover*. Our definition of a  $\delta$ -cover is precisely the following:

A finite site  $\Gamma \subset [0, 1]^d$  is a  $\delta$ -cover of  $[0, 1]^d$  if for every  $y \in [0, 1]^d$  there exist  $x, z \in \Gamma \cup \{0\}$  with  $\text{vol}([0, z]) - \text{vol}([0, x]) \leq \delta$  and  $x_i \leq y_i \leq z_i$  for all  $i \in \{1, \dots, d\}$ .

Let  $N(d, \delta)$  denote the number of elements in a minimal  $\delta$ -cover of  $[0, 1]^d$ . In this paper we were able to improve (3) to

$$d_{\infty}^*(n, d) \leq k'd^{1/2}n^{-1/2}\ln(n)^{1/2}, \quad (4)$$

where  $k'$  is smaller than  $k$ , by deducing reasonably good upper bounds for  $N(d, \delta)$ . Using a derandomized version of Hoeffding's inequality [19], we give a deterministic construction of a point set satisfying (4). For the construction of  $n = O(\varepsilon^{-2}d(\ln \ln(d) + \ln(1/\varepsilon)))$  points with star-discrepancy at most  $\varepsilon > 0$ ,

this algorithm has running time exponential in  $d$ . We leave open the problem whether there is a polynomial-time algorithm in  $d$  achieving our bound.

We prove a lower bound for the minimum cardinality  $N(d, \delta)$  of a  $\delta$ -cover which gives together with our upper bound

$$d^{1/2}e^{-d}\delta^{-d} + O(|\ln(\delta)|^{d-1}) \leq N(d, \delta) \leq d^{-1/2}e^d\delta^{-d} + O(\delta^{-d+1}), \quad (5)$$

where the implicit constants of the  $O$ -notation may depend on  $d$ , but not on  $\delta$ . This lower bound is significant as it shows that the bound on the star-discrepancy in (4) cannot be improved with the Hoeffding approach and our  $\delta$ -cover technique beyond  $d_\infty^*(n, d) = O(d^{1/2}n^{-1/2} \ln(n/d)^{1/2})$ .

Away from the application to geometric discrepancy, the problem of finding small  $\delta$ -covers seems to be an interesting problem on its own as it is related to the NP-hard set cover problem in combinatorics (see, e.g., [18] for a discussion of this problem). Furthermore, it is related to the covering and the  $L^1(\lambda)$ -packing number of anchored boxes in the  $d$ -dimensional unit cube (see Remark 2.10).

The paper is organized as follows: In Section 2 we construct small  $\delta$ -covers and establish in this way upper bounds for the smallest possible cardinality of  $\delta$ -covers. In Section 3 the impact on the star-discrepancy is given and in Section 4 we provide the derandomized constructive algorithm.

## 2 Construction of Small $\delta$ -Covers

Let  $d \in \mathbb{N}_{\geq 2}$ . Put  $[d] = \{1, \dots, d\}$ . Let  $\mathbf{1}$  be the  $d$ -dimensional vector  $(1, \dots, 1)$ . For  $x, y \in [0, 1]^d$  we write  $x \leq y$  if  $x_i \leq y_i$  holds for all  $i \in [d]$ . We write  $[x, y] = \prod_{i \in [d]} [x_i, y_i]$  and use corresponding notations for open and half-open intervals. For a point  $x \in [0, 1]^d$  we denote by  $V_x$  the volume of the box  $[0, x]$ .

We restate the definition of  $\delta$ -covers in a slightly more general form than in Section 1:

**Definition 2.1.** Let  $\delta > 0$ . A pair  $(x, z)$  of points  $x, z \in [0, 1]^d$  is called a  $\delta$ -covering pair if  $x \leq z$  and  $V_z - V_x \leq \delta$ . In this case we call  $[x, z]$  a  $\delta$ -covering box. Let  $S \subseteq [0, 1]^d$ . We say that some finite subset  $\Gamma$  of  $S$  is a  $\delta$ -cover of  $S$  if for all  $y \in S$  there exist  $x, z \in \Gamma \cup \{0\}$  such that  $y \in [x, z]$  and  $(x, z)$  is a  $\delta$ -covering pair. We put

$$N(S, \delta) = \min\{|\Gamma| \mid \Gamma \text{ } \delta\text{-cover of } S\} \quad \text{and} \quad N(d, \delta) = N([0, 1]^d, \delta).$$

The following example serves as an illustration of the definition and gives us a first simple bound for  $N(d, \delta)$ :

**Example 2.2.** Let  $\Gamma_m$  be the equidistant grid  $\{1/m, 2/m, \dots, 1\}^d$  in  $[0, 1]^d$ , where  $m = \lceil d/\delta \rceil$ . It is easy to see that  $\Gamma_m$  is a  $\delta$ -cover of  $[0, 1]^d$ . We have  $|\Gamma_m| = m^d$ , hence  $N(d, \delta) \leq \lceil d/\delta \rceil^d$ .

Now we derive a better bound for  $N(d, \delta)$  as in Example 2.2 by constructing a *non-equidistant grid*  $\Gamma$  of the form

$$\Gamma = \{x_0, x_1, \dots, x_{\kappa(\delta, d)}\}^d, \quad (6)$$

where  $x_0, x_1, \dots, x_{\kappa(\delta, d)}$  is a decreasing sequence in  $]0, 1]$ . We calculate this sequence recursively in the following way: Put  $x_0 := 1$  and  $x_1 := (1 - \delta)^{1/d}$ . If  $x_i > \delta$ , then define  $x_{i+1} := (x_i - \delta)x_1^{1-d}$ . If  $x_{i+1} \leq \delta$ , then put  $\kappa(\delta, d) := i+1$ , otherwise proceed by calculating  $x_{i+2}$ .

It is easy to see that  $x_0, x_1, \dots$  is a decreasing sequence with  $x_i - x_{i+1} \leq x_{i+1} - x_{i+2}$ , since  $x_i - x_{i+1} = (x_{i+1} - x_{i+2})x_1^{d-1}$ . Therefore  $\kappa(\delta, d)$  is finite.

**Theorem 2.3.** *Let  $d \geq 2$ , and let  $0 < \delta < 1$ . Let  $\Gamma = \{x_0, x_1, \dots, x_{\kappa(\delta, d)}\}^d$  be as in (6). Then  $\Gamma$  is a  $\delta$ -cover of  $[0, 1]^d$ , and consequently*

$$N(d, \delta) \leq |\Gamma| \leq (\kappa(\delta, d) + 1)^d, \quad (7)$$

where

$$\kappa(\delta, d) = \left\lceil \frac{d}{d-1} \frac{\ln(1 - (1 - \delta)^{1/d}) - \ln(\delta)}{\ln(1 - \delta)} \right\rceil. \quad (8)$$

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  $\delta$  approaches 0.

*Proof.* Let  $x^* \in [0, 1]^d$ . Since  $\mathbf{1} \in \Gamma$ , there exists a uniquely determined minimal  $\bar{x} = (x_{i_1}, \dots, x_{i_d}) \in \Gamma$ ,  $i_1, \dots, i_d \in \{0, \dots, \kappa(\delta, d)\}$ , with  $x^* \leq \bar{x}$ .

*Case 1:*  $i_j < \kappa(\delta, d)$  for all  $j \in [d]$ . Then  $\underline{x} := (x_{i_1+1}, \dots, x_{i_d+1}) \in \Gamma$ , and  $\underline{x} \leq x^*$ , as  $\bar{x}$  is the minimal point in  $\Gamma$  with  $x^* \leq \bar{x}$ . The difference

$$\bar{x}_1 \dots \bar{x}_d - \underline{x}_1 \dots \underline{x}_d = \prod_{j=1}^d x_{i_j} - (1 - \delta)^{1-d} \prod_{j=1}^d (x_{i_j} - \delta)$$

is at most  $\delta$  if and only if

$$\prod_{j=1}^d (x_{i_j} - \delta) - \left( \prod_{j=1}^d x_{i_j} - \delta \right) (1 - \delta)^{d-1} \geq 0.$$

It can easily be shown by induction on  $d$  that the last inequality holds even for arbitrary  $x_{i_1}, \dots, x_{i_d} \in [\delta, 1]$ .

*Case 2:* There exists an index  $j \in [d]$  with  $i_j = \kappa(\delta, d)$ . Then we consider  $\underline{x} := 0$ . Obviously  $\underline{x} \leq x^* \leq \bar{x}$  and

$$\bar{x}_1 \dots \bar{x}_d - \underline{x}_1 \dots \underline{x}_d = x_{i_1} \dots x_{i_d} \leq x_{\kappa(\delta, d)} \leq \delta.$$

Thus we have shown that  $\Gamma$  is a  $\delta$ -cover of  $[0, 1]^d$ . The recursion formula for calculating  $x_0, \dots, x_{\kappa(\delta, d)}$  implies

$$\begin{aligned} x_i &= (1 - \delta)^{i \frac{1-d}{d}} - \delta \sum_{\alpha=1}^i (1 - \delta)^{\alpha \frac{1-d}{d}} \\ &= (1 - \delta)^{i \frac{1-d}{d}} - \delta (1 - \delta)^{\frac{1-d}{d}} \frac{1 - (1 - \delta)^{i \frac{1-d}{d}}}{1 - (1 - \delta)^{\frac{1-d}{d}}}. \end{aligned}$$

Thus  $x_{i+1} \leq \delta$  if and only if

$$i + 1 \geq \frac{d}{d-1} \frac{\ln(1 - (1 - \delta)^{\frac{1}{d}}) - \ln(\delta)}{\ln(1 - \delta)}.$$

This establishes (8). Furthermore, an elementary analysis reveals that the function  $F : ]0, 1[ \rightarrow \mathbb{R}$ , defined by

$$F(\delta) = \delta \frac{\ln(1 - (1 - \delta)^{1/d}) - \ln(\delta)}{\ln(1 - \delta)},$$

is monotonic decreasing with  $\lim_{\delta \rightarrow 0} F(\delta) = \ln(d)$ . Hence the estimate for  $\kappa(\delta, d)$ , given in Theorem 2.3, holds and is sharp.  $\square$

We want to prove an additional upper bound for  $N(d, \delta)$  with a better asymptotic behavior in  $d$ . In our application we shall use both bounds, i.e., (7) and (9), depending on the dimension  $d$  (see Theorem 3.2 and Remark 3.4). Let us introduce further definitions:

**Definition 2.4.** Let  $S \subset [0, 1]^d$  and  $\Gamma$  be a  $\delta$ -cover of  $S$ . Then

$$M(\Gamma) := \min(|A| + |E|),$$

where the minimum is taken over all subsets  $A, E$  of  $\Gamma$  with the property that for every  $y \in S$  there exists a  $\delta$ -covering pair  $(x, z)$ ,  $x \in A \cup \{0\}$ ,  $z \in E$ , with  $y \in [x, z]$ . We put

$$\tilde{N}(S, \delta) = \min\{M(\Gamma) \mid \Gamma \text{ } \delta\text{-cover of } S\} \quad \text{and} \quad \tilde{N}(d, \delta) = \tilde{N}([0, 1]^d, \delta).$$

Furthermore, we denote  $(1 - \delta)^{1/d}$  by  $a(d, \delta)$ . For  $0 \leq a \leq b \leq 1$  let

$$S^d([a, b]) = [0, b]^d \setminus [0, a]^d.$$

We list some elementary observations:

**Lemma 2.5.** (i) For all  $S \subset [0, 1]^d$ :  $N(d, \delta) \leq \tilde{N}(d, \delta) \leq 2N(d, \delta)$ .

(ii) *Subadditivity:* If  $S_1, S_2 \subseteq [0, 1]^d$ , then  $\tilde{N}(S_1 \cup S_2, \delta) \leq \tilde{N}(S_1, \delta) + \tilde{N}(S_2, \delta)$ .

(iii) *Scaling:* If  $\lambda \in \mathbb{R}_{>0}$  and  $S, \lambda S \subseteq [0, 1]^d$ , then  $\tilde{N}(S, \delta) = \tilde{N}(\lambda S, \lambda^d \delta)$ .

**Lemma 2.6.** Let  $d \geq 2$ ,  $\delta \in ]0, 1]$  and  $\delta' = 1 - (1 - \delta)^{(d-1)/d}$ . Then

$$\tilde{N}(S^d([a(d, \delta), 1]), \delta) \leq d\tilde{N}(d-1, \delta').$$

*Proof.* Note that  $a := a(d-1, \delta') = a(d, \delta)$ . Let  $x', z' \in [0, 1]^{d-1}$  such that  $(x', z')$  is a  $\delta'$ -covering pair. Put  $x = (x', a)$  and  $z = (z', 1)$ . Then  $x \leq z$ . If  $V_{x'} \geq 1 - \delta'$ , then  $V_z - V_x = V_{z'} - aV_{x'} \leq 1 - (1 - \delta)^{1/d}(1 - \delta') = \delta$ . If  $V_{x'} \leq 1 - \delta'$ , then  $V_z - V_x = V_{z'} - V_{x'} + (1 - a)V_{x'} \leq \delta' + (1 - a)(1 - \delta') = \delta$ .

Thus  $(x, z)$  is a  $\delta$ -covering pair. Hence  $\tilde{N}([0, 1]^{d-1} \times [a, 1], \delta) \leq \tilde{N}(d-1, \delta')$ , and the lemma follows from Lemma 2.5(ii).  $\square$

In the rest of this section, all  $O$ -notation refers to the variable  $\delta^{-1}$  only. So the implicit constants may depend on  $d$ .

**Theorem 2.7.** Let  $d \in \mathbb{N}$  and  $0 < \delta < 1$ . Then

$$\tilde{N}(d, \delta) \leq 2 \frac{d^d}{d!} \left( \delta^{-1} + \frac{d+1}{4} \right)^d. \quad (9)$$

In particular,  $N(d, \delta) \leq \sqrt{\frac{2}{\pi d}} e^d \delta^{-d} + O(\delta^{-d+1})$ .

*Proof.* Put  $n = \lceil \delta^{-1} \rceil$ ,  $\alpha(d) = 2d^d/d!$  and  $\beta(d) = \frac{d+1}{4}$ . We proceed by induction. Let  $d = 1$ . Then we define  $A = \{1/n, \dots, (n-1)/n\}$  and  $E = A \cup \{1\}$ . The set  $\Gamma = A \cup E$  is clearly a  $\delta$ -cover with  $M(\Gamma) \leq 2\delta^{-1} + 1$ . Let now  $d \geq 2$ , and put  $a_i = (1 - i\delta)^{1/d}$  for  $i = 0, \dots, n-1$  and  $a_n = 0$ .



Furthermore, let  $\delta_i = \delta/a_{i-1}^d$  for all  $i \in [n-1]$ . Since  $a_i/a_{i-1} = a(d, \delta/a_{i-1}^d)$ , we conclude from Lemma 2.5 and 2.6

$$\begin{aligned} \tilde{N}(d, \delta) &\leq \sum_{i=1}^n N(S^d([a_i, a_{i-1}]), \delta) \leq \sum_{i=1}^{n-1} \tilde{N}(S^d([\frac{a_i}{a_{i-1}}, 1]), \delta_i) + 1 \\ &\leq d \sum_{i=1}^{n-1} \tilde{N}(d-1, \delta'_i) + 1 \leq d \alpha(d-1) \sum_{i=1}^{n-1} ((\delta'_i)^{-1} + \beta(d-1))^{d-1} + 1. \end{aligned}$$

Since  $\delta'_i = 1 - (1 - \delta_i)^{(d-1)/d} \geq \frac{d-1}{d} \delta_i$ , we can majorize the last sum by

$$T := \sum_{i=1}^{n-1} \left( \frac{d}{d-1} (\delta^{-1} - i + 1) + \beta(d-1) \right)^{d-1}.$$

The inequality  $(\gamma + \tau)^{d-1} + (\gamma - \tau)^{d-1} \geq 2\gamma^{d-1}$  for all  $\gamma \geq 0$  shows that

$$\begin{aligned} T &\leq \int_{1/2}^{n-1/2} \left( \frac{d}{d-1} (\delta^{-1} - x + 1) + \beta(d-1) \right)^{d-1} dx \\ &\leq \frac{d^{d-2}}{(d-1)^{d-1}} \left[ \left( \delta^{-1} - x + 1 + \frac{d-1}{d} \beta(d-1) \right)^d \right]_{x=n-1/2}^{1/2} \\ &\leq \frac{d^{d-2}}{(d-1)^{d-1}} ((\delta^{-1} + \beta(d))^d - \beta(d)^d). \end{aligned}$$

Hence

$$\tilde{N}(d, \delta) \leq 2 \frac{d^{d-1}}{(d-1)!} ((\delta^{-1} + \beta(d))^d - \beta(d)^d) + 1 \leq 2 \frac{d^d}{d!} (\delta^{-1} + \beta(d))^d.$$

The inequality  $d! \geq \sqrt{2\pi d} d^d e^{-d}$  verifies our estimate for  $N(d, \delta)$ .  $\square$

At the end of this section we state a lower bound for  $N(d, \delta)$  and show how  $N(d, \delta)$  is related to the so-called covering number and the packing number of anchored boxes in the  $d$ -dimensional unit cube.

**Theorem 2.8.** *Let  $\delta \in ]0, 1]$ . Then*

$$N(d, \delta) \geq \frac{2}{5} \frac{d!}{d^d} \delta^{-d} - \frac{2}{5} d! \sum_{k=0}^{d-1} \frac{d^k |\ln(d\delta)|^k}{k!}. \quad (10)$$

*In particular,  $N(d, \delta) \geq \sqrt{d} e^{-d} \delta^{-d} + O(|\ln(\delta)|^{d-1})$ .*

Since a not much weaker bound can be derived directly from a result in [10] (see Remark 2.10), we give just the proof idea and omit tedious calculations. For  $z \in [0, 1]^d$  let  $A_\delta(z)$  be the set of all  $x \in [0, 1]^d$  such that  $(x, z)$  is a  $\delta$ -covering pair. It can be shown that  $\text{vol}(A_\delta(z))$  depends only on  $V_z$ . Let  $U(d, \delta)$  denote the set  $\{y \in [0, 1]^d \mid V_y > \delta\}$ . For all  $y, z \in U(d, \delta)$  with  $V_y \leq V_z$  follows  $\text{vol}(A_\delta(y)) \geq \text{vol}(A_\delta(z))$ . (This has to be shown only in the geometrical setting  $y \leq z$ , where it is evident.) Furthermore, one can derive  $\text{vol}(A_\delta(z)) \leq \frac{5}{2d!} \frac{\delta^d}{V_z^{d-1}}$  for all  $z \in U(d, d\delta)$  (e.g. by expanding  $\text{vol}(A_\delta(z))$  into a power series in  $\delta/V_z$ ). The following auxiliary lemma shows how we can use the last estimate to prove Theorem 2.8.

**Lemma 2.9.** *If  $\delta \in ]0, 1]$ , then*

$$N(d, \delta) \geq \int_{U(d, \delta)} \text{vol}(A_\delta(z))^{-1} dz. \quad (11)$$

*Proof.* Let  $\Gamma$  be a  $\delta$ -cover of  $[0, 1]^d$ . For each  $y \in U(d, \delta)$  we choose (in a measurable way) a  $\delta$ -covering pair  $(x(y), z(y))$  in  $\Gamma$  with  $x(y) \leq y \leq z(y)$ . Then  $\text{vol}(A_\delta(y)) \geq \text{vol}(A_\delta(z(y)))$ . Let  $E(\Gamma) = \{z(y) \mid y \in U(d, \delta)\}$ . If  $z \in E(\Gamma)$ , then  $\{y \mid z = z(y)\} \subset A_\delta(z)$ . These observations result in

$$\begin{aligned} |\Gamma| &\geq |E(\Gamma)| \geq \sum_{z \in E(\Gamma)} \frac{\text{vol}(\{y \mid z = z(y)\})}{\text{vol}(A_\delta(z))} = \sum_{z \in E(\Gamma)} \int_{\{y \mid z = z(y)\}} \text{vol}(A_\delta(z))^{-1} dy \\ &\geq \int_{U(d, \delta)} \text{vol}(A_\delta(z(y)))^{-1} dy \geq \int_{U(d, \delta)} \text{vol}(A_\delta(y))^{-1} dy. \end{aligned}$$

□

Thus

$$N(d, \delta) \geq \int_{U(d, d\delta)} \text{vol}(A_\delta(z))^{-1} dz \geq \frac{2d!}{5} \delta^{-d} \int_{U(d, d\delta)} V_z^{d-1} dz.$$

Hence Theorem 2.8 follows from the identity

$$\int_{U(d, d\delta)} V_z^{d-1} dz = d^{-d} \left( 1 - (d\delta)^d \sum_{k=0}^{d-1} \frac{(d \ln((d\delta)^{-1}))^k}{k!} \right),$$

which can be proved by induction on  $d$  using a suitable integral transform.

**Remark 2.10.** Consider  $\mathcal{C} = \{[0, x[ \mid x \in [0, 1]^d\}$ , endowed with the metric  $d_\lambda(C, C') = \text{vol}(C \Delta C')$ , where  $\lambda$  denotes the Lebesgue measure and  $\Delta$  the symmetric difference of two sets. Then the *covering number*  $N(\mathcal{C}, d_\lambda, \varepsilon)$  is the smallest number of closed  $\varepsilon$ -balls  $\{C' \in \mathcal{C} \mid d_\lambda(C', C) \leq \varepsilon\}$  that cover  $\mathcal{C}$ . A subset  $T$  of  $\mathcal{C}$  is  $\varepsilon$ -*separated*, if  $d_\lambda(C, C') > \varepsilon$  for distinct  $C, C' \in T$ , and the  $L^1(\lambda)$ -*packing number*  $M(\mathcal{C}, d_\lambda, \varepsilon)$  is the cardinality of the largest  $\varepsilon$ -separated subset of  $\mathcal{C}$ . It is easy to verify that

$$N(\mathcal{C}, d_\lambda, \varepsilon) \leq M(\mathcal{C}, d_\lambda, \varepsilon) \leq N(\mathcal{C}, d_\lambda, \varepsilon/2) \leq N(d, \varepsilon) + 1.$$

Thus, e.g., our upper bounds for  $N(d, \delta)$  lead to the upper bound

$$M(\mathcal{C}, d_\lambda, \varepsilon) \leq \sqrt{\frac{2}{\pi d}} e^d \varepsilon^{-d} + O(\varepsilon^{-d+1}).$$

For fixed  $d$  and small  $\varepsilon$  this bound improves the bound that follows from a celebrated result of Haussler [6, Cor.1]:  $M(\mathcal{C}, d_\lambda, \varepsilon) \leq (d+1)2^d e^{d+1} \varepsilon^{-d}$ . On the other hand, a lower bound for  $N(\mathcal{C}, d_\lambda, \varepsilon)$  induces a lower bound for  $N(d, \delta)$ . After we achieved the result of Theorem 2.8, we became aware of a preprint version of [10]. There, in the course of the proof of Theorem 2, the estimate  $N(\mathcal{C}, d_\lambda, \varepsilon) \geq d! (8ed)^{-d} \varepsilon^{-d}$  was established, inducing the weaker lower bound  $N(d, \delta) \geq \sqrt{2\pi d} (4e)^{-d} \delta^{-d} - 1$ .

### 3 Applications to Star-Discrepancy

In this section we use our constructions of small  $\delta$ -covers to prove upper bounds for the  $L^\infty$ -star-discrepancy and its inverse. We observe the following simple approximation property:

**Lemma 3.1.** *Let  $\Gamma$  be a  $\delta$ -cover of  $[0, 1]^d$ . Then for all  $t_1, \dots, t_n \in [0, 1]^d$*

$$d_\infty^*(t_1, \dots, t_n) \leq \max_{x \in \Gamma} \left| V_x - \frac{1}{n} \sum_{i=1}^n 1_{[0, x[}(t_i) \right| + \delta. \quad (12)$$

*Proof.* Let  $x \in [0, 1]^d$ . Then there exist  $\underline{x}, \bar{x} \in \Gamma \cup \{0\}$  with  $\underline{x} \leq x \leq \bar{x}$  and  $V_{\bar{x}} - V_{\underline{x}} \leq \delta$ . Therefore (12) follows from

$$V_{\bar{x}} - \delta - \frac{1}{n} \sum_{i=1}^n 1_{[0, \bar{x}[}(t_i) \leq V_x - \frac{1}{n} \sum_{i=1}^n 1_{[0, x[}(t_i) \leq V_{\underline{x}} + \delta - \frac{1}{n} \sum_{i=1}^n 1_{[0, \underline{x}[}(t_i).$$

□

Note that similar results were proved in [8, Subsec. 2.1], and [12, Sec. 5]. Lemma 3.1 combined with our results on  $\delta$ -covers yield the following:

**Theorem 3.2.** *Let  $d \geq 2$  and  $\varepsilon > 0$ .*

(i) *If  $\varepsilon \leq 8/(d+1)$ , then*

$$n_{\infty}^*(\varepsilon, d) \leq \left\lceil 2\varepsilon^{-2} \left( d \ln \left( \frac{4e}{\varepsilon} \right) + \ln(2) \right) \right\rceil. \quad (13)$$

(ii) *For all  $0 < \varepsilon \leq 1$  we have, with  $\kappa(\varepsilon/2, d)$  as in Theorem 2.3,*

$$n_{\infty}^*(\varepsilon, d) \leq \left\lceil 2\varepsilon^{-2} (d \ln(\kappa(\varepsilon/2, d) + 1) + \ln(2)) \right\rceil \quad (14)$$

*and, with  $\rho = 3 \ln(3) / \sqrt{2(3 \ln(3) + \ln(2))} < 1.167$ ,*

$$d_{\infty}^*(n, d) \leq \sqrt{2} n^{-1/2} (d \ln(\lceil \rho n^{1/2} \rceil + 1) + \ln(2))^{1/2}. \quad (15)$$

*Proof.* We combine our upper bounds for the cardinality of minimal  $\delta$ -covers from Section 2 with the probabilistic argument from [8], proof of Theorem 1: For  $\delta > 0$  let  $\Gamma$  be a  $\delta$ -cover of  $[0, 1]^d$ . Let  $\tau_1, \dots, \tau_n$  be independent random variables, uniformly distributed in  $[0, 1]^d$ . For  $x \in [0, 1]^d$  we define

$$\xi_x^{(i)} = \xi_x^{(i)}(\tau_i) := V_x - 1_{[0, x]}(\tau_i), \quad i = 1, \dots, n.$$

The range of  $\xi_x^{(i)}$  is contained in an interval of length one and the expectation  $\mathbb{E}(\xi_x^{(i)})$  is zero. Thus Hoeffding's inequality (see, e.g., [15, p.191-2]) implies

$$\mathbb{P} \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_x^{(i)} \right| \geq \delta \right\} \leq 2 \exp(-2\delta^2 n).$$

In the case where  $x \in \{0, \mathbf{1}\}$ , the probability on the left hand side of the last inequality is even zero. Since  $\mathbf{1}$  is contained in  $\Gamma$ , we get

$$\mathbb{P} \{ d_{\infty}^*(\tau_1, \dots, \tau_n) \leq 2\delta \} \geq \mathbb{P} \left\{ \max_{x \in \Gamma} \left| \frac{1}{n} \sum_{i=1}^n \xi_x^{(i)} \right| \leq \delta \right\} > 1 - 2|\Gamma| \exp(-2\delta^2 n).$$

Therefore the condition

$$2\delta^2 n \geq \ln(|\Gamma|) + \ln(2) \quad (16)$$

implies  $\mathbb{P}\{d_\infty^*(\tau_1, \dots, \tau_n) \leq 2\delta\} > 0$ . Hence  $n_\infty^*(\varepsilon, d) \leq \lceil \frac{1}{2\delta^2}(\ln(|\Gamma|) + \ln(2)) \rceil$  for  $\varepsilon = 2\delta$ . Therefore (7) and (9) lead to (14) and (13).

To derive from (16) an estimate for  $d_\infty^*(n, d)$ , we need more information about the dependence of  $|\Gamma|$  on  $\delta$ . So let  $\Gamma$  be a non-equidistant grid as in (6). Then any solution of

$$\delta^2 \geq \frac{1}{2n} \left( d \ln \left( \left\lceil \frac{d}{d-1} \frac{\ln(d)}{\delta} \right\rceil + 1 \right) + \ln(2) \right) \quad (17)$$

fulfills (16) and leads to  $d_\infty^*(n, d) \leq 2\delta$ . Let  $\delta = \delta(n)$  be a half times the right hand side of (15). Then  $\delta$  fulfills (17) if

$$\rho n^{1/2} \geq \frac{d}{d-1} \frac{\ln(d)}{\delta}.$$

The last inequality holds for all  $n \in \mathbb{N}$  if and only if

$$\rho \geq \frac{\sqrt{2}d}{d-1} \frac{\ln(d)}{(d \ln(3) + \ln(2))^{1/2}}.$$

An elementary analysis shows that the right hand side takes its maximum in  $d = 3$  and this maximum is nothing but  $\rho$ .  $\square$

**Remark 3.3.** The proof of Theorem 3.2 shows that our upper bounds for the star-discrepancy and its inverse will improve if we are able to derive better bounds for  $N(d, \delta)$ . Nevertheless, our lower bound (10) stresses that this proof method cannot lead to a better result than  $n_\infty^*(\varepsilon, d) = O(d\varepsilon^{-2} \ln(1/\varepsilon))$  and  $d_\infty^*(n, d) = O(d^{1/2}n^{-1/2} \ln(n/d)^{1/2})$ .

**Remark 3.4.** If we consider errors  $\varepsilon \leq 8/(d+1)$ , then some calculations show that for dimension  $d \leq 215$  estimate (14) is better than (13), whereas for  $d \geq 225$ , it is preferable to use estimate (13). (Recall that the proof of (14) made use of (7), while (13) was derived with the help of (9).)

**Remark 3.5.** A similar probabilistic argument as in the proof of our Theorem 3.2 was used in the recent papers [9, 12]. But instead of directly applying Hoeffding's inequality, the authors used Bennett's inequality [15, p.192] to derive another upper bound of Hoeffding type. A closer look at the random variables under consideration reveals that the direct use of Hoeffding's inequality, as stated in [15, Cor.B3], leads to a better result. For example the

independent random variables  $\zeta_{y,U}^{(i)}, i = 1, \dots, n$ , in [9], proof of Theorem 3, fulfill  $a_i \leq \zeta_{y,U}^{(i)} \leq b_i$  with  $b_i - a_i = 1$ . Therefore [15, Cor. B3] gives

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n \zeta_{y,U}^{(i)}\right| \geq \delta\right\} \leq 2 \exp(-2n\delta^2).$$

This result is a little bit better than the result in [9] derived from Bennett's inequality:

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n \zeta_{y,U}^{(i)}\right| \geq \delta\right\} \leq 2 \exp(-cn\delta^2), \quad \text{where } c < 1.3.$$

Similarly, the bound in [15] is sharper than the bound stated in [12, Prop.5.1].

At the end of this section we want to evaluate the right hand side of estimate (14) explicitly for some values of  $d$  and  $\varepsilon$ . We take the same values as in [8], so it is easier to compare our results with the bounds for the inverse given there at the end of Section 2.

$n_\infty^*(0.45, 5) \leq 116$	$n_\infty^*(0.1, 5) \leq 3828$
$n_\infty^*(0.45, 10) \leq 244$	$n_\infty^*(0.1, 10) \leq 8003$
$n_\infty^*(0.45, 20) \leq 514$	$n_\infty^*(0.1, 20) \leq 16648$
$n_\infty^*(0.45, 40) \leq 1103$	$n_\infty^*(0.1, 40) \leq 34679$
$n_\infty^*(0.45, 60) \leq 1686$	$n_\infty^*(0.1, 60) \leq 53020$
$n_\infty^*(0.45, 80) \leq 2291$	$n_\infty^*(0.1, 80) \leq 71777$

The bounds in [8] were achieved by using the same technique that we adapted in the proof of Theorem 3.2 and by analyzing the average behavior of the  $L^p$ -star-discrepancy for even integers  $p$ . Our bounds are smaller by factors between 5 and 8.1 than the bounds in [8] that make use of Hoeffding's inequality, and they are still smaller than the bounds resulting from the average  $L^p$ -star-discrepancy analysis: roughly by a factor 3 for  $\varepsilon = 0.45$  and 1.6 for  $\varepsilon = 0.1$ .

## 4 Deterministic Construction of Low-Discrepancy Points

In this section we give a deterministic algorithm for the construction of points satisfying the  $L^\infty$ -star-discrepancy bound of Theorem 3.2 by derandomizing the probabilistic argument. We invoke an algorithmic version of the Hoeffding inequality proved in [19].

### 4.1 The Basics of Derandomization

Here we explain the basic idea behind the algorithm, in fact give its basic form, which will lead to the deterministic construction in the next subsection. In the literature it is often referred to as the *method of conditional probabilities*.

Let  $r, n \in \mathbb{N}$ . We consider the probability space  $(\Omega, \mathbb{P})$ , where  $\Omega = [r]^n$ , the powerset  $\mathcal{P}(\Omega)$  is the  $\sigma$ -algebra, and  $\mathbb{P}$  is a product measure of probability measures on  $[r]$ . Let  $E \subseteq \Omega$  be an event with  $\mathbb{P}(E) > 0$ , and let  $E^c$  denote the complement of  $E$ . For  $y \in \Omega$ ,  $l \in [n]$  and  $\omega_1, \dots, \omega_l \in [r]$  the conditional probability of  $E^c$  under the condition that the first  $l$  components  $y_1, \dots, y_l$  of the random vector  $y$  are  $\omega_1, \dots, \omega_l$  is denoted by

$$\mathbb{P}(E^c \mid \omega_1, \dots, \omega_l) := \mathbb{P}(E^c \mid y_1 = \omega_1, \dots, y_l = \omega_l).$$

The following simple procedure constructs a vector in  $E$ .

**Algorithm** CONDPROB

*Input:* An event  $E \subseteq \Omega$  with  $\mathbb{P}(E) > 0$ .

*Output:* A vector  $x \in E$ .

For  $l = 1, \dots, n$  do:

If  $x_1, \dots, x_{l-1} \in [r]$  have been selected, set  $y_l = x_l$ , where  $x_l$  minimizes the function  $\omega \mapsto \mathbb{P}(E^c \mid x_1, \dots, x_{l-1}, y_l = \omega)$ ,  $\omega \in [r]$ .

**Proposition 4.1.** *The algorithm CONDPROB is correct.*

*Proof.* Let  $l \in [n]$  and  $x_1, \dots, x_{l-1} \in [r]$ . The conditional probabilities in CONDPROB can be written as convex combinations:

$$\mathbb{P}(E^c \mid x_1, \dots, x_{l-1}) = \sum_{\omega \in [r]} \mathbb{P}\{y_l = \omega\} \cdot \mathbb{P}(E^c \mid x_1, \dots, x_{l-1}, y_l = \omega). \quad (18)$$

By the choice of the components of  $x$  and the assumption  $\mathbb{P}(E) > 0$ ,

$$1 > \mathbb{P}(E^c) \geq \mathbb{P}(E^c | x_1) \geq \cdots \geq \mathbb{P}(E^c | x_1, \dots, x_n) \in \{0, 1\},$$

so  $\mathbb{P}(E^c | x_1, \dots, x_n) = 0$ , and consequently  $x \in E$ .  $\square$

The efficiency of this algorithm depends on the efficient computation of the conditional probabilities. In general, it seems to be hopeless to compute conditional probabilities directly. But for the purpose of derandomization it suffices to compute upper bounds for the conditional probabilities, which take over the role of the conditional probabilities. Such upper bounds were introduced by Spencer [17] in the hyperbolic cosine algorithm, and were later defined in a rigorous way as so called pessimistic estimators by Raghavan [16]. We give a little modified definition, covering multi-valued random variables.

**Definition 4.2.** Let  $\mathcal{U}$  be a family consisting of functions  $U_l : [r]^l \mapsto \mathbb{Q}$ ,  $l \in [n]$ , and a constant  $U_0 \in \mathbb{Q}$ . Let  $E \subseteq \Omega$  be an event with  $\mathbb{P}(E) \geq \theta$  for some  $0 < \theta < 1$ .  $\mathcal{U}$  is called a *pessimistic estimator* for the event  $E$  if for each  $l \in [n]$  the following conditions are satisfied:

- (i)  $\mathbb{P}(E^c) \leq U_0 \leq 1 - \theta$ .
- (ii)  $\mathbb{P}(E^c | \omega_1, \dots, \omega_l) \leq U_l(\omega_1, \dots, \omega_l)$  for all  $\omega_1, \dots, \omega_l \in [r]$ .
- (iii) Given  $\omega_1, \dots, \omega_{l-1}$  there exists an  $\omega_l \in [r]$  such that  $U_l(\omega_1, \dots, \omega_l) \leq U_{l-1}(\omega_1, \dots, \omega_{l-1})$ .
- (iv) Each value  $U_l(\omega_1, \dots, \omega_l)$  can be computed in time polynomially bounded in  $n, r$  and  $\log(1/\theta)$ .

With a pessimistic estimator we have a polynomial-time implementation of the algorithm CONDPROB.

**Algorithm DERAND**

*Input:* An event  $E \subseteq \Omega$  with  $\mathbb{P}(E) \geq \theta > 0$  and a pessimistic estimator  $\mathcal{U}$  for  $E$ .

*Output:* A vector  $x \in E$ .

For  $l = 1, \dots, n$  do:

If  $x_1, \dots, x_{l-1} \in [r]$  have been selected, choose  $x_l \in [r]$  as the minimizer of the function  $\omega \mapsto U_l(x_1, \dots, x_{l-1}, \omega)$ ,  $\omega \in [r]$ .



**Proposition 4.3.** *The algorithm DERAND runs in polynomial time in  $n, r$  and  $\log(1/\theta)$ , and the output  $x$  is contained in  $E$ .*

*Proof.* Since  $\mathcal{U}$  is a pessimistic estimator, each  $U_l(x_1, \dots, x_{l-1}, \omega)$ ,  $\omega \in [r]$ , can be computed in polynomial time, thus the minimizer  $x_l$  can be computed in polynomial time. The vector  $x$  satisfies

$$\mathbb{P}(E^c \mid x_1, \dots, x_n) \leq U_n(x_1, \dots, x_n) \leq \dots \leq U_0 \leq 1 - \theta < 1,$$

hence  $\mathbb{P}(E^c \mid x_1, \dots, x_n) = 0$ , and therefore  $x \in E$ . □

If we want to apply the algorithm DERAND, we have to circumnavigate two obstacles: First, we have to prove a non-zero probability statement of the event under consideration. This is often done with the help of large deviation inequalities for sums of random variables—the approach that we shall use in the next subsection. Furthermore, we need a suitable pessimistic estimator. If the random variables are independent such estimators can be constructed. Again, we shall comment on this in the next subsection.

## 4.2 Applications to Low-Discrepancy Points

Let  $n \in \mathbb{N}$  and  $\delta \in ]0, 1/3]$ . Let  $m = \lceil d/\delta \rceil$ , and let  $\Gamma_m$  be the equidistant grid with mesh size  $1/m$  as defined in Example 2.2. We put  $r := m^d$  and define the shifted grid  $\Gamma_m^s = \{g_1, \dots, g_r\}$  to be the set  $\Gamma_m - \frac{1}{m}\mathbf{1}$ . Let us consider  $n$  mutually independent random variables  $X_1, \dots, X_n$  with values in  $\Gamma_m^s$ . Suppose that the distribution of each  $X_j$  is uniform, i.e.,

$$\mathbb{P}\{X_j = g_k\} = 1/r \quad \text{for all } k \in [r].$$

The randomized algorithm for the generation of  $n$  points in  $[0, 1]^d$  simply is to take the outcome of the random variables  $X_1, \dots, X_n$  as the points under consideration.

We can analyze the discrepancy behavior of random point sets of this type using the Hoeffding inequality. In order to use the derandomization results in [19], let us adopt the notation there: Let  $X_{jk}$ ,  $1 \leq j \leq n$ ,  $1 \leq k \leq r$ , be the 0/1-random variable which is 1 if  $X_j = g_k$  and 0 else. Let  $N := N(d, \delta)$ , and let  $\Gamma = \{x_1, \dots, x_N\}$  be a minimal  $\delta$ -cover. For  $i \in [N]$  put

$$b_{ik} = \begin{cases} 1 & \text{if } g_k \in [0, x_i[ \\ 0 & \text{else.} \end{cases}$$

The (signed) discrepancy of  $[0, x_i[$  then is

$$\psi_i := V_{x_i} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^r b_{ik} X_{jk}. \quad (19)$$

**Lemma 4.4.** *We have  $|\mathbb{E}(\psi_i)| \leq \delta$  for all  $i \in [N]$ .*

*Proof.* We consider an arbitrary  $i \in [N]$ . Let  $J(i) := \{k \mid g_k \in \Gamma_m^s \cap [0, x_i[ \}$ . We have for all  $j \in [n]$

$$\mathbb{E} \left( \sum_{k=1}^r b_{ik} X_{jk} \right) = \sum_{k \in J(i)} \mathbb{P}\{X_{jk} = 1\} = \sum_{k \in J(i)} \mathbb{P}\{X_j = g_k\} = \frac{|J(i)|}{r}.$$

Let  $x_i^*$  be the uniquely determined point in  $\Gamma_m$  with  $\Gamma_m^s \cap [0, x_i^*[ = \Gamma_m^s \cap [0, x_i[$ . We get  $|J(i)| = rV_{x_i^*}$  and

$$\mathbb{E}(\psi_i) = V_{x_i} - \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left( \sum_{k=1}^r b_{ik} X_{jk} \right) = V_{x_i} - V_{x_i^*}.$$

Since  $\Gamma_m$  is in particular a  $\delta$ -cover of  $[0, 1]^d$ , we have  $|V_{x_i} - V_{x_i^*}| \leq \delta$ .  $\square$

For  $i \in [N]$  let  $\lambda_i \in \mathbb{R}$  and  $E_i$  be the event  $\{|\psi_i - \mathbb{E}(\psi_i)| \leq \lambda_i\}$ , and let  $E_i^c$  be the complement of  $E_i$ . Hoeffding's inequality implies the following:

**Lemma 4.5.** *For all  $i = 1, \dots, N$  we have  $\mathbb{P}(E_i^c) \leq 2e^{-2\lambda_i^2 n}$ .*

The derandomization problem is to construct a vector  $t = (t_1, \dots, t_n) \in (\Gamma_m^s)^n$  such that  $t$  satisfies all events  $E_i$ , i.e.,  $t \in \bigcap_{i=1}^N E_i$ . If

$$\sum_{i=1}^N 2e^{-2\lambda_i^2 n} \leq 1 - \theta \quad \text{for some } 0 < \theta \leq 1, \quad (20)$$

then by Lemma 4.5

$$\mathbb{P}(\bigcap_{i=1}^N E_i) = 1 - \mathbb{P}(\bigcup_{i=1}^N E_i^c) \geq \theta.$$

In that case  $\bigcap_{i=1}^N E_i$  is non-empty and at least some  $t \in \bigcap_{i=1}^N E_i$  exists. We have

**Theorem 4.6.** *If (20) is satisfied, then a vector  $t \in \cap_{i=1}^N E_i$  can be constructed in  $O(rn^2N \ln(\frac{rnN}{\theta}))$  time.*

Note that Theorem 4.6 is a special case of Theorem 2.13 in [19]. The observation which has to be made is that Theorem 2.13 in [19] is also valid if we use the Hoeffding bound instead of the Angluin-Valiant bounds (which are used in [19] for the proof of Theorem 2.13).

The proof of Theorem 4.6 involves the conditional probability method and the computation of a pessimistic estimator for the events  $E_1^c, \dots, E_N^c$  (see the previous section). Here we give a brief sketch of the form of the pessimistic estimator. We deal with the deviation above resp. below the expectation  $\mathbb{E}(\psi_i)$  separately. Let  $E_i^+$  be the event  $\psi_i > \mathbb{E}(\psi_i) + \lambda_i$  and  $E_i^-$  the event  $\psi_i < \mathbb{E}(\psi_i) - \lambda_i$ ,  $i = 1, \dots, N$ . Under the condition that the first  $l$  random variables  $X_1, \dots, X_l$  have values  $\omega_1, \dots, \omega_l \in \Gamma_m^s$ , we get from Markov's inequality

$$\mathbb{P}(E_i^\pm | \omega_1, \dots, \omega_l) \leq e^{-t_i \lambda_i} \mathbb{E}(e^{\pm t_i (\psi_i - \mathbb{E}(\psi_i))} | \omega_1, \dots, \omega_l), \quad (21)$$

where  $t_i$  is a positive scalar that has to be chosen in an appropriate way. We may define  $\tilde{U}_i^\pm(\omega_1, \dots, \omega_l)$  as the right hand side of (21) and write

$$\tilde{U}_i(\omega_1, \dots, \omega_l) := \tilde{U}_i^+(\omega_1, \dots, \omega_l) + \tilde{U}_i^-(\omega_1, \dots, \omega_l), i = 1, \dots, N.$$

The functions

$$\tilde{U}(\omega_1, \dots, \omega_l) := \sum_{i=1}^N \tilde{U}_i(\omega_1, \dots, \omega_l), 1 \leq l \leq N,$$

form a pessimistic estimator for the event  $E := \cap_{i=1}^N E_i$  in the sense of Definition 4.2 if we neglect the efficient computability condition (iv). The main work in [19] is to show that  $\tilde{U}$  can be approximated by a low-degree polynomial involving only rational parameters. This leads to the replacement of  $\tilde{U}$  by an efficiently computable function  $U$ , which still is a pessimistic estimator for the event  $E$ .

The application to the star-discrepancy problem is now straightforward. In the situation where we take  $\lambda_i = \delta$  for all  $i \in [N]$ , (20) is equivalent to

$$2\delta^2 n \geq \ln(N) + \ln\left(\frac{2}{1-\theta}\right).$$

Thus we can use Theorem 4.6 with  $\theta = 1/2$  to obtain the following result:

**Theorem 4.7.** *Let  $d \geq 2$  and  $\delta \in ]0, 1/3]$ . Let  $N = N(d, \delta)$ . If*

$$2\delta^2 n \geq \ln(N) + \ln(4), \quad (22)$$

*then points  $\tau_1, \dots, \tau_n \in [0, 1]^d$  satisfying*

$$d_\infty^*(\tau_1, \dots, \tau_n) \leq 3\delta$$

*can be constructed in  $O(d^d \delta^{-d} n^2 N \ln(d^d \delta^{-d} n N))$  time.*

*Proof.* We apply Theorem 4.6 and recall that  $r = \lceil \frac{d}{\delta} \rceil^d$ . This gives the construction of  $t = (t_1, \dots, t_n)$  such that  $|\psi_i(t) - \mathbb{E}(\psi_i)| \leq \delta$  for all  $i = [N]$ . Hence Lemma 4.4 implies  $|\psi_i(t)| \leq 2\delta$  for all  $i$ . Put  $\tau_j := X_j(t)$  for  $j = 1, \dots, n$ . By definition of  $\psi_i$ ,  $i = 1, \dots, N$ , and Lemma 3.1,

$$d_\infty^*(\tau_1, \dots, \tau_n) \leq \max_{1 \leq i \leq N} |\psi_i(t)| + \delta.$$

So  $d_\infty^*(\tau_1, \dots, \tau_n) \leq 2\delta + \delta = 3\delta$ .  $\square$

Let  $\varepsilon > 0$  be given and let  $\delta = \varepsilon/3$ . According to Theorem 2.3, we have  $N(d, \delta) \leq (\lceil \frac{d}{d-1} \frac{\ln(d)}{\delta} \rceil + 1)^d$  for  $d \geq 2$ . Therefore Theorem 4.7 ensures that we can construct points  $\tau_1, \dots, \tau_n \in [0, 1]^d$  with  $d_\infty^*(\tau_1, \dots, \tau_n) \leq \varepsilon$  and

$$n \leq \left\lceil \frac{9}{2\varepsilon^2} \left( d \ln \left( \left\lceil \frac{3d}{d-1} \frac{\ln(d)}{\varepsilon} \right\rceil + 1 \right) + \ln(4) \right) \right\rceil,$$

i.e.,  $n = O(\varepsilon^{-2} d (\ln \ln(d) + \ln(\varepsilon^{-1})))$ . Let us cite Problem 1(c) stated by Heinrich [7]:

*Problem:* For each  $\varepsilon > 0$  and  $d \in \mathbb{N}$ , give a construction of a point set  $\{t_1, \dots, t_n\} \subseteq [0, 1]^d$  with  $n \leq c_\varepsilon d^{\kappa_\varepsilon}$  and  $d_\infty^*(t_1, \dots, t_n) \leq \varepsilon$ , where  $c_\varepsilon$  and  $\kappa_\varepsilon$  are positive constants which may depend on  $\varepsilon$ , but not on  $d$ .

The discussion above shows that our algorithm formally solves the problem, but the drawback is that we just can ensure a running time smaller than  $O(C^d d^d \ln(d)^d \varepsilon^{-2(d+2)} \ln(\varepsilon^{-1})^3)$ ,  $C$  a constant. (In [7] the term “construction” is not specified. In particular, there is nothing written about computing time—but probably the author had in mind a computing time polynomial in  $\varepsilon^{-1}$  and  $d$ .)

Let now  $n$  and  $d \in \mathbb{N}$  be given. Similar as in the proof of Theorem 3.2, we can show that for

$$\delta = \frac{1}{\sqrt{2n}} \left( d \ln(\lceil \sigma n^{1/2} \rceil + 1) + \ln(4) \right)^{1/2}$$

condition (22) holds. Here  $\sigma < 1.1$ , as an elementary analysis reveals. It follows from Theorem 4.7 that we can construct points  $\tau_1, \dots, \tau_n$  with

$$d_\infty^*(\tau_1, \dots, \tau_n) \leq O(d^{1/2} n^{-1/2} \ln(n)^{1/2}). \quad (23)$$

Thus our construction formally solves Problem 1(d) from [7], which is

*Problem:* For each  $n, d \in \mathbb{N}$ , give a construction of a point set  $\{t_1, \dots, t_n\} \subseteq [0, 1]^d$  with  $d_\infty^*(t_1, \dots, t_n) \leq cd^\kappa n^{-\alpha}$ , where  $c, \kappa$  and  $\alpha$  are positive constants not depending on  $n$  and  $d$ .

Also here we have the drawback of a large running time, which is bounded from above by  $O(C^d n^{d+2} \ln(d)^d / \ln(n)^{d-1})$ .

### 4.3 Comparison with other Methods

#### A Trivial Search Algorithm

Let  $\varepsilon > 0$  and  $d \geq 2$  be given, and let  $\Gamma$  be a  $\varepsilon/2$ -cover. We know from the proof of Theorem 3.2 and the discussion in the previous subsection that, with probability at least  $\theta$ , we can find points  $t_1, \dots, t_n$  in  $\Gamma \cup \{0\}$  with discrepancy  $d_\infty^*(t_1, \dots, t_n) \leq \varepsilon$  and  $n = \lceil 2\varepsilon^{-2}(\ln(|\Gamma|) + \ln(2/(1-\theta))) \rceil$ . (If we take, e.g.,  $\theta = 1 - 2|\Gamma|^{-1}$ , we have  $n = \lceil 4\varepsilon^{-2} \ln(|\Gamma|) \rceil$ .) So a trivial algorithm could test all possible combinations of  $n$  points from  $\Gamma \cup \{0\}$  whether they have discrepancy smaller  $\varepsilon$  or not.

For a better comparison with the discussion at the end of the previous subsection, let  $\Gamma$  be the non-equidistant grid (6). If we want to test whether a given  $n$ -point configuration  $t_1, \dots, t_n \in \Gamma \cup \{0\}$  has discrepancy smaller  $\varepsilon$ , it suffices (according to Lemma 3.1) to show

$$\left| V_x - \frac{1}{n} \sum_{i=1}^n 1_{[0,x]}(t_i) \right| \leq \frac{\varepsilon}{2} \quad \text{for all } x \in \Gamma.$$

So we have to carry out  $O(nd)$  arithmetic operations and comparisons for each  $x \in \Gamma$ , which results in costs of order  $O(C^d \ln(d)^d \varepsilon^{-d-2} \ln(\varepsilon^{-1}))$ ,  $C > 2$  a suitable constant. (Note that we here already took advantage of our results from Section 2. If one uses an equidistant grid, these costs would be of order  $O(C^d d^d \varepsilon^{-d-2} \ln(\varepsilon^{-1}))$ , which is not very far away from the order of the whole running time of our derandomized algorithm.)

The number of possible  $n$ -point configurations in  $\Gamma$  can be estimated by

$$\binom{|\Gamma|}{n} \geq \left( \frac{|\Gamma|}{n} \right)^n \geq (C^d \ln(d)^d \varepsilon^{2-d})^{c\varepsilon^{-2}d},$$

where  $c, C$  are suitable constants. (Again, we would get a worse result if we take an equidistant grid instead of our non-equidistant grid  $\Gamma$ .) Even if we choose  $\theta$  of the order  $1 - 2|\Gamma|^{-1}$ , we cannot ensure that the search algorithm needs less than

$$(1 - \theta)(C^d \ln(d)^d \varepsilon^{2-d})^{c\varepsilon^{-2d}} \geq (C^d \ln(d)^d \varepsilon^{2-d})^{c(1-o(1))\varepsilon^{-2d}}$$

tests before it returns a sufficiently good  $n$ -point set. (Although it is extremely unlikely that we need such a large number of tests.)

Hence the worst-case running time of the trivial search algorithm exceeds the running time of our derandomized algorithm, since it is not only exponential in  $d$ , but exponential in  $(d/\varepsilon)^2$ .

### Small Sample Spaces

In the literature one can find constructions of points having small discrepancy with respect to all axis-parallel boxes in the  $d$ -dimensional unit cube. This kind of discrepancy is sometimes called *unanchored* or *extreme discrepancy*, and its precise definition is

$$d_\infty(t_1, \dots, t_n) := \sup_{x, y \in [0, 1]^d} \left| \text{vol}([x, y]) - \frac{1}{n} \sum_{i=1}^n 1_{[x, y]}(t_i) \right|$$

for points  $t_1, \dots, t_n \in [0, 1]^d$ . The quantities  $d_\infty(n, d)$  and  $n_\infty(d, \varepsilon)$  are defined in the natural way. According to [8, 10], we have the same bounds

$$cd\varepsilon^{-1} \leq n_\infty(d, \varepsilon) \leq Cd\varepsilon^{-2}$$

as for the star-discrepancy (with maybe different constants). As pointed out in [4], our bounds for  $\delta$ -covers and the bounds from Theorem 3.2 can also be transferred to the situation of the extreme discrepancy—roughly one has to replace  $d$  by  $2d$  on the right hand side of the estimates. (A similar transference result holds for bounds based on the average  $L^p$ -discrepancy, see [5].) Thus it seems reasonable to compare constructions for the extreme discrepancy with our deterministic algorithm for the star-discrepancy.

Even et al. [3] state efficient constructions of small sample spaces  $S_1, S_2 \subseteq [0, 1]^d$  with extreme discrepancy at most  $\varepsilon$  and  $|S_1| = (d/\varepsilon)^{O(\log(1/\varepsilon))}$ ,  $|S_2| = (d/\varepsilon)^{O(\log(d))}$ . They also constructed a sample space  $S_3$  whose discrepancy with respect to all axis-parallel boxes that are non-trivial in at most  $k$  dimensions (i.e., boxes of the form  $\prod_{i=1}^d r_i$ , where we have  $r_j = [0, 1[$  for at least

$d - k$  indices  $j$ ) is less than  $\delta > 0$ . The size of  $S_3$  is polynomial in  $\log(d)$ ,  $2^k$  and  $1/\delta$ , and the construction is based on small bias probability spaces as created in [13]. Chari, Rohatgi and Srinivasan [1] proved by dimension reduction that the sample space  $S_3$  has already extreme discrepancy less than  $\varepsilon$  if one chooses

$$k = O(\log(1/\varepsilon)) \quad \text{and} \quad \delta = \frac{\varepsilon}{2} \left( \frac{k}{de^2} \right)^k.$$

This results in a construction of

$$n = \text{poly} \left( 1/\varepsilon, (d/\log(1/\varepsilon))^{\log(1/\varepsilon)} \right)$$

points  $t_1, \dots, t_n$  with  $d_\infty(t_1, \dots, t_n) \leq \varepsilon$ , improving the bounds of Even et al.. Since the computing time of this construction is bounded from below by  $n$ , it is also not polynomial in  $d$  and  $\varepsilon^{-1}$ . Moreover, their bounds on  $n$  are far away from the desired order  $O(\varepsilon^{-2}d)$ . (Nevertheless, also this construction solves formally Problem 1(c) from [7].) Our construction for the star-discrepancy gives bounds on  $n$  with nearly optimal order of  $O(\varepsilon^{-2}d(\ln \ln(d) + \ln(1/\varepsilon)))$  in the dimension  $d$ . The running time of our construction is exponential in the dimension  $d$  (i.e.,  $O(C^d d^d \ln(d)^d \varepsilon^{-2d-4})$ , neglecting log-terms).

## An Open Problem

It is a challenging open question whether or not there is a polynomial-time algorithm in  $d$  achieving our bound. This may be not the case unless  $P = NP$ .

In view of the results of Even et al., Chari et al. and our discussion, it would be of great interest to determine the computational complexity of the problem and to exhibit the threshold for  $n$  as a function of  $d$  at which a polynomial-time construction in  $d$  is possible.

## References

- [1] S. Chari, P. Rohatgi, A. Srinivasan, Improved algorithms via approximation of probability distributions, *J. Comput. System Sci* 61 (2000) 81–107.
- [2] M. Drmota, R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer, Berlin, 1997.
- [3] G. Even, O. Goldreich, M. Luby, N. Nisan, B. Veličković, Approximations of general independent distributions, in: *Proceedings of the 24th ACM Symp. on Theory of Computing (STOC)*, 1992, pp. 10–16.
- [4] M. Gnewuch (joint with B. Doerr), Geometric discrepancies and  $\delta$ -covers, *Oberwolfach Reports* 1 (2004) 687–690.
- [5] M. Gnewuch, Bounds for the average  $L^p$ -extreme and the  $L^\infty$ -extreme discrepancy, *Berichtsreihe des Mathematischen Seminars der Universität Kiel*, Report 05-4, 2005.
- [6] D. Haussler, Sphere packing numbers for subsets of the Boolean  $n$ -cube with bounded Vapnik-Chervonenkis dimension, *J. Combin. Theory Ser. A* 69 (1995) 217–232.
- [7] S. Heinrich, Some open problems concerning the star-discrepancy, *J. Complexity* 19 (2003) 416–419.
- [8] S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, *Acta Arith.* 96 (2001) 279–302.
- [9] F. J. Hickernell, I. H. Sloan, G. W. Wasilkowski, On tractability of weighted integration over bounded and unbounded regions in  $\mathbb{R}^s$ , *Math. Comp.* 73 (2004) 1885–1901.
- [10] A. Hinrichs, Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy, *J. Complexity* 20 (2004) 477–483.



- [11] J. Matoušek, *Geometric Discrepancy*, Springer, Berlin, 1999.
- [12] H. N. Mhaskar, On the tractability of multivariate integration and approximation by neural networks, *J. Complexity* 20 (2004) 561–590.
- [13] J. Naor, M. Naor, Small-bias probability spaces: efficient constructions and applications, *SIAM J. Comput.* 22(4) (1993), 838–856.
- [14] H. Niederreiter, *Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [15] D. Pollard, *Convergence of Stochastic Processes*, Springer, Berlin, 1984.
- [16] P. Raghavan, Probabilistic construction of deterministic algorithms: approximating packing integer programs, *J. Comput. System Sci.* 37 (1988) 130–143.
- [17] J. Spencer, *Ten lectures on the probabilistic method*, SIAM, Philadelphia, 1987.
- [18] A. Srivastav, Derandomization in Combinatorial Optimization, in : Pardalos, Rajasekaran, Reif, Rolim (Eds.), *Handbook of Randomized Computing, Volume II*, Kluwer Academic Publishers, Dordrecht, 2001, 731–842.
- [19] A. Srivastav, P. Stangier, Algorithmic Chernoff-Hoeffding inequalities in integer programming, *Random Structures Algorithms* 8 (1996) 27–58.