

Max-Planck-Institut  
für Mathematik  
in den Naturwissenschaften  
Leipzig

Bracketing numbers for axis-parallel boxes and  
applications to geometric discrepancy

by

*Michael Gnewuch*

Preprint no.: 21

2007





# Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy

Michael Gnewuch

Institute of Computer Science, University of Kiel,  
Christian-Albrechts-Platz 4, 24098 Kiel, Germany  
email: mig@informatik.uni-kiel.de

January 8, 2007

## Abstract

In the first part of this paper we derive lower bounds and constructive upper bounds for the bracketing numbers of anchored and unanchored axis-parallel boxes in the  $d$ -dimensional unit cube.

In the second part we apply these results to geometric discrepancy. We derive upper bounds for the inverse of the star and the extreme discrepancy with explicitly given small constants and an optimal dependence on the dimension  $d$ , and provide corresponding bounds for the star and the extreme discrepancy itself. These bounds improve known results from [B. Doerr, M. Gnewuch, A. Srivastav. Bounds and constructions for the star-discrepancy via  $\delta$ -covers. *J. Complexity* 21 (2005), 691-709], [M. Gnewuch. Bounds for the average  $L^p$ -extreme and the  $L^\infty$ -extreme discrepancy, *Electron. J. Combin.* 12 (2005), Research Paper 54] and [H. N. Mhaskar. On the tractibility of multivariate integration and approximation by neural networks. *J. Complexity* 20 (2004), 561-590].

We also discuss an algorithm from [E. Thiémond, An algorithm to compute bounds for the star discrepancy, *J. Complexity* 17 (2001), 850-880] to approximate the star-discrepancy of a given  $n$ -point set. Our lower bound on the bracketing number of anchored boxes, e.g., leads directly to a lower bound of the running time of Thiémond's algorithm. Furthermore, we show how one can use our results to modify the algorithm to approximate the extreme discrepancy of a given set.

## 1 Bracketing numbers for axis-parallel boxes

Let  $d \in \mathbb{N}$ . Let  $L^1([0, 1]^d)$  be the set of real valued Lebesgue integrable functions on the  $d$ -dimensional unit cube  $[0, 1]^d$ , and let  $\mathcal{F}$  be a subset of  $L^1([0, 1]^d)$ . For  $\delta > 0$  and  $f, g \in \mathcal{F}$

we call the set

$$[f, g]_{\mathcal{F}} := \{h \in \mathcal{F} \mid f \leq h \leq g \text{ everywhere on } [0, 1]^d\}$$

a  $\delta$ -bracket of  $\mathcal{F}$  if its weight  $W([f, g]_{\mathcal{F}})$  satisfies

$$W([f, g]_{\mathcal{F}}) := \int_{[0, 1]^d} (g(x) - f(x)) dx \leq \delta.$$

A finite subset  $\Gamma$  of  $\mathcal{F}$  is a (*one-sided*)  $\delta$ -cover of  $\mathcal{F}$  if for every  $h \in \mathcal{F}$  there exists a  $\delta$ -bracket  $[f, g]_{\mathcal{F}}$  with  $f, g \in \Gamma$  and  $h \in [f, g]_{\mathcal{F}}$ . A  $\delta$ -bracketing cover of  $\mathcal{F}$  is a set of  $\delta$ -brackets whose union is  $\mathcal{F}$ .

By  $N_{[\ ]}(\mathcal{F}, \delta)$  we denote the *bracketing number* of  $\mathcal{F}$ , i.e., the smallest number of  $\delta$ -brackets whose union is  $\mathcal{F}$ , and by  $N(\mathcal{F}, \delta)$  we denote the smallest cardinality of all  $\delta$ -covers of  $\mathcal{F}$ . The notion of bracketing is well established in the theory of empirical processes, see, e.g., [14, 16]. The notion of one-sided  $\delta$ -covers was introduced in [11].

It is easy to see that we have the following relation:

$$N(\mathcal{F}, \delta) \leq 2N_{[\ ]}(\mathcal{F}, \delta) \leq N(\mathcal{F}, \delta)(N(\mathcal{F}, \delta) + 1). \quad (1)$$

Let us introduce further helpful notation: Put  $[d] := \{1, \dots, d\}$ . 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 notation for open and half-open intervals. We put  $V_x := \lambda^d([0, x])$  and  $V_{x,y} := \lambda^d([x, y])$ , where  $\lambda^d$  is the  $d$ -dimensional Lebesgue measure. Let us denote the characteristic function of a set  $A \subseteq \mathbb{R}^d$  by  $1_A$ . In this paper we consider the subsets

$$\mathcal{C}_d := \{1_{[0,x]} \mid x \in [0, 1]^d\} \quad \text{and} \quad \mathcal{R}_d := \{1_{[x,y]} \mid x, y \in [0, 1]^d\}$$

of  $L^1([0, 1]^d)$ . The elements of  $\mathcal{C}_d$  are called *anchored (axis-parallel) boxes* or simply *corners*. The elements of  $\mathcal{R}_d$  are called *unanchored (axis-parallel) boxes*. (Here the word “unanchored” is of course meant in the sense of “not necessarily anchored”.) It is easy to see that

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq N_{[\ ]}(\mathcal{R}_d, \delta) \quad \text{and} \quad N(\mathcal{C}_d, \delta) \leq N(\mathcal{R}_d, \delta) \quad (2)$$

holds for all  $\delta > 0$ . Indeed, let  $f = 1_{[x,z]}$ ,  $g = 1_{[x',z']}$ , and let  $[f, g]_{\mathcal{R}_d}$  be a  $\delta$ -bracket of  $\mathcal{R}_d$  with  $[f, g]_{\mathcal{R}_d} \cap \mathcal{C}_d \neq \emptyset$ . This implies  $x' = 0$ , thus  $g \in \mathcal{C}_d$ . Consider  $f_{-x} := 1_{[0,z-x]}$ . Obviously  $[f_{-x}, g]_{\mathcal{C}_d}$  is a  $\delta$ -bracket of  $\mathcal{C}_d$ . Let  $h = 1_{[0,y]} \in [f, g]_{\mathcal{R}_d} \cap \mathcal{C}_d$ . From  $f \leq h$  follows  $f_{-x} \leq h$ , hence  $h \in [f_{-x}, g]_{\mathcal{C}_d}$ . Therefore  $[f, g]_{\mathcal{R}_d} \cap \mathcal{C}_d \subseteq [f_{-x}, g]_{\mathcal{C}_d}$ . This establishes (2).

Let us identify the functions  $1_{[0,x]}$  in  $\mathcal{C}_d$  with the corresponding points  $x \in [0, 1]^d$  and the functions  $1_{[x,y]}$  in  $\mathcal{R}_d$  with the corresponding sets  $[x, y] \subseteq [0, 1]^d$ . According to this convention, we identify the bracket  $[1_{[0,x]}, 1_{[0,y]}]_{\mathcal{C}_d}$  with the  $d$ -dimensional box  $[x, y]$ .

## 1.1 A lower bound for $N_{[\ ]}(\mathcal{C}_d, \delta)$

In [3, Thm. 2.8] the following lower bound for  $N(\mathcal{C}_d, \delta)$  was stated:

$$N(\mathcal{C}_d, \delta) \geq \sqrt{d} e^{-d} \delta^{-d} + O(|\ln(\delta)|^{d-1}).$$

According to (1), multiplying the right hand side by  $1/2$  gives us a lower bound for  $N_{[\cdot]}(\mathcal{C}_d, \delta)$ . Earlier, in [10, Proof of Thm. 2], Hinrichs proved a lower bound for the so-called covering number of  $\mathcal{C}_d$ , resulting in

$$N_{[\cdot]}(\mathcal{C}_d, \delta) \geq \sqrt{\pi d/2} (4e)^{-d} \delta^{-d}$$

(cf. [3, Remark 2.10]). In this section we will derive a better lower bound for  $N_{[\cdot]}(\mathcal{C}_d, \delta)$  with a coefficient in front of the most significant term  $\delta^{-d}$  that is a constant and therefore not exponentially decreasing in the dimension  $d$ .

As we will show in Lemma 1.2, the bracketing number  $N_{[\cdot]}(\mathcal{C}_d, \delta)$  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  $\delta$ -bracket with maximum volume containing  $x$ . Thus for a lower bound of  $N_{[\cdot]}(\mathcal{C}_d, \delta)$  it is helpful to determine first  $\lambda^d(B_\delta(x))$  for each  $x \in [0, 1]^d$ . Let us start with the case where  $V_x \geq \delta$ . We define the set  $U(d, \delta) := \{y \in [0, 1]^d \mid V_y \geq \delta\}$  and the function

$$h_\delta : U(d, \delta) \rightarrow [0, 1], z \mapsto (1 - (1 - \delta/V_z)^{1/d})^d V_z. \quad (3)$$

Expanding  $(1 - \delta/V_z)^{1/d}$  into a power series, we obtain for  $V_z > \delta$

$$h_\delta(z) = d^{-d} \frac{\delta^d}{V_z^{d-1}} \left( 1 + \sum_{k=1}^{\infty} \frac{(k-1/d)\dots(1-1/d)}{(k+1)!} \left( \frac{\delta}{V_z} \right)^k \right)^d, \quad (4)$$

and for  $d \geq 2$  the right hand side of (4) is a strictly decreasing function in  $V_z$ .

**Lemma 1.1.** *Let  $d \geq 2$ ,  $\delta \in (0, 1]$ , and let  $z \in [0, 1]^d$  with  $V_z \geq \delta$ . Put*

$$x = x(z, \delta) := \left( 1 - \frac{\delta}{V_z} \right)^{1/d} z. \quad (5)$$

*Then  $[x, z]$  is the uniquely determined  $\delta$ -bracket having maximum volume of all  $\delta$ -brackets containing  $z$ . Its volume is  $V_{x,z} = h_\delta(z)$ .*

*Proof.* Let us first prove that the point  $x$  in (5) is the uniquely determined maximum of the function

$$g : \{\xi \in [0, z] \mid V_\xi = V_z - \delta\} \rightarrow \mathbb{R}, \xi \mapsto V_{\xi,z}.$$

If  $V_z = \delta$ , then  $x = 0$ , and 0 is obviously the unique point where  $g$  takes its maximum. So let  $V_z > \delta$ . Put  $f : [0, z] \rightarrow [0, \infty)$ ,  $f(\xi) = V_\xi$ , and  $M := \{\xi \mid f(\xi) = V_z - \delta\}$ . Let  $y$  be a local maximum of  $g$  on the compact set  $[0, z] \cap M$ . It is obvious that  $y_i \neq 0$  and  $y_i \neq z_i$  for each  $i \in [d]$ , i.e.,  $y \in (0, z) \cap M$ . Thus  $\text{grad } f(y) = V_y(y_1^{-1}, \dots, y_d^{-1}) \neq 0$ , which implies the existence of a Lagrangian multiplier  $\lambda \in \mathbb{R}$  with  $\text{grad } g(y) = \lambda \text{grad } f(y)$ . Since

$$\text{grad } g(y) = -V_{y,z}((z_1 - y_1)^{-1}, \dots, (z_d - y_d)^{-1}),$$

$y$  and  $z$  have to be necessarily linearly dependend. From this and  $V_y = V_z - \delta$  we obtain  $y = (1 - \delta/V_z)^{1/d} z = x$ . The whole statement of Lemma 1.1 follows now from  $V_{x,z} = (1 - (1 - \delta/V_z)^{1/d})^d V_z$  and the fact that  $V_{x(z),z}$  is a strictly decreasing function with respect to the parameter  $V_z$ , see (4).  $\square$

Let us now consider the case where  $V_x \leq \delta$ . Let us assume that the  $\delta$ -bracket  $B_\delta(x) = [v, w]$  has the maximum volume of all  $\delta$ -brackets containing  $x$ . Then obviously  $V_w \geq \delta$ , and we find an  $u \in B_\delta(x)$  with  $V_u = \delta$  and  $x \in [0, u]$ . Thus  $u$  is contained in the  $\delta$ -brackets  $[0, u]$  and  $B_\delta(x)$ . According to Lemma 1.1  $[0, u]$  has the maximum volume of all  $\delta$ -brackets containing  $u$ , which implies  $\lambda^d([0, u]) \geq \lambda^d(B_\delta(x))$ . From the definition of  $B_\delta(x)$  we get additionally  $\lambda^d(B_\delta(x)) \geq \lambda^d([0, u])$ . Thus

$$\lambda^d(B_\delta(x)) = \delta \quad \text{for all } x \in U(d, \delta)^c,$$

where  $U(d, \delta)^c$  denotes the complement of  $U(d, \delta)$  in  $[0, 1]^d$ .

**Lemma 1.2.** *Let  $d \geq 2$  and  $\delta \in (0, 1]$ . We have*

$$N_{[\cdot]}(\mathcal{C}_d, \delta) \geq \int_{U(d, \delta)} h_\delta(z)^{-1} dz + \delta^{-1} \lambda^d(U(d, \delta)^c).$$

*Proof.* Let  $\mathcal{B}$  be a finite set of  $\delta$ -brackets whose union is  $[0, 1]^d$ . For each  $y \in [0, 1]^d$  choose a bracket  $Q(y) \in \mathcal{B}$  with  $y \in Q(y)$  in such a way that  $C(Q) := \{y \mid Q = Q(y)\}$  is measurable for all  $Q \in \mathcal{B}$ . Clearly  $C(Q) \subseteq Q$  for all  $Q \in \mathcal{B}$  and  $(C(Q))_{Q \in \mathcal{B}}$  forms a partition of  $[0, 1]^d$ . According to Lemma 1.1, we get  $h_\delta(y) \geq \lambda^d(Q(y))$  for any  $y$  with  $V_y \geq \delta$ . Thus

$$\begin{aligned} |\mathcal{B}| &\geq \sum_{Q \in \mathcal{B}} \frac{\lambda^d(C(Q))}{\lambda^d(Q)} = \sum_{Q \in \mathcal{B}} \int_{C(Q)} \lambda^d(Q)^{-1} dy = \int_{[0, 1]^d} \lambda^d(Q(y))^{-1} dy \\ &= \int_{U(d, \delta)} \lambda^d(Q(y))^{-1} dy + \int_{U(d, \delta)^c} \lambda^d(Q(y))^{-1} dy \geq \int_{U(d, \delta)} h_\delta(y)^{-1} dy + \delta^{-1} \lambda^d(U(d, \delta)^c). \end{aligned}$$

□

**Lemma 1.3.** *Let  $f \in L^1([0, 1]^d)$  such that there exists an  $\tilde{f} \in L^1([0, 1])$  with  $f(z) = \tilde{f}(V_z)$  for all  $z \in [0, 1]^d$ . Then we have*

$$\int_{U(d, \delta)} f(z) dz = \frac{1}{(d-1)!} \int_\delta^1 \tilde{f}(\vartheta) \ln(\vartheta^{-1})^{d-1} d\vartheta. \quad (6)$$

*Proof.* Let us consider the transformation

$$\Phi : (0, \infty)^d \rightarrow (0, \infty)^d, \quad (z_1, \dots, z_{d-1}, z_d) \mapsto (z_1, \dots, z_{d-1}, V_z)$$

and its inverse function given by  $\Phi^{-1}(\zeta_1, \dots, \zeta_{d-1}, \vartheta) = (\zeta_1, \dots, \zeta_{d-1}, \vartheta/V_\zeta)$ , where  $V_\zeta = \zeta_1 \dots \zeta_{d-1}$ . The Jacobian determinant of  $\Phi^{-1}$  is  $\det(D\Phi^{-1}(\zeta_1, \dots, \zeta_{d-1}, \vartheta)) = V_\zeta^{-1}$ . Then

$$\begin{aligned} \int_{U(d, \delta)} f(z) dz &= \int_\delta^1 \int_{U(d-1, \vartheta)} (f \circ \Phi^{-1})(\zeta, \vartheta) |\det D\Phi^{-1}(\zeta, \vartheta)| d\zeta d\vartheta \\ &= \int_\delta^1 \left( \int_{U(d-1, \vartheta)} V_\zeta^{-1} d\zeta \right) \tilde{f}(\vartheta) d\vartheta. \end{aligned} \quad (7)$$

If  $d = 2$ , then obviously  $\int_{U(1,\vartheta)} V_\zeta^{-1} d\zeta = \ln(\vartheta^{-1})$ . If  $d \geq 3$ , then (7) implies

$$\int_{U(d-1,\vartheta)} V_\zeta^{-1} d\zeta = \int_\vartheta^1 \left( \int_{U(d-2,\sigma)} V_\eta^{-1} d\eta \right) \sigma^{-1} d\sigma.$$

Hence a simple induction gives us  $\int_{U(d-1,\vartheta)} V_\zeta^{-1} d\zeta = \ln(\vartheta^{-1})^{d-1}/(d-1)!$ .  $\square$

**Remark 1.4.** With the help of Lemma 1.3 one can easily calculate the quantity  $\lambda^d(U(d, \delta)^c)$  appearing in Lemma 1.2. From  $\lambda^d(U(d, \delta)^c) = 1 - \lambda^d(U(d, \delta))$  we get

$$\lambda^d(U(d, \delta)^c) = \delta \sum_{k=0}^{d-1} \frac{\ln(\delta^{-1})^k}{k!}.$$

Notice that for fixed  $\delta$  we have  $\lim_{d \rightarrow \infty} \lambda^d(U(d, \delta)^c) = \delta \exp(\ln(\delta^{-1})) = 1$ .

Let now  $\psi(x) := (x/d)^d (1 - (1-x)^{1/d})^{-d}$ . The function  $\psi$  is obviously holomorphic on  $\{x \in \mathbb{C} \mid |x| < 1\} \setminus \{0\}$ . Expanding  $(1-x)^{1/d}$  into a power series for  $|x| < 1$  we get, similarly as in (4),

$$\psi(x) = \left( 1 + \sum_{k=1}^{\infty} \frac{(k-1/d) \dots (1-1/d)}{(k+1)!} x^k \right)^{-d}, \quad (8)$$

which shows that  $\lim_{x \rightarrow 0} \psi(x) = 1$ . Thus  $\psi$  is holomorphic on  $\{x \in \mathbb{C} \mid |x| < 1\}$  and can there be represented by a convergent power series  $\psi(x) = 1 + \sum_{k=1}^{\infty} \alpha_k x^k$ . If we put  $\psi(1) = d^{-d}$ , then  $\psi$  is also continuous in  $x = 1$ . In particular, we can write  $\psi(x) = 1 + x\bar{\psi}(x)$  for all  $x \in [0, 1]$ , where  $\bar{\psi}$  is a continuous function on  $[0, 1]$ . Then Lemma 1.3 and identity (4) give us

$$\begin{aligned} \int_{U(d,\delta)} h_\delta(z)^{-1} dz &= \delta^{-d} \frac{d^d}{(d-1)!} \int_\delta^1 (\vartheta \ln(\vartheta^{-1}))^{d-1} \psi(\delta/\vartheta) d\vartheta \\ &= \delta^{-d} \frac{d^d}{(d-1)!} \left\{ \int_\delta^1 (\vartheta \ln(\vartheta^{-1}))^{d-1} d\vartheta + \delta \int_\delta^1 \vartheta^{d-2} \ln(\vartheta^{-1})^{d-1} \bar{\psi}(\delta/\vartheta) d\vartheta \right\} \\ &\geq \delta^{-d} \frac{d^d}{(d-1)!} \left\{ \int_\delta^1 (\vartheta \ln(\vartheta^{-1}))^{d-1} d\vartheta - \delta \max_{0 \leq x \leq 1} |\bar{\psi}(x)| \int_\delta^1 \vartheta^{d-2} \ln(\vartheta^{-1})^{d-1} d\vartheta \right\}. \end{aligned}$$

Now

$$\begin{aligned} \int_\delta^1 (\vartheta \ln(\vartheta^{-1}))^{d-1} d\vartheta &= \left[ (d-1)! \left( \frac{\vartheta}{d} \right)^d \sum_{k=0}^{d-1} \frac{d^k \ln(\vartheta^{-1})^k}{k!} \right]_{\vartheta=\delta}^1 \\ &= \frac{(d-1)!}{d^d} \left( 1 - \delta^d \sum_{k=0}^{d-1} \frac{d^k \ln(\delta^{-1})^k}{k!} \right). \end{aligned}$$

Furthermore,

$$\begin{aligned} \int_{\delta}^1 \vartheta^{d-2} \ln(\vartheta^{-1})^{d-1} d\vartheta &= \left[ \frac{\vartheta^{d-1} \ln(\vartheta^{-1})^{d-1}}{d-1} \right]_{\vartheta=\delta}^1 + \int_{\delta}^1 \vartheta^{d-2} \ln(\vartheta^{-1})^{d-2} d\vartheta \\ &= \frac{(d-2)!}{(d-1)^{d-1}} \left( 1 - \delta^{d-1} \sum_{k=0}^{d-1} \frac{(d-1)^k \ln(\delta^{-1})^k}{k!} \right). \end{aligned}$$

Thus we get from Lemma 1.2 the following lower bound for the bracketing number of anchored boxes:

**Theorem 1.5.** *For  $d \geq 2$  and  $\delta \in (0, 1]$  there exists a constant  $c_d$ , which may depend on  $d$ , but not on  $\delta$ , with*

$$N_{[\cdot]}(\mathcal{C}_d, \delta) \geq \delta^{-d} (1 - c_d \delta). \quad (9)$$

**Remark 1.6.** To get a more explicit lower bound than (9) one may expand  $\psi$  into a McLaurin series. For  $|x| < 1$  we have

$$\psi(x) = \sum_{k=0}^{\infty} \frac{\psi^{(k)}(0)}{k!} x^k$$

and  $\psi^{(0)}(0) = \psi(0) = 1$ ,  $\psi^{(1)}(0) = -(d-1)/2$ , and

$$\psi^{(2)}(0) = (d-1) \left( (d+1) \frac{1-1/d}{4} - \frac{2-1/d}{3} \right).$$

However, the larger  $k$  the more the number of summands of the explicit expression of  $\psi^{(k)}$  blows up. Alternatively, one can also get a bound with explicitly given constants by estimating more rigorously. Let  $c := (2 \ln 2)^{-2} > 0.52$ . It is straightforward to establish

$$N_{[\cdot]}(\mathcal{C}_d, \delta) \geq c\delta^{-d} - c d^d \sum_{k=0}^{d-1} \frac{|d \ln(d\delta)|^k}{k!} \quad \text{for all } d \geq 2 \text{ and all } \delta \in (0, 1] \quad (10)$$

by using the estimates

$$h_{\delta}(z) \leq d^{-d} \frac{\delta^d}{V_z^{d-1}} \left( \sum_{k=0}^{\infty} \frac{1}{k+1} \left( \frac{\delta}{V_z} \right)^k \right)^d$$

and

$$\int_{U(d,\delta)} h_{\delta}(z)^{-1} dz \geq \left( \frac{d}{\delta} \right)^d \left( \sum_{k=0}^{\infty} \frac{d^{-k}}{k+1} \right)^{-d} \int_{U(d,\delta)} V_z^{d-1} dz;$$

an elementary analysis proves that the minimum of

$$\left( \sum_{k=0}^{\infty} \frac{d^{-k}}{k+1} \right)^{-d} = (-d \ln(1 - 1/d))^{-d}$$

is taken in  $d = 2$ . Although the leading coefficient  $c$  in (10) is smaller than the leading coefficient 1 in (9), it is still a constant and in particular not decreasing with  $d$ .

## 1.2 A lower bound for $N_{[\cdot]}(\mathcal{R}_d, \delta)$

Due to (2) the right hand side of (9) is also a lower bound of  $N_{[\cdot]}(\mathcal{R}_d, \delta)$ , but one would suspect it to be a rather poor one. Here we derive a better lower bound for  $N_{[\cdot]}(\mathcal{R}_d, \delta)$ .

For our calculations it is convenient to identify  $\mathcal{R}_d$  as a measure space with the set

$$\tilde{\mathcal{R}}_d := \{(x, y) \in [0, 1]^{2d} \mid x \leq y\}$$

endowed with the probability measure  $d\mu^d(x, y) := 2^d dx dy$ . Since the set  $D_d := \{(x, x) \in [0, 1]^{2d}\}$  has  $\mu^d$ -measure zero, it is not important that our map of identification  $\Phi : \tilde{\mathcal{R}}_d \rightarrow \mathcal{R}_d, (x, y) \mapsto [x, y]$  is not injective on  $D_d$ .

As we will state in Lemma 1.8, the bracketing number  $N_{[\cdot]}(\mathcal{R}_d, \delta)$  is bounded from below by the average  $\mu^d(B(x, y, \delta))$  over all  $(x, y) \in \tilde{\mathcal{R}}_d$ , where  $B(x, y, \delta)$  is a  $\delta$ -bracket of  $\mathcal{R}_d$  with maximum  $\mu^d$ -measure containing  $[x, y]$ . Previously we calculate the value of  $\mu^d(B(x, y, \delta))$ . For  $\delta \in (0, 1]$  let us define  $\tilde{U}(d, \delta) := \{(x, y) \in [0, 1]^{2d} \mid x < y, V_{x,y} > \delta\}$  and

$$\tilde{h}_\delta : \tilde{U}(d, \delta) \rightarrow [0, 1], (x, y) \mapsto 2^{-d} \left( 1 - \left( 1 - \frac{\delta}{V_{x,y}} \right)^{1/d} \right)^{2d} V_{x,y}^2.$$

A power series expansion (cf. (4)) leads to

$$\tilde{h}_\delta(x, y) = 2^{-d} d^{-2d} \frac{\delta^{2d}}{V_{x,y}^{2(d-1)}} \left( 1 + \sum_{k=1}^{\infty} \frac{(k-1/d) \dots (1-1/d)}{(k+1)!} \left( \frac{\delta}{V_{x,y}} \right)^k \right)^{2d}. \quad (11)$$

**Lemma 1.7.** *Let  $d \geq 2$ , and let  $x, y \in [0, 1]^d$  with  $V_{x,y} > \delta$ . Put*

$$\xi = \xi(x, y, \delta) := \frac{1}{2} \left( \left( 1 + \left( 1 - \frac{\delta}{V_{x,y}} \right)^{1/d} \right) x + \left( 1 - \left( 1 - \frac{\delta}{V_{x,y}} \right)^{1/d} \right) y \right)$$

and

$$\eta = \eta(x, y, \delta) := \frac{1}{2} \left( \left( 1 + \left( 1 - \frac{\delta}{V_{x,y}} \right)^{1/d} \right) y + \left( 1 - \left( 1 - \frac{\delta}{V_{x,y}} \right)^{1/d} \right) x \right).$$

Then  $B = B(x, y, \delta) := [[\xi, \eta], [x, y]]_{\mathcal{R}_d}$  is the uniquely determined  $\delta$ -bracket having maximum  $\mu^d$ -measure of all  $\delta$ -brackets containing  $[x, y]$ . Its measure is  $\mu^d(B) = \tilde{h}_\delta(x, y)$ .

*Sketch of the proof.* Similar as in Lemma 1.1 one can use the elementary properties of Lagrangian multipliers to prove that the function

$$\tilde{g} : \{(v, w) \mid x \leq v \leq w \leq y, V_{v,w} = V_{x,y} - \delta\} \rightarrow \mathbb{R}, (v, w) \mapsto \mu^d([v, w], [x, y])_{\mathcal{R}_d}$$

takes its uniquely determined maximum in  $(\xi, \eta)$ . By direct calculation one gets  $\mu^d(B) = \tilde{h}_\delta(x, y)$  and  $\tilde{h}_\delta(x, y)$  is obviously a strictly decreasing function in  $V_{x,y}$ , see (11). From this follows easily  $\mu^d(B) > \mu^d(B')$  for all  $\delta$ -brackets  $B' \neq B$  containing  $[x, y]$ .  $\square$

Let now  $x, y \in [0, 1]^d$  with  $x \leq y$  and  $V_{x,y} \leq \delta$ . Let  $u, v \in [0, 1]^d$  with  $u \leq x \leq y \leq v$  and  $V_{u,v} = \delta$ . Then  $[\emptyset, [u, v]]_{\mathcal{R}_d}$  has maximum  $\mu^d$ -measure of all  $\delta$ -brackets containing  $[x, y]$ , namely  $\mu^d([\emptyset, [u, v]]_{\mathcal{R}_d}) = V_{u,v}^2 = \delta^2$ . Notice that this measure differs from the value

$$\lim_{\delta < V_{x,y} \rightarrow \delta} \mu^d(B(x, y, \delta)) = 2^{-d} \delta^2.$$

Similar as in the proof of Lemma 1.2 one can verify the next lemma.

**Lemma 1.8.** *Let  $d \geq 2$  and  $\delta \in (0, 1]$ . Then we have*

$$N_{[\cdot]}(\mathcal{R}_d, \delta) \geq \int_{\tilde{U}(d, \delta)} \tilde{h}_\delta(x, y)^{-1} d\mu^d(x, y) + \delta^{-2} \mu^d(\tilde{U}(d, \delta)^c),$$

where  $\tilde{U}(d, \delta)^c$  denotes the complement of  $\tilde{U}(d, \delta)$  in  $\tilde{\mathcal{R}}_d$ .

**Lemma 1.9.** *For all  $\delta \in (0, 1)$  and all  $l \in \mathbb{N}_0$  we have*

$$\int_{\tilde{U}(d, \delta)} V_{x,y}^l d\mu^d(x, y) = \frac{2^d}{(l+1)^d (l+2)^d} + 2^d \delta^{l+1} \sum_{k=0}^{d-1} (\alpha_{d,k}^{(l)} + \beta_{d,k}^{(l)} \delta) \frac{(\ln \delta)^k}{k!},$$

where the constants  $\alpha_{d,k}^{(l)}, \beta_{d,k}^{(l)}$ ,  $k = 1, \dots, d-1$ , do not depend on  $\delta$ .

*Sketch of the proof.* We put

$$I(d, l, \delta) := 2^{-d} \int_{\tilde{U}(d, \delta)} V_{x,y}^l d\mu^d(x, y).$$

We have

$$I(1, l, \delta) = \int_0^{1-\delta} \int_{x+\delta}^1 (y-x)^l dy dx = \frac{1}{(l+1)(l+2)} - \frac{1}{l+1} \delta^{l+1} + \frac{1}{l+2} \delta^{l+2}$$

and

$$I(d+1, l, \delta) = \int_0^{1-\delta} \int_{x+\delta}^1 (y-x)^l I\left(d, l, \frac{\delta}{y-x}\right) dy dx.$$

Therefore it is straightforward to verify by induction that

$$I(d, l, \delta) = \frac{1}{(l+1)^d (l+2)^d} + \delta^{l+1} \sum_{k=0}^{d-1} (\alpha_{d,k}^{(l)} + \beta_{d,k}^{(l)} \delta) \frac{(\ln \delta)^k}{k!},$$

with coefficients not depending on  $\delta$ . More precisely, we have for an arbitrary  $l \in \mathbb{N}$

$$\begin{aligned} \alpha_{1,0}^{(l)} &= -\frac{1}{l+1}, & \beta_{1,0}^{(l)} &= \frac{1}{l+2}, \\ \alpha_{d+1,0}^{(l)} &= -\frac{1}{(l+1)^{d+1} (l+2)^d} + \sum_{\nu=0}^{d-1} \left( -\alpha_{d,\nu}^{(l)} + (-1)^\nu \beta_{d,\nu}^{(l)} \right), \\ \beta_{d+1,0}^{(l)} &= \frac{1}{(l+1)^d (l+2)^{d+1}} - \sum_{\nu=0}^{d-1} \left( -\alpha_{d,\nu}^{(l)} + (-1)^\nu \beta_{d,\nu}^{(l)} \right), \\ \alpha_{d+1,k}^{(l)} &= -\sum_{\nu=k-1}^{d-1} \alpha_{d,\nu}^{(l)}, & \beta_{d+1,k}^{(l)} &= -\sum_{\nu=k-1}^{d-1} (-1)^{\nu+k} \beta_{d,\nu}^{(l)}, \end{aligned}$$

for  $k = 1, \dots, d$ . □

With the help of Lemma 1.9 we can now estimate the integral  $\int_{\tilde{U}(d,\delta)} \tilde{h}_\delta(x, y)^{-1} d\mu^d(x, y)$ . Let  $\psi$  be the function from (8). Then there exists a continuous function  $\tilde{\psi}$  with  $\psi(x)^2 = 1 + x\tilde{\psi}$  for all  $x \in [0, 1]$ . Thus we have

$$\begin{aligned} \int_{\tilde{U}(d,\delta)} \tilde{h}_\delta(x, y)^{-1} d\mu^d(x, y) &= 2^d d^{2d} \delta^{-2d} \int_{\tilde{U}(d,\delta)} V_{x,y}^{2(d-1)} \psi(\delta/V_{x,y})^2 d\mu^d(x, y) \\ &= 2^d d^{2d} \delta^{-2d} \left\{ \int_{\tilde{U}(d,\delta)} V_{x,y}^{2(d-1)} d\mu^d(x, y) + \delta \int_{\tilde{U}(d,\delta)} V_{x,y}^{2d-3} \tilde{\psi}(\delta/V_{x,y}) d\mu^d(x, y) \right\} \\ &\geq 2^d d^{2d} \delta^{-2d} \left\{ \int_{\tilde{U}(d,\delta)} V_{x,y}^{2(d-1)} d\mu^d(x, y) - \delta \max_{0 \leq x \leq 1} |\tilde{\psi}(x)| \int_{\tilde{U}(d,\delta)} V_{x,y}^{2d-3} d\mu^d(x, y) \right\}. \end{aligned}$$

From Lemma 1.8 and Lemma 1.9 we now get the following theorem:

**Theorem 1.10.** *Let  $d \geq 2$  and  $\delta \in (0, 1)$ . There exists a constant  $c_d$ , not depending on  $\delta$ , with*

$$N_{[\ ]}(\mathcal{R}_d, \delta) \geq \delta^{-2d} \left( \left(1 - \frac{1}{2d}\right)^{-d} - c_d \delta \right).$$

We have  $16/9 \geq (1 - 1/(2d))^{-d} \geq \sqrt{e}$  and  $\lim_{d \rightarrow \infty} (1 - 1/(2d))^{-d} = \sqrt{e}$ .

**Remark 1.11.** One can prove the following analog of Lemma 1.3: Let  $g \in L^1([0, 1]^{2d})$  such that there exists a  $\tilde{g} \in L^1([0, 1])$  with  $g(x, y) = \tilde{g}(V_{x,y})$  for all  $(x, y) \in \tilde{R}_d$ . Then

$$\int_{\tilde{U}(d,\delta)} g(x, y) d\mu^d(x, y) = \int_{\delta}^1 \tilde{g}(\vartheta) \varphi_d(\vartheta) d\vartheta$$

for all  $\delta \in (0, 1)$ , where

$$\varphi_d(\vartheta) = (-1)^{d-1} 2^d \sum_{j=0}^{d-1} \left( \binom{2(d-1)-j}{d-1} ((-1)^j - \vartheta) \frac{\ln(\vartheta^{-1})^j}{j!} \right).$$

Since the density function  $\varphi_d$  is rather complicated, it seemed to us more convenient to use Lemma 1.9 to prove Theorem 1.10.

### 1.3 An upper bound for $N_{[\ ]}(\mathcal{C}_d, \delta)$

In [3, Thm. 2.7] an upper bound for  $N(\mathcal{C}_d, \delta)$  was derived implying

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \frac{d^d}{d!} \left( \delta^{-1} + \frac{d+1}{4} \right)^d. \quad (12)$$

The major drawback of this bound, in particular with regard to the applications we have in mind, see Section 2, is its super-exponential dependence on  $d$ . Here we give a modification of the construction in the proof of [3, Thm. 2.7], which is more sophisticated.

Before we start to prove the new upper bound for  $N_{[\ ]}(\mathcal{C}_d, \delta)$ , we introduce further useful notation.

**Definition 1.12.** Let  $\delta \in (0, 1]$  and  $S$  be a subset of  $[0, 1]^d$ . We define  $N_{[\ ]}(S, \delta)$  to be the smallest number of brackets  $[x, z]$ ,  $x, z \in [0, 1]^d$ , whose union contains  $S$ . Let  $a(d, \delta) := (1 - \delta)^{1/d}$  and  $\delta(d, k) := 1 - (1 - \delta)^{(d-k)/d}$ . We shall apply the shorthands  $\delta^{(k)} := \delta(d, k)$  and  $\delta' := \delta^{(1)}$ . Furthermore, let  $S^d([a, b]) := [0, b]^d \setminus [0, a]^d$  for  $0 \leq a \leq b \leq 1$ .

Some elementary observations are listed in the following lemma.

**Lemma 1.13.** *Let  $d \in \mathbb{N}$  and  $\delta \in (0, 1]$ .*

1. *If  $S, T \subseteq [0, 1]^d$ , then  $N_{[\ ]}(S \cup T, \delta) \leq N_{[\ ]}(S, \delta) + N_{[\ ]}(T, \delta)$ .*
2. *If  $\lambda > 0$  and  $S, \lambda S \subseteq [0, 1]^d$ , then  $N_{[\ ]}(S, \delta) = N_{[\ ]}(\lambda S, \lambda^d \delta)$ .*

The lemma below is a generalization of [3, Lemma 2.6].

**Lemma 1.14.** *Let  $d \geq 2$ ,  $\delta \in (0, 1]$  and  $k \in [d - 1]$ . For every subset  $S \subseteq [0, 1]^{d-k}$  we have*

$$N_{[\ ]}(S \times [a(d, \delta), 1]^k, \delta) \leq N_{[\ ]}(S, \delta^{(k)}). \quad (13)$$

*Proof.* Let first  $k = 1$  and put  $a = a(d, \delta)$ . Let  $x', z' \in [0, 1]^{d-1}$  such that  $[x', z']$  is a  $\delta'$ -bracket in dimension  $d - 1$ . Define the  $d$ -dimensional vectors  $x = (x', a)$ ,  $z = (z', 1)$ . Then  $x \leq z$ . If  $V_{x'} \geq 1 - \delta'$ , then  $V_z - V_x = V_{z'} - aV_{x'} \leq 1 - a(1 - \delta') = \delta$ . 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  $\delta$ -bracket in dimension  $d$  and  $[x, z] \cap (S \times [a, 1]) = ([x', z'] \cap S) \times [a, 1]$ . Therefore  $N_{[\ ]}(S \times [a, 1], \delta) \leq N_{[\ ]}(S, \delta')$ .

Suppose now, we have already shown (13) for a fixed  $k \in [d - 1]$  and any  $S \subseteq [0, 1]^{d-k}$ . Then, since  $a(d, \delta) = a(d - k, \delta^{(k)})$ , we get for  $S' \subseteq [0, 1]^{d-k-1}$

$$\begin{aligned} N_{[\ ]}(S' \times [a(d, \delta), 1]^{k+1}, \delta) &= N_{[\ ]}(S' \times [a(d, \delta), 1], \delta^{(k)}) \\ &= N_{[\ ]}(S' \times [a(d - k, \delta^{(k)}), 1], \delta^{(k)}) \leq N_{[\ ]}(S', (\delta^{(k)})'). \end{aligned}$$

Note that  $(\delta^{(k)})' = \delta^{(k)}(d - k, 1)$ , since the set  $S' \times [a(d - k, \delta^{(k)}), 1]$  is actually a  $(d - k)$ -dimensional one. From  $(\delta^{(k)})' = \delta^{(k+1)}$  we get  $N_{[\ ]}(S' \times [a(d, \delta), 1]^{k+1}, \delta) \leq N_{[\ ]}(S', \delta^{(k+1)})$ .  $\square$

**Theorem 1.15.** *Let  $d \in \mathbb{N}$  and  $0 < \delta \leq 1$ . Then*

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq 2^{d-1} \frac{d^d}{d!} (\delta^{-1} + 1)^d. \quad (14)$$

*Proof.* We put  $n := \lceil \delta^{-1} \rceil$  and proceed by induction. If  $d = 1$ , then we have obviously  $N_{[\ ]}(\mathcal{C}_d, \delta) \leq n \leq \delta^{-1} + 1$ .

Consider now  $d \geq 2$ . Define  $a_i := (1 - i\delta)^{1/d}$  for  $i = 0, \dots, n-1$  and  $a_n := 0$ . Furthermore, let  $\delta_i := \delta/a_{i-1}^d$  for all  $i \in [n-1]$ . From  $a_i/a_{i-1} = a(d, \delta_i)$  and Lemma 1.13 we get

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \sum_{i=1}^n N_{[\ ]}(S^d([a_i, a_{i-1}]), \delta) \leq \sum_{i=1}^{n-1} N_{[\ ]}(S^d([a(d, \delta_i), 1]), \delta_i) + 1.$$

For  $i \in [n-1]$  we can subdivide  $S^d([a(d, \delta_i), 1])$  into sets that are, after suitable permutations of coordinates, of the form  $[0, a(d, \delta_i)]^{d-k} \times [a(d, \delta_i), 1]^k$  for  $k = 1, \dots, d$ . Since bracketing numbers are obviously invariant under permutations of coordinates, Lemma 1.13 and 1.14 ensure

$$\begin{aligned} N_{[\ ]}(S^d([a(d, \delta_i), 1]), \delta_i) &\leq \sum_{k=1}^d \binom{d}{k} N_{[\ ]}([0, a(d, \delta_i)]^{d-k} \times [a(d, \delta_i), 1]^k, \delta_i) \\ &\leq \sum_{k=1}^{d-1} \binom{d}{k} N_{[\ ]}\left(\mathcal{C}_{d-k}, \frac{\delta_i^{(k)}}{1 - \delta_i^{(k)}}\right) + 1. \end{aligned}$$

Our induction hypothesis, a change in the order of summation, and the inequality  $\delta_i^{(k)} \geq \frac{d-k}{d}\delta_i$  lead to

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \sum_{k=1}^{d-1} \binom{d}{k} 2^{d-k-1} \frac{d^{d-k}}{(d-k)!} \left( \sum_{i=1}^{n-1} (\delta^{-1} - i + 1)^{d-k} \right) + \delta^{-1} + 1.$$

Due to  $\gamma^{d-k} \leq ((\gamma + \tau)^{d-k} + (\gamma - \tau)^{d-k})/2$  for all  $\gamma \geq 0$  and  $\tau \in [-1/2, 1/2]$ , we get

$$\sum_{i=1}^{n-1} (\delta^{-1} - i + 1)^{d-k} \leq \int_{1/2}^{n-1/2} (\delta^{-1} - x + 1)^{d-k} dx \leq \frac{(\delta^{-1} + 1/2)^{d-k+1}}{d-k+1},$$

implying

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \sum_{k=0}^{d-2} \binom{d}{k+1} 2^{d-k-2} \frac{d^{d-k-1}}{(d-k)!} (\delta^{-1} + 1/2)^{d-k} + \delta^{-1} + 1.$$

If we compare the coefficients in the last term and in

$$2^{d-1} \frac{d^d}{d!} (\delta^{-1} + 1)^d = \sum_{k=0}^d \binom{d}{k} 2^{d-k-1} \frac{d^d}{d!} (\delta^{-1} + 1/2)^{d-k},$$

it becomes evident that (14) holds. □

**Remark 1.16.** The comparison of coefficients at the end of the proof of Theorem 1.15 indicates that our estimate (14) can to some extent be improved. If we are, e.g. just interested in the coefficient of the highest power of  $\delta^{-1}$ , then it is easy to see that  $N_{[\ ]}(\mathcal{C}_d, \delta) \leq \frac{d^d}{d!} \delta^{-d} + O(\delta^{-d+1})$  (cf. also [3, Theorem 2.7]). Furthermore, one can show

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \alpha(d) \frac{d^d}{d!} (\delta^{-1} + 1)^d \quad \text{with } \alpha(d) < 2^{d-1} \text{ for all } d \geq 2. \quad (15)$$

Nevertheless, for the subsequent investigation we would like to have an estimate of the form

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq C^d \frac{d^d}{d!} (\delta^{-1} + 1)^d, \quad C > 0 \text{ a constant independent of } d. \quad (16)$$

If we start with the induction hypothesis (15), our proof approach gives us

$$N_{[\ ]}(\mathcal{C}_d, \delta) \leq \sum_{k=0}^{d-2} \binom{d}{k+1} \alpha(d-k-1) \frac{d^{d-k-1}}{(d-k)!} (\delta^{-1} + 1/2)^{d-k} + \delta^{-1} + 1.$$

A comparison with the coefficients of

$$\alpha(d) \frac{d^d}{d!} (\delta^{-1} + 1)^d = \sum_{k=0}^d \binom{d}{k} \alpha(d) \frac{d^d (\delta^{-1} + 1/2)^{d-k}}{2^k}$$

shows that  $\alpha(d)$  has to satisfy conditions of the form

$$\frac{2^k}{k+1} \frac{(d-1) \dots (d-k)}{d^k} \alpha(d-k-1) \leq \alpha(d)$$

for  $k = 0, \dots, d-2$ . Hence we find for every given  $\varepsilon > 0$  integers  $d_0$  and  $k$  (where  $d_0 \gg k$ ) such that for all  $d \geq d_0$

$$\left( \frac{\alpha(d)}{\alpha(d-k-1)} \right)^{1/k+1} \geq 2 - \varepsilon.$$

This makes clear that, without reasonable modifications, our way of proof could not lead to an estimate of the form (16) with a constant  $C < 2$ .

## 1.4 An upper bound for $N_{[\ ]}(\mathcal{R}_d, \delta)$

In this subsection we provide two lemmas, which demonstrate how one can use  $\delta/2$ -covers and bracketing covers of  $\mathcal{C}_d$  to construct  $\delta$ -covers and bracketing covers of  $\mathcal{R}_d$  respectively. Let  $\mathbf{1}$  be the  $d$ -dimensional vector  $(1, \dots, 1)$ .

**Lemma 1.17.** *Let  $\delta > 0$ . If  $\Gamma \subseteq \mathcal{C}_d$  is a  $\delta/2$ -cover of  $\mathcal{C}_d$ , then*

$$\tilde{\Gamma} = \{[x, z] \mid \mathbf{1} - x, z \in \Gamma\}$$

*is a  $\delta$ -cover of  $\mathcal{R}_d$ . In particular,  $N(\mathcal{R}_d, \delta) \leq N(\mathcal{C}_d, \delta/2)^2$ .*

*Proof.* Let  $x, z \in [0, 1]^d$ . As  $\Gamma$  is a  $\delta/2$ -cover of  $\mathcal{C}_d$ , we find points  $x', x''$  and  $\underline{z}, \bar{z} \in \Gamma$  with  $x' \leq \mathbf{1} - x \leq x'', \underline{z} \leq z \leq \bar{z}$  and  $V_{x''} - V_{x'} \leq \delta/2$  and  $V_{\bar{z}} - V_{\underline{z}} \leq \delta/2$ . Defining  $\underline{x} := \mathbf{1} - x''$  and  $\bar{x} := \mathbf{1} - x'$  gives us  $\underline{x} \leq x \leq \bar{x}$  and  $V_{\underline{x}, \mathbf{1}} - V_{\bar{x}, \mathbf{1}} \leq \delta/2$ . Hence, since  $\bar{z} \leq \mathbf{1}$  and  $0 \leq \underline{x}$ ,

$$V_{\underline{x}, \bar{z}} - V_{\bar{x}, \underline{z}} = V_{\underline{x}, \bar{z}} - V_{\bar{x}, \bar{z}} + V_{\bar{x}, \bar{z}} - V_{\bar{x}, \underline{z}} \leq V_{\underline{x}, \mathbf{1}} - V_{\bar{x}, \mathbf{1}} + V_{\bar{z}} - V_{\underline{z}} \leq \delta.$$

Since  $[\underline{x}, \bar{z}], [\bar{x}, \underline{z}] \in \tilde{\Gamma}$  and  $[x, z] \in [[\bar{x}, \underline{z}], [\underline{x}, \bar{z}]]_{\mathcal{R}_d}$ ,  $\tilde{\Gamma}$  is a  $\delta$ -cover of  $\mathcal{R}_d$ .  $\square$

It is straightforward to prove a corresponding lemma in terms of bracketing:

**Lemma 1.18.** *Let  $\delta > 0$ , and let  $x^{(1)}, \dots, x^{(n)}, z^{(1)}, \dots, z^{(n)} \in [0, 1]^d$  such that  $[x^{(1)}, z^{(1)}], \dots, [x^{(n)}, z^{(n)}]$  form a  $\delta/2$ -bracketing cover of  $\mathcal{C}_d$ . Then the non-empty sets of*

$$[[\mathbf{1} - x^{(i)}, x^{(j)}], [\mathbf{1} - z^{(i)}, z^{(j)}]]_{\mathcal{R}_d}, \quad i, j \in [n], \quad (17)$$

*form a  $\delta$ -bracketing cover of  $\mathcal{R}_d$ . In particular,  $N_{[\cdot]}(\mathcal{R}_d, \delta) \leq N_{[\cdot]}(\mathcal{C}_d, \delta/2)^2$ .*

## 2 Applications to geometric discrepancy

Let  $P$  be an  $n$ -point set in  $[0, 1]^d$ . We define the *star discrepancy* of  $P$  by

$$d_{\infty}^*(P) = \sup_{C \in \mathcal{C}_d} \left| \lambda^d(C) - \frac{1}{n} |P \cap C| \right|,$$

where  $|P \cap C|$  denotes the cardinality of the finite set  $P \cap C$ . The *extreme* (or *unanchored*) *discrepancy* of  $P$  is given by

$$d_{\infty}^e(P) = \sup_{C \in \mathcal{R}_d} \left| \lambda^d(C) - \frac{1}{n} |P \cap C| \right|.$$

The smallest possible star discrepancy of any  $n$ -point configuration in  $[0, 1]^d$  is

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

The *inverse of the star discrepancy* is given by

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

Similarly, we define the smallest possible extreme discrepancy of any  $n$ -point set  $d_{\infty}^e(n, d)$  and its inverse  $n_{\infty}^e(\varepsilon, d)$ .

It is well known that discrepancy is related to the error of multivariate numerical integration of certain function classes (see, e.g., [4, 9, 12, 13]). For this application it is of interest to calculate the discrepancy of a given  $n$ -point set efficiently up to some admissible error  $\delta$ . Furthermore, it is desirable to have useful bounds for the smallest possible discrepancy of any  $n$ -point set and to be able to construct point sets of moderate size (in particular, non-exponentially in  $d$ ) satisfying these bounds. The classical upper bounds of the form  $d_{\infty}^*(n, d) \leq C_d \ln(n)^{d-1} n^{-1}$  (which also hold for the extreme discrepancy) are not very useful for high-dimensional integration, since the constant  $C_d$  depends crucially on  $d$ , and  $\ln(n)^{d-1} n^{-1}$  is an increasing function in  $n$  for  $n \leq e^{d-1}$ . In particular, we cannot use the classical bounds to get helpful information about the discrepancy of point sets of moderate size.

## 2.1 Low-discrepancy sets of moderate size

In [8] Heinrich, Novak, Wasilkowski, and Woźniakowski proved the bounds

$$d_{\infty}^*(n, d) \leq C\sqrt{\frac{d}{n}} \quad \text{and} \quad n_{\infty}^*(\varepsilon, d) \leq \lceil C^2 d \varepsilon^{-2} \rceil, \quad (18)$$

where  $C$  is a universal constant. The proof uses a theorem of Talagrand on empirical processes [14, Thm. 6.6] combined with a celebrated upper bound of Haussler on so-called covering numbers of Vapnik-Červonenkis classes [7]. Since the theorem of Talagrand holds not only under a condition on the covering number of the set system  $\mathcal{S}$  under consideration, but also under the alternative condition that the  $\delta$ -bracketing number of  $\mathcal{S}$  is bounded from above by  $(C\delta^{-1})^d$ ,  $C$  some constant [14, Thm. 1.1], we can reprove (18) by using our bracketing result Theorem 1.15 instead of the result of Haussler.

An advantage of (18) is that the dependence of the inverse of the discrepancy on  $d$  is optimal. This was verified in [8] by a lower bound for the inverse, which was improved by Hinrichs [10] to  $n_{\infty}^*(d, \varepsilon) \geq c_0 d \varepsilon^{-1}$ . A disadvantage of (18) is that so far no good estimate for the constant  $C$  has been published<sup>1</sup>.

Bracketing numbers for axis-parallel boxes can also be used to derive different bounds for the discrepancy and its inverse with an optimal behavior in the dimension  $d$  and explicitly given, small constants. The idea is to “discretize” the discrepancy and use the following approximation property:

Let  $\Gamma$  be a  $\delta$ -cover of  $\mathcal{C}_d$ . Then for all finite subsets  $P$  of  $[0, 1]^d$  we have

$$d_{\infty}^*(P) \leq \max_{x \in \Gamma} \left| V_x - \frac{|P \cap [0, x]|}{|P|} \right| + \delta. \quad (19)$$

(See, e.g., [3, Lemma 3.1].) An analogous bound holds for the extreme discrepancy of  $P$  and arbitrary  $\delta$ -covers of  $\mathcal{R}_d$ . The discretization of the set of test boxes can be employed in the following probabilistic approach (see, e.g., [3, Thm. 3.2]): Let  $\tau_1, \dots, \tau_n$  be uniformly distributed, independent random variables in  $[0, 1]^d$ . Due to (19) we have

$$\Pr\{d_{\infty}^*(\tau_1, \dots, \tau_n) \leq 2\delta\} > 0 \quad \text{if} \quad \Pr\{d_{\Gamma}^*(\tau_1, \dots, \tau_n) \leq \delta\} > 0. \quad (20)$$

Since  $\Gamma$  is a finite set, we can successfully use the large deviation bound known as Hoeffding’s inequality to deduce that the last inequality in (20) is satisfied if  $2\delta^2 n > \ln |\Gamma| + \ln 2$  holds. Let now  $\Gamma$  be a  $\delta$ -cover of minimal cardinality. Then (1) and (14) imply the following theorem:

**Theorem 2.1.** *Let  $d \geq 2$  and  $\varepsilon \in (0, 1]$ . Then*

$$n_{\infty}^*(\varepsilon, d) \leq \left\lceil 2\varepsilon^{-2} \left( d \ln \left( \frac{6e}{\varepsilon} \right) + \ln(2) \right) \right\rceil. \quad (21)$$

---

<sup>1</sup>A. Hinrichs presented at the Dagstuhl Seminar 04401 “Algorithms and Complexity for Continuous Problems” in 2004 a more direct approach to prove (18) with  $C = 10$ . He estimated the expected discrepancy of random points with the help of Dudley’s metric entropy bound and Haussler’s result on packing numbers of VC-classes. Since there exist versions of Dudley’s bound in terms of bracketing, one can derive similar bounds with the help of Thm. 1.15.

If  $n \geq 2d \ln(6e)$ , we have

$$d_{\infty}^*(n, d) \leq n^{-1/2} \left( d \ln \left( C \frac{n}{d} \right) + 2 \ln(2) \right)^{1/2}, \quad (22)$$

where  $C \leq 18e^2 / \ln(6e)$ .

In [3, Thm. 3.2(i)] we proved almost the same bound for the inverse of the star discrepancy, but under the additional constraint  $\varepsilon \leq 8/(d+1)$ . (This is due to the fact that we used the somehow unpractical bound (12) there.) In this respect, (21) improves upon [3, Thm. 3.2(i)]<sup>2</sup>. Observe that our bound (22) is asymptotically better than the bounds for the star discrepancy in [3, Thm. 3.2]. Recall that the inverse of the star discrepancy (as well as the inverse of the extreme discrepancy) depends linearly on the dimension  $d$ . Thus the practically most relevant choice of  $n$  seems to be  $n$  proportional to  $d$ . In this case (22) behaves asymptotically as the bound for the star discrepancy in (18).

Using the approximation property with respect to the set system  $\mathcal{R}_d$ , the upper bound (14) and Lemma 1.17, one can easily modify the probabilistic approach described above to derive similar results for the extreme discrepancy:

**Theorem 2.2.** *Let  $d \geq 2$  and  $\varepsilon \in (0, 1]$ . Then*

$$n_{\infty}^e(\varepsilon, d) \leq \left\lceil 2\varepsilon^{-2} \left( 2d \ln \left( \frac{10e}{\varepsilon} \right) + \ln(2) \right) \right\rceil. \quad (23)$$

If  $n \geq 4d \ln(10e)$ , we have

$$d_{\infty}^e(n, d) \leq \sqrt{2} n^{-1/2} \left( d \ln \left( C \frac{n}{d} \right) + \ln(2) \right)^{1/2}, \quad (24)$$

where  $C \leq 25e^2 / \ln(10e)$ .

The upper bound (23) for the inverse of the extreme discrepancy improves upon the estimate

$$n_{\infty}^e(\varepsilon, d) \leq 9 \cdot 2^{5(1+1/2k)} k^{1-1/k} d \varepsilon^{-2-1/k} \quad \text{for all } k \in \mathbb{N},$$

which was proved in [5] with the help of upper bounds for the average  $L^p$ -extreme discrepancy,  $2 \leq p < \infty$ .

The upper bound (24) for the extreme discrepancy improves upon the bound given by Mhaskar [11]. Theorem 3.1(a) in [11] is a quite general result on the discrepancy of set systems of axis-parallel boxes (or “cells”) in  $\mathbb{R}^d$ . For the sake of explicit constants it was derived by a probabilistic approach similar to the one described above (cf. also [3, Remark 3.5]). Theorem 3.1(a) provides the following bound for the extreme discrepancy (we have to choose the parameters  $R = 1/2 = R_1$  and  $\mu$  as the Lebesgue measure on  $[-1/2, 1/2]^d$  there, resulting in  $M = 1 = \gamma$ ):

---

<sup>2</sup>Notice also that bound (21) is better than bound (18) with, e.g.,  $C = 10$  for all  $\varepsilon \geq e^{-47}$ , i.e., for all practically interesting values of  $\varepsilon$ .

Let  $G = 4/(3 \ln 3 - 2) \simeq 3.0868$  and  $B(d) = \ln(2^{2d^2+3d+1} d^{2d})$ . If  $n \geq GB(d)$ , then

$$d_\infty^\varepsilon(n, d) \leq 2\sqrt{G} n^{-1/2} \left( B(d) + d \ln \left( \frac{n}{GB(d)} \right) \right)^{1/2}. \quad (25)$$

Apart from the constants this bound is not as good as (24), since  $B(d) \geq \Omega(d^2)$ . In particular, the bound (25) is not applicable (for large  $d$ ) if  $n$  depends linearly on  $d$ .

A comparison of Theorem 2.1 and Theorem 2.2 shows that one get the bounds (23) and (24) for the extreme discrepancy from the corresponding bounds of the star discrepancy more or less by replacing  $d$  with  $2d$  (cf. also [6]). Similar transference results hold for bounds based on the average  $L^p$ -discrepancy [5] and for the bounds in (18) (see [8]). In fact (18) holds for (almost) arbitrary systems  $\mathcal{S}$  of measurable subsets of  $[0, 1]^d$ —one just has to replace  $d$  by the Vapnik-Červonenkis (VC) dimension of  $\mathcal{S}$ . It is easy to see that the VC dimension of  $\mathcal{C}_d$  is  $d$ , and the VC dimension of  $\mathcal{R}_d$  is  $2d$ .

The probabilistic approach via  $\delta$ -covers and Hoeffding's inequality described above has the advantage that it is known how to derandomize it to construct small samples satisfying bounds like (21), (22), (23), and (24). Such a derandomized algorithm for the star discrepancy was provided in [3]; it is essentially a point-by-point construction using the method of conditional probabilities and so-called pessimistic estimators. Unfortunately, it is not trivial to implement and the proven upper bound for the worst case running time is exponential in  $d$ . B. Doerr and the author found a different approach [2] based on special  $\delta$ -covers (or bracketing covers) and on recent results on generating randomized roundings with cardinality constraints [1]. Compared with the algorithm in [3] the new algorithm is easier to implement and has a reasonably better worst case running time. Nevertheless it is still exponential in  $d$ . (This is maybe not too surprising, since all the known deterministic algorithms for the seemingly easier problem of approximating the star discrepancy of arbitrary given point sets have a running time exponential in  $d$ —see also the discussion in the next subsection.)

Although the new algorithm seems to be a step into the right direction, further improvements are desirable (see also the discussion in [3]).

## 2.2 Approximating the discrepancy of a given set

In [15] Thiéard described and tested an algorithm that calculates in moderate dimension  $d$  for a given  $n$ -point set its star discrepancy up to an admissible error  $\delta$ . The algorithm uses the following idea:

Let  $\mathcal{B} = \mathcal{B}_\delta$  be a  $\delta$ -bracketing cover of  $[0, 1]^d$ . If we define for each  $B = [x, y] \in \mathcal{B}$  half-open boxes  $B^- := [0, x)$  and  $B^+ := [0, y)$ , and furthermore for each finite set  $P \subset [0, 1]^d$  with  $|P| = n$

$$B(\mathcal{B}, P) := \max_{B \in \mathcal{B}} \max \left\{ \frac{|P \cap B^+|}{n} - \lambda^d(B^-), \lambda^d(B^+) - \frac{|P \cap B^-|}{n} \right\},$$

and

$$C(\mathcal{B}, P) := \max_{B \in \mathcal{B}} \max \left\{ \left| \frac{|P \cap B^-|}{n} - \lambda^d(B^-) \right|, \left| \frac{|P \cap B^+|}{n} - \lambda^d(B^+) \right| \right\},$$

then it is easy to see that

$$C(\mathcal{B}, P) \leq d_\infty^*(P) \leq B(\mathcal{B}, P) \quad \text{and} \quad B(\mathcal{B}, P) - C(\mathcal{B}, P) \leq \delta.$$

(This is more or less a reformulation of the approximation property (19).) For a given  $n$ -point set  $P$ , and a given admissible error  $\delta$ , Thiéard's algorithm generates a  $\delta$ -bracketing cover  $\mathcal{B}_\delta$  of  $[0, 1]^d$  and calculates  $B(\mathcal{B}_\delta, P)$  and  $C(\mathcal{B}_\delta, P)$ .

The costs of generating  $\mathcal{B}_\delta$  are of order  $\Theta(d|\mathcal{B}_\delta|)$ . If we count the number of points in  $B^- \cap P$  and  $B^+ \cap P$  for each  $B \in \mathcal{B}_\delta$  in a naive way, this results in an overall running time of  $\Theta(dn|\mathcal{B}_\delta|)$  for the whole algorithm. As Thiéard pointed out, this orthogonal range counting can be done more effectively by employing data structures based on so-called range trees. This approach reduces in moderate dimension  $d$  the time  $O(dn)$  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

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

But since we cannot get rid of the factor  $|\mathcal{B}_\delta|$ , and the major part of the costs is due to  $|\mathcal{B}_\delta|$ , we are not interested in these technicalities here and refer to [15] for details of the implementation.

Our bound (9) tells us that

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

implying a lower bound for any algorithm based on the approach of Thiéard, regardless of the particular  $\delta$ -bracketing cover generated. Note that even the time for generating the  $\delta$ -bracketing cover is bounded from below by  $\Omega(d\delta^{-d})$ . The question is now, given an  $n$ -point set  $P$  in dimension  $d$ , what is a reasonable choice of the admissible error  $\delta$ ? Since the expected star-discrepancy of uniformly distributed, independent random variables  $\tau_1, \dots, \tau_n$  is of order  $O(d^{1/2}n^{-1/2})$  (see Subsection 2.1), and since only choices  $\delta \leq O(d_\infty^*(P))$  seem to be of interest, it is natural to choose  $\delta \leq O(n^{-1/2})$ . Notice that if one prefers to choose  $\delta < n^{-1}$ , then one should not use Thiéard's algorithm, but calculate the discrepancy exactly. Indeed, using the grid  $G = G_1 \times \dots \times G_d$ , where  $G_i$  contains 1 and the  $i$ th coordinates of all points of  $P$ , one gets

$$d_\infty^*(P) = \max_{x \in G} \left\{ V_x - \frac{|P \cap [0, x]|}{n}, \frac{|P \cap [0, x]|}{n} - V_x \right\}$$

and the right hand side can be computed in  $O(d(n+1)^d)$  time. The choice  $\delta \sim n^{-1/2}$  and the lower bound (9) lead to  $|\mathcal{B}_\delta| \geq \Omega(n^{d/2})$ , which shows that we can expect only in moderate dimension  $d$  good estimates for low-discrepancy points by using the algorithm of Thiéard.

For the special  $\delta$ -bracketing cover  $\mathcal{P}_\delta^d$  that is used in his algorithm, Thiéard proved the upper bound

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

This leads to

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

a weaker bound than the bounds  $|\mathcal{B}_\delta^d| \leq e^d \varepsilon^{-d} + C_d \varepsilon^{-d+1}$  and  $|\mathcal{B}_\delta^d| \leq 2^{d-1} e^d (\varepsilon^{-1} + 1)^d$  we established for our construction  $\mathcal{B}_\delta^d$  from the proof of Theorem 1.15.

Although the approach of Thiéard has limitations, it would be of practical interest to close the gap between the lower bound (9) and the constructive upper bounds for the cardinality of  $\delta$ -bracketing covers. Thiéard's empirical data suggests that  $N_{[\cdot]}(\mathcal{C}_d, \delta) \leq d \delta^{-d}$  holds. (However, the discussion in Remark 1.16 underlines, that even if such an estimate holds in low-dimensions, it does not necessarily hold for all dimensions.)

If one wants to approximate the extreme discrepancy of a given  $n$ -point set  $P$  up to an admissible error  $\delta$ , one can modify Thiéard's algorithm in the following way: Generate a  $\delta/2$ -bracketing cover  $\mathcal{B}$  of  $\mathcal{C}_d$ . Instead of considering  $B^- = [0, x)$  and  $B^+ = [0, y)$  for each  $B = [x, y] \in \mathcal{B}$ , it is more convenient to work with the set

$$\Gamma := \{z \mid \exists B = [x, y] \in \mathcal{B} : z = x \vee z = y\},$$

which is a  $\delta/2$ -cover of  $\mathcal{C}_d$ . According to Lemma 1.17 the set

$$\tilde{\Gamma} = \{[x, z] \mid \mathbf{1} - x, z \in \Gamma\}$$

is a  $\delta$ -cover of  $\mathcal{R}_d$ . We have

$$\max_{[x,z] \in \tilde{\Gamma}} \left| V_{x,y} - \frac{|P \cap [x,y]|}{n} \right| \leq d_\infty^e(P) \leq \max_{[x,z] \in \tilde{\Gamma}} \left| V_{x,y} - \frac{|P \cap [x,y]|}{n} \right| + \delta.$$

For counting the points in  $P \cap [x, z)$ ,  $[x, z) \in \tilde{\Gamma}$ , we again may use orthogonal range counting based on range trees, as proposed in [15]. The running time of this modified algorithm is of the same order as the running time of the original algorithm, apart from the fact that we have to substitute the factor  $|\mathcal{B}|$  by  $|\mathcal{B}|^2$ .

Let us finish by mentioning a disadvantage of generating the  $\delta$ -cover  $\tilde{\Gamma}$  via Lemma 1.17: In Thiéard's original algorithm the  $\delta$ -bracketing cover  $\mathcal{P}_\delta^d$  of  $\mathcal{C}_d$  has not to be stored, since counting the number of points of  $P$  that lie inside a  $\delta$ -bracket  $B$  can be done on the fly, i.e., directly after  $B$  is generated. But for the generation of  $\tilde{\Gamma}$  we obviously have to store the whole set  $\Gamma$ , which has cardinality at least  $\Omega((2\delta^{-1})^d)$ . Here maybe more sophisticated constructions could help to overcome this drawback.

### Acknowledgment.

The major part of the work was done while the author was staying at the University of Kiel, supported by the Deutsche Forschungsgemeinschaft under Grant SR7/10-1, and at the Max Planck Institute for Mathematics in the Sciences in Leipzig.

## References

- [1] B. Doerr, Generating randomized roundings with cardinality constraints and derandomizations, in: B. Durand, W. Thomas (Eds.), Proceedings of the 23rd Annual Symposium on Theoretical Aspects of Computer Science (STACS'06), LNCS 3884, 2006, pp. 571-583.

- [2] B. Doerr, M. Gnewuch, Construction of low-discrepancy point sets of small size by bracketing covers and dependent randomized rounding, *Berichtreihe des Mathematischen Seminars der Universität Kiel*, Report 06-14, 2006.
- [3] B. Doerr, M. Gnewuch, A. Srivastav, Bounds and constructions for the star-discrepancy via  $\delta$ -covers, *J. Complexity* 21 (2005) 691-709.
- [4] M. Drmota, R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics, vol. 1651, Springer, Berlin, 1997.
- [5] M. Gnewuch, Bounds for the average  $L^p$ -extreme and the  $L^\infty$ -extreme discrepancy, *Electron. J. Combin.* 12 (2005), Research Paper 54, 11 pp.
- [6] M. Gnewuch (Joint with B. Doerr), Geometric discrepancies and  $\delta$ -covers, *Oberwolfach Rep.* 1 (2004) 687-690.
- [7] D. Haussler, Sphere packing numbers for subsets of the Boolean  $n$ -cube with bounded Vapnik-Chervonenkis dimension, *J. Comb. Theory A* 69 (1995) 217-232.
- [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) 1885-1905.
- [10] A. Hinrichs, Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy, *J. Complexity* 20 (2004) 477-483.
- [11] H. N. Mhaskar, On the tractability of multivariate integration and approximation by neural networks, *J. Complexity* 20 (2004) 561-590.
- [12] H. Niederreiter, *Pseudo Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [13] E. Novak, H. Woźniakowski, When are integration and discrepancy tractable?, in: R. A. DeVore, A. Iserles, E. Süli (Eds.), *Foundations of Computational Mathematics*, Cambridge University Press, 2001, pp. 211-266.
- [14] M. Talagrand, Sharper bounds for Gaussian and empirical processes, *Ann. Prob.* 22 (1994) 28-76.
- [15] E. Thiérmard, An algorithm to compute bounds for the star discrepancy, *J. Complexity* 17 (2001) 850-880.
- [16] A. W. van der Vaart, J. A. Wellner, *Weaker Convergence and Empirical Processes*, Springer Series in Statistics, Springer, New York, 1996.