A Generalized Faulhaber Inequality, Improved Bracketing Covers, and Applications to Discrepancy

Michael Gnewuch∗ Hendrik Pasing† Christian Weiß‡

April 21, 2021

Abstract

We prove a generalized Faulhaber inequality to bound the sums of the $j$-th powers of the first $n$ (possibly shifted) natural numbers. With the help of this inequality we are able to improve the known bounds for bracketing numbers of $d$-dimensional axis-parallel boxes anchored in 0 (or, put differently, of lower left orthants intersected with the $d$-dimensional unit cube $[0,1]^d$). We use these improved bracketing numbers to establish new bounds for the star-discrepancy of negatively dependent random point sets and its expectation. We apply our findings also to the weighted star-discrepancy.

Keywords: Faulhaber’s formula, sums of powers, bracketing number, covering number, negative correlation, Monte Carlo point sets, pre-asymptotic bound, tractability, star-discrepancy, weighted star-discrepancy.

1 Introduction

Already in the 17th century, Faulhaber found a formula to express the $p$-th power of the first $n$ positive integers. His work did not get much recognition until it was rediscovered in the second half of the 20th century, see most notably [Knu93]. Nowadays, Faulhaber’s formula is a standard tool in $p$-adic analysis and a research topic on its own, see e.g. [Lev19] and [BMV20] for two recent examples. In this paper, we demonstrate an application of Faulhaber sums in a new context: we derive a generalized Faulhaber inequality to improve known bounds for bracketing numbers of $d$-dimensional axis-parallel boxes anchored in 0. Recall that Faulhaber’s formula is for $j \in \mathbb{N}$ given by

$$
\sum_{i=1}^{n} i^j = \frac{n^{j+1}}{j+1} + \frac{n^j}{2} + \sum_{k=2}^{j} (-1)^k \frac{B_k}{k!} (j)_{k-1} n^{j-k+1},
$$

∗Institut für Mathematik, Universität Osnabrück, Germany (michael.gnewuch@uni-osnabrueck.de).
†Institut Naturwissenschaften, Hochschule Ruhr West, Germany (hendrik.pasing@hs-ruhrwest.de).
‡Institut Naturwissenschaften, Hochschule Ruhr West, Germany (christian.weiss@hs-ruhrwest.de).
where $B_k$ is the $k$-th Bernoulli number and $(j)_{k-1} := \frac{j!}{(j-k+1)!}$ the Pochhammer symbol. Also note that all odd Bernoulli numbers with $k \geq 3$ are equal to 0. Furthermore, an easy calculation yields that the coefficient of the power $n^{j-1}$ on the right hand side is $\frac{1}{12}$.

However, we are not faced with simple sums of positive integers in the context of bracketing numbers. Instead a shifted version, namely sums of the form $\sum_{i=1}^{n}(i+r)^j$ with some shift $0 \leq r \leq 1$, appears. As a first main result we therefore derive the following generalized Faulhaber inequality:

$$\sum_{i=1}^{n}(i+r)^j \leq \frac{(n+r)^{j+1}}{j+1} + \frac{(n+r)^j}{2} + \frac{j(n+r)^{j-1}}{12},$$

see Theorem 2.4. Note that the inequality only involves positive coefficients on both sides. This is an essential technical feature which we need in our analysis of bracketing numbers. However, its potential use is not limited to the topic of this paper but might be applied to improve other (discrepancy) bounds.

Let us now describe the other main results of this paper in more detail. To this purpose we have to state the necessary definitions and concepts.

**Bracketing numbers.** The concept of bracketing numbers is well known in empirical process theory, see, e.g., [vdVW96]. We use here the definitions and notations as presented in [DGS05] and [Gne08a]: Let $A \subset [0,1]^d$. For any $\delta \in (0,1]$ a finite set of points $\Gamma \subset [0,1]^d$ is called a $\delta$-cover of $A$, if for every $y \in A$, there exist $x, z \in \Gamma \cup \{0\}$ such that $x \leq y \leq z$ and $\lambda_d([0,z]) - \lambda_d([0,x]) \leq \delta$, where $\lambda_d$ denotes the $d$-dimensional Lebesgue-measure. For $x, y \in \mathbb{R}^d$, the expression $x \leq y$ is meant as a component-wise inequality. If $x \leq y$ the interval $[x,y]$ is defined by $[x,y] := [x_1,y_1] \times [x_2,y_2] \times \ldots \times [x_d,y_d]$. We use corresponding notation for (half-)open $d$-dimensional intervals. The number $N(A,\delta)$ denotes the smallest cardinality of a $\delta$-cover in $A$. The bracketing number $N_{[1]}(A,\delta)$ of $A$ is the smallest number of closed axis-parallel boxes (or “brackets”) of the form $[x,y]$ with $x, y \in [0,1]^d$, satisfying $\lambda_d([0,y]) - \lambda_d([0,x]) \leq \delta$, whose union contains $A$. It is easily checked that the numbers $N(A,\delta)$ and $N_{[1]}(A,\delta)$ satisfy the relations

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

We will rely on the first inequality. Furthermore, we set $N_{[1]}(d,\delta) := N_{[1]}([0,1]^d,\delta)$. In dimension $d = 1$, the identity $N_{[1]}(1,\delta) = [\delta^{-1}] = N([0,1]^1,\delta)$ holds as can be seen by the help of the $\delta$-cover $\Gamma := \{1/\lceil [\delta^{-1}] \rceil, 2/\lceil [\delta^{-1}] \rceil, \ldots \}$.

For higher dimension, the following theorem was proven in [Gne08a].

**Theorem 1.1** ([Gne08a], Theorem 1.15). For any $d \geq 1$ and $1 \geq \delta > 0$ we have

$$N_{[1]}(d,\delta) \leq 2^{d-1}d! \delta^{-1} + 1)^d.$$
The result was further improved in [PW20] to:

**Theorem 1.2** ([PW20], Proposition 2.3). For any \( d \geq 1 \) and \( 1 \geq \delta > 0 \) we have

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

In this paper, we establish a smaller upper bound for the bracketing number in dimension \( d = 2 \) by an explicit geometric construction (Theorem 2.2). By comparing our proof to the simple argument in dimension \( d = 1 \), it can be seen that direct geometric arguments to approximate bracketing numbers get more and more complicated as the dimension increases. Still we are able to significantly improve the general upper bound from Theorem 1.1 and Theorem 1.2, respectively, for all \( d \in \mathbb{N} \) to

\[
N_{[\,]}(d, \delta) \leq \max \left( 1, 1.1^{d-101} \right) \frac{d^d}{d!} (\delta^{-1} + 1)^d ,
\]

see Theorem 2.5. Note that the involved power of \( 2^k \) with \( k = d - 1 \) and \( k = d - 2 \), respectively, is reduced to \( \max (1, 1.1^{d-101}) \). The proofs of Theorem 1.1 and Theorem 1.2 utilized simple integral criteria for an estimation of the bracketing number. Our improvements significantly rely on the use of the generalized Faulhaber inequality, Theorem 2.4.

**Discrepancy.** One advantage of \( \delta \)-covers is that they can be used to approximate the star-discrepancy of a sequence or set. Let \( P \) be an \( N \)-point set in \([0,1]^d \). Then the **star-discrepancy** of \( P \) is defined by

\[
D^*_N(P) := \sup_{B \subset [0,1]^d} \left| \frac{1}{N} |P \cap B| - \lambda_d(B) \right| ,
\]

where the supremum is taken over all intervals \( B = [0, a) \subset [0,1]^d \) and where \( |A| \) denotes the number of elements in a finite set \( A \). For more details we refer the reader to [DP10], [Nie92]. As a link to bracketing numbers we have the following lemma which is easy to prove.

**Lemma 1.3** ([DGS05], Lemma 3.1). Let \( P \subset [0,1]^d \) be an \( N \)-point set, \( \delta > 0 \), and \( \Gamma \) be a \( \delta \)-cover of \([0,1]^d \). Then

\[
D^*_N(P) \leq \max_{x \in \Gamma} D_N(P, [0, x)) + \delta ,
\]

where

\[
D_N(P, [0, x)) := \left| \frac{1}{N} |P \cap [0, x)| - \lambda_d([0, x)) \right| .
\]

Lemma 1.3 can be useful for several applications: It can be applied to prove (theoretical) upper discrepancy bounds (see., e.g., [Ais11] [Gne12] [NP18] and Section 3.2), to construct algorithms that approximate the discrepancy of given input sets \( P \) up to a user-specified error tolerance (see., e.g., [Thi01] [DGW14]) and algorithms that generate point
sets that exhibit a small star-discrepancy (see, e.g., [DGW09, DGW10]). Similar results based on δ-covers of [0,1]^d are known for other quality criteria for N-point sets, as, e.g., the weighted star-discrepancy (cf. [Ais14, HPS08, WGH20] and Section 3.3), the discrepancy with respect to all axis-parallel boxes, sometimes also called extreme discrepancy (see, e.g., [Gne08a]), or the dispersion (see, e.g., [Rud18, HKKR20]), a quality measure that has important applications in optimization theory (cf., e.g., [BGKTV17, Nie92]).

In all these applications it is essential to have good bounds for the bracketing number. Moreover, for constructing algorithms explicit bracketing or δ-covers of small size are essential; the smaller the δ-covers, the faster the running time of the algorithms (or, alternatively, the larger the size of the instances that can be treated). That is why improved cardinality bounds and constructions of δ-covers are highly relevant.

The star-discrepancy is closely related to the construction of optimal integration rules via the Koksma-Hlawka inequality, namely: For every N-point set \( P \subset [0,1)^d \) we have

\[
\left| \int_{[0,1]^d} f(x) d\lambda_d(x) - \frac{1}{N} \sum_{p \in P} f(p) \right| \leq D_N^*(P) V(f),
\]

where \( V(f) \) is the variation of \( f \) in the sense of Hardy and Krause. If all partial mixed derivatives of \( f \) are continuous on \([0,1]^d\) then

\[
V(f) = \sum_u \int_{[0,1]^u} \left| \frac{\partial^{u_0} f}{\partial x_u}(x_u, 1) \right| dx_u,
\]

where the sum is taken over all subsets \( u \subset \{1, 2, \ldots, d\} \) and \((x_u, 1)\) is the \(d\)-dimensional vector whose \(i\)-th component is \(x_i\) if \(i \in u\) and 1 otherwise, see [KN74], Chapter 2.

It follows from the Koksma-Hlawka inequality that integration rules with small star-discrepancy yield small integration errors. The task of high-dimensional integration occurs in different applications, e.g. in computational finance (see e.g. [CMO97], [GW09], [Gla03], [Pas94], [WN19]). Therefore, it is of interest to know sharp bounds for the smallest achievable star-discrepancy

\[
D^*(N, d) := \inf \{ D_N^*(P) | P \subset [0,1]^d, |P| = N \}
\]

or, equivalently, for the inverse of the star-discrepancy

\[
N^*(\epsilon, d) := \inf \{ N \in \mathbb{N} | D^*(N, d) \leq \epsilon \}
\]

which is the minimum number of sample points that guarantees a discrepancy bound of at most \( \epsilon \). If an N-point set \( P \subset [0,1]^d \) exhibits a star-discrepancy of order

\[
D_N(P) = O(N^{-1}(\log N)^{d-1}),
\]

then it is called a low-discrepancy point set. It is conjectured that this is the best possible rate of convergence. In fact, this is trivially true in dimension one and it is known
to be true for dimensions two by the work of Schmidt, [Sch72]. For arbitrary dimension \( d \) lower bounds show that (1) is sharp up to logarithmic factors, see [BL08], [BLV08]. Due to the exponential dependence of the asymptotic bound (1) on \( d \) it does not provide helpful information about the star-discrepancy of sample sets of moderate size, cf., e.g., [Gne12]. Therefore, researchers started to prove pre-asymptotic bounds for the star-discrepancy.

The best known upper and lower bounds for the smallest achievable star-discrepancy with explicitly given dependence on the number of points as well as on the dimension are of the form

\[
D^*_N(P) \leq C\sqrt{\frac{d}{N}}
\]

for some constant \( C \). This implies

\[
N^*(\varepsilon, d) \leq \lceil C^2d\varepsilon^{-2} \rceil
\]

for all \( d, N \in \mathbb{N} \) and \( \varepsilon \in (0, 1] \). The theoretical existence of such a \( C \) was shown in [HNWW01] but no concrete value for \( C \) was calculated therein. In [Ais11], a new proof of the result was given including the first explicit upper bound \( C = 10 \) (or more precisely \( C = 9.65 \)). In [PW20], the value was improved to \( C = 9 \), in a previous version of the recent preprint [Doe21] to \( C = 2.7868 \), and in [GH19] even to \( C = 2.5287 \). In the course of this paper, we will show that indeed \( C \leq 2.4968 \) holds (Corollary 3.7). Even small improvements of the constant \( C \) are not only of pure theoretical interest but lead to a smaller necessary number of points \( N \) to guarantee a given level of precision for integration (observe that due to (2) \( N \) scales quadratically in \( C \)). Moreover the constant can be used to check whether a given (random or deterministic) point set or sequence is good in the sense that its empirically observed star-discrepancy is close to or even smaller than \( C\sqrt{\frac{d}{N}} \).

For instance, the authors of [GGP20] conclude that the sequence generated by a secure bit generator is good up to dimension at least \( d = 15 \), because its empirically observed discrepancy is smaller than \( 1 \cdot \sqrt{\frac{d}{N}} \) also for relatively big \( N \).

It was proved in [Hin04] that there exist \( c, \varepsilon_0 > 0 \) such that

\[
N^*(\varepsilon, d) \geq cd\varepsilon^{-1}
\]

for all \( 0 < \varepsilon \leq \varepsilon_0, d \in \mathbb{N} \). Note that there is a gap between the lower and the upper bound on \( N^*(\varepsilon, d) \) with regard to the dependence on \( \varepsilon^{-1} \). In [Doe14], it was shown that this gap does not exist for typical Monte Carlo points. More precisely, there exists a constant \( K > 0 \) such that the expected star-discrepancy of a Monte Carlo \( N \)-point sample fulfills

\[
\mathbb{E}[D^*_N(X)] \geq K\sqrt{\frac{d}{N}}
\]

and additionally there is the probabilistic error bound

\[
\mathbb{P}\left(D^*_N(X) < K\sqrt{\frac{d}{N}}\right) \leq \exp(-\Theta(d)).
\]
An analogous result for Latin hypercube sampling instead of Monte Carlo sampling holds for dimension \(d \geq 2\), see [DDG18]. In the course of this paper we complement estimate (3) by showing that Monte Carlo \(N\)-point samples satisfy

\[
E[D^*_N(X)] \leq 2.55648 \sqrt{\frac{d}{N}},
\]

see Corollary 3.10. This is an immediate consequence of Proposition 3.9 which shows how to bound the expected discrepancy of arbitrary random points with the help of probabilistic discrepancy bounds. Estimate (5) may serve as an indicator for a given random point set whether its discrepancy is better or worse than average. We close this paper by studying the so-called weighted star-discrepancy and derive probabilistic upper bounds for it, see Proposition 3.16 as well as Corollaries 3.17 and 3.18. The general upper bound in Proposition 3.16 gives immediately the upper bound in Corollary 3.17, which results in an improvement of a result of Aistleitner, see [Ais14, Theorem 1]. Furthermore, Corollary 3.18 improves a bound from [WGH20].

2 Improved Bracketing Number

In this section, we derive new upper bounds for bracketing numbers. In dimension \(d = 2\), we do this by an explicit geometric construction. Roughly speaking, we get an additional factor \(\log(2)\) implying an improvement of approximately \(30\%\) of the highest order constant in comparison to [PW20], Proposition 2.3. For arbitrary higher dimension we are able to reduce the involved power of \(2^k\) from \(2^{d-1}\) in Theorem 1.1 and \(2^{d-2}\) in Theorem 1.2, respectively, to \(\max(1, 1.1^{d-101})\) in Theorem 2.5.

2.1 Constructive Bound in Dimension \(d = 2\)

While it is straightforward to calculate the bracketing number for dimension \(d = 1\), namely \(N_{\|}(1, \delta) = \lceil \delta^{-1} \rceil\), the task is already much harder in dimension 2. The general construction of the bracketing cover for our improved bound of \(N_{\|}(2, \delta)\) follows the lines of [Gnc08b]. Here we provide a more careful analysis as in [Gnc08b] to derive a bound on \(N_{\|}(2, \delta)\) with explicitly given constants. The construction goes as follows:

In a first step, we cover the diagonal of \([0, 1]^2\) with \(\delta\)-brackets by choosing the points \(a_q := (a_q, a_q), 0 \leq q \leq n-1\), as points of our \(\delta\)-cover, where \(a_q := \sqrt{1 - q\delta}\) and \(n := \lceil \delta^{-1} \rceil\). Furthermore, we put \(a_n := (0, 0)\). Then

\[
\lambda_2([0, a_{q-1}]) - \lambda_2([0, a_q]) = \delta \quad \text{for all } 1 \leq q \leq n.
\]

We observe that

\[
N_{\|}(2, \delta) \leq \sum_{q=1}^{n} N_{\|}([0, a_{q-1}] \setminus [0, a_q], \delta).
\]
In a second step, we cover each of the sets \([0,a_{q-1}] \setminus [0,a_q], 1 \leq q \leq n - 1\), with \(\delta\)-brackets and add the upper right corner points of the brackets to our \(\delta\)-cover \(\Gamma\). The construction of our \(\delta\)-cover leads to the upper bound on \(N[\delta](\{0,a_{q-1}\} \setminus [0,a_q], \delta)\) presented in Lemma 2.1 below; the construction itself is described in detail in the proof of Lemma 2.1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Unit square with marked points}
\end{figure}

\textbf{Lemma 2.1.} Let \(0 < \delta < 1\) be arbitrary and choose \(q \leq n\). For \(\delta_q := \frac{\delta}{1-(q-1)\delta}\) we have

\[
N[\delta_q](\{0,a_{q-1}\} \setminus [0,a_q], \delta) \leq 2 \left[\frac{-2 \cdot \log(2)}{\log(1-\delta_q)}\right] + 1 =: 2f(\delta_q) - 1.
\]

For \(q = n\) we even have \(N[\delta_q](\{0,a_{n-1}\} \setminus \{0\}, \delta) = 1\).

\textbf{Proof.} At first, we consider the case \(q = 1\). Then we define \(b_0 := 1\) and let \(b_i\) be 1 minus the sum over the widths of the \(i\) first \(\delta\)-brackets covering \([0,1]^2 \setminus ([0,1] \times [0,a_1])\). Furthermore, let \(k_i := (b_{i-1},1)\) and \(l_i := (b_i,b_1)\) be the points as indicated in Figure 1. We choose the \(b_i\) such that the condition

\[
\lambda_2([0,k_i]) - \lambda_2([0,l_i]) = \delta
\]

is satisfied. Note that \(b_1 = a_1\). If \(b_i \leq \delta\), then \([l_j,k_j], j = 1,\ldots,i\), and \([0,k_{i+1}]\) cover \([0,1]^2 \setminus ([0,1] \times [0,a_1])\), hence \(N[\delta_q](\{0,1\}^2 \setminus ([0,1] \times [0,a_1]), \delta) \leq f = f(\delta) := i+1\). Otherwise, \(b_{i+1}\) is defined according to (7) and hence given by the equation

\[
\delta = b_i \cdot 1 - b_{i+1} \cdot b_1
\]
which is equivalent to
\[ b_{i+1} = \frac{b_i - \delta}{b_1}. \]  
(8)

From equation (8) it follows inductively that
\[ b_{i+1} = b_1^{-i+1} - \delta \sum_{j=1}^{i} b_1^{-j} = b_1^{-i+1} - \delta \cdot b_1^{-i} \cdot \frac{b_1^{-1} - 1}{b_1^{-i} - 1} . \]  
(9)

Now the aim is to calculate the minimal \( i \) such that \( b_i \leq \delta \) – recall that this \( i \) satisfies \( f = f(\delta) = i + 1 \). From (9) it follows that the latter condition is equivalent to
\[ b_1 \cdot \frac{1 - b_1}{1 - b_i^1} \leq \delta, \]
which yields
\[ i \geq \frac{\log(1 - \delta^{-1}(b_1 - b_i^1))}{\log(b_1)} . \]

Plugging in \( b_1 = \sqrt{1 - \delta} \) leads to
\[ f = \left[ \frac{\log \left( 1 - \delta^{-1}(\sqrt{1 - \delta} - (1 - \delta)) \right)}{\log \left( \sqrt{1 - \delta} \right)} \right] + 1 \]
\[ = \left[ 2 \frac{\log(1 - \sqrt{1 - \delta}) - \log(\delta)}{\log(1 - \delta)} \right] + 1. \]

Applying the estimate \( \sqrt{1 - x} \leq 1 - \frac{x}{2} \) for \( 0 \leq x \leq 1 \), we obtain
\[ f \leq \left[ \frac{-2 \log(2)}{\log(1 - \delta)} \right] + 1 \]

Clearly, we can cover \([0, 1]^2 \setminus ([0, a_1] \times [0, 1])\) with \( f \) different \( \delta \)-brackets by simply applying the reflection \( \Psi(x_1, x_2) = (x_2, x_1) \) to the \( \delta \)-brackets we used to cover \([0, 1]^2 \setminus ([0, 1] \times [0, a_1])\). Consequently, we can cover \([0, 1]^2 \setminus [0, a]\) with \( 2f - 1 \) different \( \delta \)-brackets. This implies the claim for \( q = 1 \). For \( q > 1 \), note that the transformation \( \Phi(x_1, x_2) = a_{q-1}^{-1} \cdot (x_1, x_2) \) maps \([0, a_{q-1}] \setminus [0, a_q] \) onto \([0, 1]^2 \setminus ([0, \sqrt{1 - \delta_q}] \times [0, \sqrt{1 - \delta_q}]) \) and the general result thus follows from the case \( q = 1 \). \( \square \)

Now we use (6) and the preparatory Lemma 2.1 to improve the bound for the bracketing number in dimension \( d = 2 \).

**Theorem 2.2.** Let \( 0 < \delta < 1 \) be arbitrary. For the bracketing number of \([0, 1]^2\) we have
\[ N_{\mid}[(2, \delta) \leq 2 \log(2)\delta^{-2} + 3(\log(2) + 1)\delta^{-1} - \left( \frac{13}{9} \log(2) - 1 \right). \]  
(10)
Proof. Let \( \kappa_i := 2f(\delta_i) - 1 \), where \( \delta_i \) is defined as in Lemma 2.1. Then (6) and Lemma 2.1 imply

\[
N_{\square}(2, \delta) \leq \sum_{i=1}^{n} \kappa_i \leq \sum_{i=1}^{n-1} 4 \log(2) \frac{1}{|\log (1 - \delta_i)|} + 3(n - 1) + 1.
\]

Recall that \(|\log(1 - x)| \geq x + \frac{x^2}{2} + \frac{x^3}{3}\) for \(0 \leq x < 1\). Hence it follows that

\[
N_{\square}(2, \delta) \leq 4 \log(2) \delta^{-1} + 3(n - 1) + 1.
\]

From \((n - 1)(\delta^{-1} + 1) - \frac{n(n-1)}{2} \leq \frac{1}{2} \delta^{-2} + \delta^{-1}\) we finally deduce

\[
N_{\square}(2, \delta) \leq 2 \log(2) (1 + \frac{\delta}{2} + \frac{\delta^2}{3})^{-1} (\delta^{-2} + 2 \delta^{-1}) + 3(n - 1) + 1
\]

\[
= 2 \log(2) (1 + \frac{\delta}{2} + \frac{\delta^2}{3})^{-1} \left(\delta^{-2} \left(1 + \frac{\delta}{2} + \frac{\delta^2}{3}\right) + \frac{3}{2} \delta^{-1} \left(1 + \frac{\delta}{2}\right) - \frac{13}{12}\right) + 3(n - 1) + 1
\]

\[
\leq 2 \log(2) \delta^{-2} + 3(\log(2) + 1) \delta^{-1} - \left(\frac{13}{9} \log(2) - 1\right).
\]

\[\square\]

**Remark 2.3.** It should be mentioned that in [Gne08b] the following constructive upper bound for the bracketing number in dimension \(d = 2\) was derived:

\[
N_{\square}(2, \delta) = \delta^{-2} + o(\delta^{-2}). \tag{11}
\]

This bound is asymptotically better than the one we presented in Theorem 2.2 and it is (essentially) optimal as was shown in [Gne08a] by the following lower bound

\[
N_{\square}(2, \delta) = \delta^{-2} - O(\delta^{-1}).
\]

The essential point in the constructive upper bound in Theorem 2.2 is that we have control over the coefficients of all powers of \(\delta^{-1}\); we even know the exact number of brackets of the underlying \(\delta\)-bracketing cover. This is not the case for the construction that leads to (11). It seems to be too involved to provide a simple formula for the exact number of its \(\delta\)-brackets.
2.2 Upper Bound on the Bracketing Number for Arbitrary Dimension

Faulhaber’s formula. An important ingredient for our improved upper bound on the bracketing number for arbitrary dimensions is a generalized version of the so-called Faulhaber formula, see [Knu93]. Using the whole Faulhaber formula directly would be of limited use in our context because it also involves negative coefficients. Furthermore in our calculations for improving the bound on the bracketing numbers, there do not appear sums of the form \( \sum_{i=1}^{n} i^j \) but sums involving a shift parameter \( 0 < r < 1 \), i.e., \( \sum_{i=1}^{n} (i+r)^j \). Therefore, we derive at first the following inequality.

**Theorem 2.4** (Generalized Faulhaber inequality). Let \( j, n \in \mathbb{N} \) and \( 0 \leq r \leq 1 \) be arbitrary. Then the following inequality holds

\[
\sum_{i=1}^{n} (i+r)^j \leq \frac{(n+r)^{j+1}}{j+1} + \frac{(n+r)^j}{2} + \frac{j(n+r)^{j-1}}{12}.
\]

(12)

**Proof.** We prove the assertion by induction on \( n \). For \( n = 1 \) we have \((1+r)^j\) on the left hand side of (12) and

\[
(1+r)^j \left( \frac{1+r}{j+1} + \frac{1}{2} + \frac{j}{12(1+r)} \right) \equiv f_r(j)
\]

on the right hand side of (12). Allowing arbitrary \( x \in (-1, \infty) \), we see that \( f_r(x) \) has a unique global minimum in \( x_r := \sqrt{12(1+r)} - 1 \). This leads for fixed \( r \in [0,1] \) to the minimal value

\[
f_r(x_r) = \frac{2}{\sqrt{12}} + \frac{1}{2} - \frac{1}{12(1+r)}.
\]

For \( r \geq 0.07736 \) we have \( f_r(x_r) > 1 \). For \( r \in [0, 0.07736) \) the minimum of \( f_r(j), j \in \mathbb{N} \), is taken in \( j = \lfloor x_r \rfloor = 2 \) or in \( j = \lceil x_r \rceil = 3 \). For both values of \( j \) we have \( f_r(j) \geq 1 \), since \( f_0(j) = 1 \) and \( \frac{\partial}{\partial r} f_r(j) \geq 0 \).

Now we consider \( n+1 \). Applying the induction hypothesis and binomial expansion to the left hand side of (12) yields

\[
\sum_{i=1}^{n+1} (i+r)^j \leq \frac{(n+r)^{j+1}}{j+1} + \frac{(n+r)^j}{2} + \frac{j(n+r)^{j-1}}{12} + (n+1+r)^j
\]

\[
= \frac{(n+r)^{j+1}}{j+1} + \frac{(n+r)^j}{2} + \frac{j(n+r)^{j-1}}{12} + \sum_{k=0}^{j} \binom{j}{k} (n+r)^k.
\]

Similarly, the right hand side of (12) can be written as

\[
\frac{1}{j+1} \sum_{k=0}^{j+1} \binom{j+1}{k} (n+r)^k + \frac{1}{2} \sum_{k=0}^{j} \binom{j}{k} (n+r)^k + \frac{j}{12} \sum_{k=0}^{j-1} \binom{j-1}{k} (n+r)^k.
\]
Next we compare the coefficients of the powers of $n + r$ on both sides. For $k = j + 1, j, j - 1, j - 2, j - 3$ the coefficients are equal on both sides, namely $\frac{1}{j+1}, \frac{3}{2}, \frac{13j}{12}, \frac{j(j-1)}{2}$ and $\frac{j(j-1)(j-2)}{6}$. Let $k = j - s \in \mathbb{N}_0$ with $j \geq s \geq 4$ be arbitrary. Then the coefficient on the left side is $\binom{j}{j-s}$. On the right side, the coefficient is

$$\frac{1}{j+1} \binom{j+1}{j-s} + \frac{1}{2} \binom{j}{j-s} + \frac{j}{12} \binom{j-1}{j-s} = \left( \frac{1}{s+1} + \frac{1}{2} + \frac{s}{12} \right) \binom{j}{j-s}.$$

The expression in the bracket is at least 1 for $s \geq 3$. This finishes the proof.

**Main result on bracketing numbers.** The generalized Faulhaber inequality from Theorem 2.4 is the main tool for improving the upper bound on the bracketing number for arbitrary dimension.

**Theorem 2.5.** For all $0 < \delta < 1$ and $d \in \mathbb{N}$ we have

$$N[\mathord{]} (d, \delta) \leq \max \left( 1.1^{d-101}, 1 \right) \frac{d^d}{d!} (\delta^{-1} + 1)^d. \quad (13)$$

In order to keep the presentation of the proof of Theorem 2.5 as clear as possible, we present two preparatory results as separate statements.

**Lemma 2.6.** Given $d \in \mathbb{N}$, the function

$$f(k) = \frac{d!}{d^{k-1}(d-k+1)!} \left( \frac{k}{12} + \frac{1}{2} \right)$$

with $k \in \mathbb{N}$ is maximal for $k = -3 + \lceil \sqrt{16 + d} \rceil$.

**Proof.** We have

$$\frac{f(k+1)}{f(k)} = \frac{k+7}{k+6} \cdot \frac{d-k+1}{d}.$$

This expression is $\leq 1$ if and only if

$$\frac{k+7}{k+6}(d-k+1) = \frac{k+7}{k+6}d + \frac{(k+7)(1-k)}{k+6} \leq d.$$

which is yet equivalent to

$$-k^2 - 6k + 7 + d \leq 0$$

or, in other words,

$$k \geq -3 + \sqrt{16 + d}.$$

This implies the claim. \hfill \Box
Lemma 2.7. Given $d \in \mathbb{N}$, the function
\[ g_d(k) = \frac{d!}{d^{k-1}(d-k+1)!} \left( \frac{k}{12} + \frac{1}{2} + \frac{1}{k+1} \right) \]
with $k \in \mathbb{N}$ takes its maximum in $k_{\max} \leq \min\{-3 + \lceil \sqrt{16 + d} \rceil, 1 + \lceil 0.0544 \cdot d \rceil\}$.

Proof. Since $r(k) := g_d(k) - f(k)$ is strictly monotone decreasing on $\mathbb{N}$, we obtain from Lemma 2.6 that $k_{\max} \leq -3 + \lceil \sqrt{16 + d} \rceil$, because $g(k) = f(k) + r(k)$ and both $f(k_1) < f(-3 + \lceil \sqrt{16 + d} \rceil)$ and $r(k_1) < r(-3 + \lceil \sqrt{16 + d} \rceil)$ for all $k > -3 + \lceil \sqrt{16 + d} \rceil$. Furthermore, we have for all $1 \leq k < d$
\[ h(k) := \frac{g_d(k+1)}{g_d(k)} = \frac{d-k+1}{d} \cdot \frac{k+1}{k+2} \cdot \frac{k^2 + 9k + 26}{k^2 + 7k + 18}, \]
and
\[ u(k) := \frac{d}{d-k+1} h(k) \]
satisfies $\lim_{k \to \infty} u(k) = 1$. It is easily checked that $u(k)$ takes its maximum in $k = 7$, for $k \in \mathbb{N}$, namely $u(7) = 1.05747126$. Hence, if $k_0 = k_0(d) \in \mathbb{N}$ satisfies the estimate
\[ \frac{d-k_0+1}{d} \cdot 1.05747126 \leq 1, \]
then all $k \geq k_0$ satisfy it, too, implying $h(k) \leq 1$ and thus $g_d(k) \geq g_d(k+1)$. This yields the necessary condition
\[ k_{\max} \leq 1 + \lceil d \cdot 0.0544 \rceil. \]

\[ \square \]

Proof of Theorem 2.5. Let $b_d := \max (1.1^{d-101}, 1)$ and $\nu := \lceil \delta^{-1} \rceil$. We proceed by induction over the dimension $d$. For $d = 1$ the claim is obviously true. Let now $d \geq 2$. Analogously to the proof of Theorem 1.15 in [Gne08a] the induction hypothesis with induction step $d-1 \to d$ yields
\[ N_{[j]}(d, \delta) \leq \sum_{k=1}^{d-1} \binom{d}{k} b_{d-k} \frac{d^{d-k}}{(d-k)!} \left( \sum_{i=1}^{n-1} (\delta^{-1} - i + 1)^{d-k} \right) + \delta^{-1} + 1 \]
(where in [Gne08a] $b_\ell$ was actually equal to $2^\ell$ for all $\ell \in \mathbb{N}$). Applying the generalized Faulhaber inequality from Theorem 2.4 with $j = d-k$ and $r = \delta^{-1} - n + 1$ we get
\[ N_{[j]}(d, \delta) \leq \sum_{k=1}^{d-1} \binom{d}{k} b_{d-k} \frac{d^{d-k}}{(d-k)!} \left( \frac{1}{d-k+1} (\delta^{-1})^{d-k+1} + \frac{1}{2} (\delta^{-1})^{d-k} + \frac{d-k}{12} (\delta^{-1})^{d-k-1} \right) \]
\[ + \delta^{-1} + 1. \]
(14)
The right hand side of inequality (13) is equal to
\[ b_d \frac{d^d}{d!} (\delta^{-1} + 1)^d = \sum_{k=0}^{d} b_d \frac{d^d}{d!} (\delta^{-1})^{d-k} = \sum_{k=0}^{d} b_d \frac{d^d}{k!(d-k)!} (\delta^{-1})^{d-k}. \] (15)

Now we compare the coefficients in front of the powers of \(\delta^{-1}\) in (14) and (15). We start with the lowest powers, \(p = 0\) and \(p = 1\). For \(p = 0\) the coefficient in (14) is
\[
\binom{d}{d-1} b_1 \frac{d^{12}}{12} + 1 = \frac{d^{12}}{12} + 1
\]
and the corresponding coefficient in (15) is \(b_d \frac{d^d}{d!}\). It is easily checked that
\[
\frac{d^{12}}{12} + 1 \leq \frac{d^d}{d!} \leq b_d \frac{d^d}{d!} \quad \text{for all } d \geq 2.
\]
For \(p = 1\) the coefficient in (14) is
\[
\frac{d^2}{2} + \frac{d^3 (d-1)}{24} + 1
\]
and the corresponding coefficient in (15) is \(b_d \frac{d^d}{(d-1)!}\), and again it is easily verified that
\[
\frac{d^2}{2} + \frac{d^3 (d-1)}{24} + 1 \leq \frac{d^d}{(d-1)!} \leq b_d \frac{d^d}{(d-1)!} \quad \text{for all } d \geq 2.
\]
For \(p = d\) the corresponding coefficients in (14) is not larger than the one in (15), since \(b_{d-1} \leq b_d\). For \(p = d - 1\), \(p \geq 2\), it can be easily checked that the coefficient in (14) is strictly less than the one in (15). Next we compare the coefficients for powers \(2 \leq p \leq d-2\), i.e., for powers \(d - k\) with \(2 \leq k \leq d - 2\). In (14) the respective coefficient is given by
\[
\frac{d^{d-k-1}}{(d-k)!} \binom{d}{k+1} b_{d-k} + b_{d-k} \frac{d}{2} \binom{d}{k} + b_{d-k+1} \frac{d^2}{12} \binom{d}{k-1}.
\]
To show that this coefficient is smaller than the corresponding one from (15) it suffices to show that
\[
\frac{b_{d-k+1}}{b_d} \frac{k!}{d^{k+1}} a_{d-k} \leq 1.
\] (16)

At first, let us consider the case \(d \leq 101\), i.e., \(b_d = 1 = b_{d-k+1}\). We obtain the inequality
\[
A_{d-k} := k! \cdot \frac{a_{d-k}}{d^{k+1}} = \frac{d!}{d^{k+1}(d-k+1)!} \cdot \frac{1}{d^2} \left( \frac{(d-k+1) \cdot (d-k)}{k+1} + \frac{d \cdot (d-k+1)}{2} + \frac{d^2 \cdot k}{12} \right)
\leq \frac{d!}{d^{k+1}(d-k+1)!} \left( \frac{1}{k+1} + \frac{1}{2} + \frac{k}{12} \right) = g_d(k) \leq g_d(k_{\max}) \leq f_0(k_{\max}),
\] (17)
where $k_{\text{max}}$ is as in Lemma 2.7. From this and Lemma 2.7 we get for $d \leq 36$ that $A_{d-k} \leq 1$, because $k_{\text{max}} \leq \min\{-3 + \lceil \sqrt{16 + d} \rceil, 1 + \lceil 0.0544 \cdot d \rceil\} < 4$ and hence $f_0(k_{\text{max}}) \leq 1$. It can be checked by a simple computer calculation of all possible values of $A_{d-k}$ that in fact $A_{d-k} \leq 1$ holds for $d \leq 101$. Thus (16) holds and consequently the coefficients in (14) are smaller than those in (15) for all powers of $\delta^{-1}$ and the claim follows for $d \leq 101$. Now let $d > 101$. For all $k$ with $d - k + 1 > 101$ we see that

$$b_{d-k+1} A_{d-k} = \frac{1}{1.1^{d-101}} A_{d-k} \leq \frac{1}{1.1^{d-101}} g_d(k_{\text{max}}) \leq \frac{1}{1.1^{d-101}} f_0(k_{\text{max}}) =: \tilde{g}_d(k_{\text{max}})$$

for all $d - 100 \leq k \leq d - 2$. A direct calculation yields $k_{\text{max}}(102) = 7$ and $g_{102}(7) < 1.01$. Hence (16) is true for $d = 102$. Note that $f_0(k)$ is monotonic increasing for $k \geq 2$ and, due to Lemma 2.7,

$$k_{\text{max}} = k_{\text{max}}(d) \leq \min\{-3 + \lceil \sqrt{16 + d} \rceil, 1 + \lceil 0.0544 \cdot d \rceil\} =: k_0(d)$$

which is monotonic increasing as well, we have that

$$\frac{b_{d-k+1}}{b_d} A_{d-k} \leq \tilde{g}_d(k_0(d)).$$

Now $k_0(103) = 7$ implies

$$\tilde{g}_{103}(k_{\text{max}}(103)) \leq \tilde{g}_{103}(k_0(103)) < 0.999.$$ 

Since $f_0(k + 1)/f_0(k) < 1.1$ for $k \geq 7$ and $|k_0(d + 1) - k_0(d)| \leq 1$ for all $d$, we see that $\tilde{g}_d(k_{\text{max}}(d)) \leq 1$ for all $d \geq 103$. This establishes (16) in the case $d > 101$ and finishes the proof. 

**Remark 2.8.** Note that our argumentation would fail when using the complete Faulhaber formula instead of the inequality from Theorem 2.4 because there appear negative terms in it.
3 Application to Discrepancy Bounds

In this section, we apply the new general upper bound on the bracketing number from Theorem 2.5 to derive new discrepancy bounds for the smallest achievable star-discrepancy and weighted star-discrepancy with explicitly given dependence on the number of points \( N \) and on the dimension \( d \). We will not restrict our analysis to classical Monte Carlo point sets but use the more general notion of negatively dependent sampling which we introduce first.

3.1 Negative Dependence of Sampling Schemes

Recall that classical Monte Carlo involves random samples which are independent and identically uniformly distributed on \([0, 1)^d\). Monte Carlo methods do not yield deterministic convergence and hence also result in random errors and the variance of the estimation has to be taken into account. Therefore, several variance reduction techniques for Monte Carlo have been established, see, e.g., [Gla03], Chapter 4. For this and other similar reasons, researchers proposed to use randomized Quasi-Monte Carlo methods, see, e.g., [CP76], [Owe95] or [L’Ec17]. A very recent research direction is to use notions of negative dependence to analyze random points sets, see, e.g., [GH19], [Lem18], [WG20], [WGH20], [WLD19]. This line of research is motivated in detail in the introductions of the papers [DDG18, GH19, Lem18]. We consider here the notions of negative dependence studied in [GH19] and [WGH20].

**Definition 3.1.** Let \( \gamma \geq 1 \). The binary random variables \( T_1, T_2, \ldots, T_N \) are called upper \( \gamma \)-negatively dependent if
\[
P \left( \bigcap_{j \in u} \{ T_j = 1 \} \right) \leq \gamma \prod_{j \in u} P ( T_j = 1 ), \quad \text{for all } u \subseteq [N]
\]
and lower \( \gamma \)-negatively dependent if
\[
P \left( \bigcap_{j \in u} \{ T_j = 0 \} \right) \leq \gamma \prod_{j \in u} P ( T_j = 0 ), \quad \text{for all } u \subseteq [N],
\]
where the set \( \{1, 2, \ldots, N\} \) is denoted by \([N]\). If both conditions are satisfied, then \( T_1, T_2, \ldots, T_N \) are called \( \gamma \)-negatively dependent.

In this paper, we are interested in binary random variables \( T_i, i = 1, \ldots, N \), of the form \( T_i = 1_A(X_i) \), where \( A \) is a Lebesgue-measurable subset of \([0, 1]^d\), \( 1_A \) denotes the characteristic function of \( A \), and \( X_1, \ldots, X_N \) are randomly chosen points in \([0, 1]^d\). If such a sequence is \( \gamma \)-negatively dependent, then it fulfills special bounds of Hoeffding- and Bernstein-type, respectively, see e.g. [Heb12].

**Definition 3.2.** A randomized point set \( P = (p_j)_{j=1}^N \) in \([0, 1)^d\) is called a sampling scheme, if every single \( p \in P \) is distributed uniformly in \([0, 1)^d\) and the vector \((p_1, \ldots, p_N)\) is exchangeable, meaning that for any permutation \( \pi \) on \([N]\) it holds that the probability distribution of \((p_1, \ldots, p_N)\) is the same as the one of \((p_{\pi(1)}, \ldots, p_{\pi(N)})\).
The assumption of exchangeability is easily satisfied. Let \( \tilde{P} \) be an \( N \)-point set such that every \( \tilde{p} \in \tilde{P} \) is uniformly distributed in \([0, 1]^d\) and let \( \pi \) be a random uniformly chosen permutation of \([N]\). Then \( P := (\tilde{p}_{\pi(j)})_{j=1}^{N} \) is already a sampling scheme.

**Definition 3.3.** For a given set \( S \) of Lebesgue measurable subsets of \([0, 1]^d\) we say that a sampling scheme \( P = (p_j)_{j=1}^{N} \) is \( S - \gamma \)-negatively dependent if for every \( Q \in S \) the random variables \((1_Q(p_j))_{j=1}^{N}\) are \( \gamma \)-negatively-dependent.

In this paper, we are interested in two special cases, namely \( S \in \{C_0^d, D_0^d\} \), where \( C_0^d := \{[0, a) : a \in [0, 1)^d\} \) is the set of boxes anchored at 0 (sometimes called corners) and \( D_0^d := \{Q \setminus R : Q, R \in C_0^d\} \) is the set of differences of boxes from \( C_0^d \).

### 3.2 Bounds for the Star-Discrepancy

For deriving discrepancy bounds, we will use an intermediate result of the proof of Theorem 4.4 in [GH19] and therefore stick to the notation therein. Hence, we need to introduce some variables. First let \( \mu \in \mathbb{N}, \mu \geq 2 \) be arbitrary and

\[
\ell_\mu := \frac{1}{1 - \sqrt{\frac{\mu + 1}{2\mu}}}.
\]

Furthermore let \( c_0, c_1 > 0 \) be two constants and choose \( K \geq \mu \) as the smallest natural number with

\[
\frac{1}{\sqrt{K^2 \mu}} \leq c_0 c_1 c_\mu \sqrt{\frac{d}{N}}
\]

and set

\[
\tau_\mu := \frac{c_1^2}{4(1 + 1/(3c_\mu))} \quad \text{implying} \quad c_1 = \sqrt{4\tau_\mu \left(1 + \frac{1}{3c_\mu}\right)}.
\]

Note that \( c_0 \) may still be chosen arbitrarily. For each \( \mu \leq k \leq K \) we choose a \( 2^{-k} \)-cover \( \Gamma_k \) of minimum size.

**Lemma 3.4 ([GH19], Proof of Theorem 4.4).** Let \( d, N \in \mathbb{N} \) and \( \rho \in [0, \infty) \). Let \( X = (X_j)_{j=1}^{N} \) be a \( D_0^d - e^{\rho d} \)-negatively dependent sampling scheme. Then there exist events \( E_k, k = \mu, \ldots, K \), with

\[
P(E_{\mu}^c) \leq 2|\Gamma_\mu|e^{\rho d}e^{-2c_0^2d} \quad (18)
\]

and

\[
P(E_{k}^c) \leq 2|\Gamma_k|e^{\rho d}e^{-2c_0^2\tau_\mu(k-1)d} \quad (19)
\]
for \( \mu < k \leq K \) and if the event

\[
E := \bigcap_{k=\mu}^{K} E_k
\]

occurs then any realization \( P \) of \( X \) satisfies

\[
D^*_N(P) \leq c_0 \left( 1 + c_1 c_\mu \sqrt{\frac{\mu}{2^\mu}} \right) \sqrt{\frac{d}{N}}.
\]

From Lemma 3.4 and Theorem 2.5 it follows that the star-discrepancy of a negatively dependent sampling scheme fulfills

\[
D^*_N(X) \leq c \sqrt{\frac{d}{N}}
\]

with high probability as long as \( c \) is sufficiently large.

**Theorem 3.5.** Let \( d, N \in \mathbb{N} \) with \( d \geq 2 \) and \( \rho \in [0, \infty) \). Let \( X = (X_j)_{j=1}^N \) be a \( D^d_{-\rho d} \)-negatively dependent sampling scheme. Then for every \( c > 2.4968 \)

\[
D^*_N(X) \leq c \sqrt{\frac{d}{N}} \tag{20}
\]

holds with probability at least \( 1 - e^{-(1.67681 c^2 - 10.45292 - \rho) d} \) implying that for every \( q \in (0, 1) \)

\[
D^*_N(X) \leq 0.7723 \sqrt{10.45292 + \rho + \frac{\log((1-q)^{-1})}{d}} \sqrt{\frac{d}{N}} \tag{21}
\]

holds with probability at least \( q \).

**Proof.** We consider the sets \( E_\mu, \ldots, E_K \), and \( E \) from Lemma 3.4. Then Theorem 2.5 and Sterling’s formula imply

\[
|\Gamma_k| \leq 2 \frac{1}{\sqrt{2\pi d}} b_d e^d (2^k + 1)^d
\]

with \( b_d := \max \left( 1, 1.1^{d-101} \right) \). Defining \( \sigma := \mu - \log(2^\mu + 1) - 1 - \frac{\log(b_d)}{d} - \frac{\log(2)}{d} \) and (18) yield

\[
\mathbb{P}(E^c_\mu) \leq \sqrt{\frac{2}{\pi d}} e^{-(2c_0^2 - \rho - \mu + \sigma)d}.
\]

Similarly, we get from (19) the inequality

\[
\mathbb{P}(E^c_k) \leq 2 \sqrt{\frac{2}{\pi d}} b_d e^d k d (1 + 2^{-k})^d \rho e^d \exp \left(-2c_0^2 \tau_\mu (k - 1) d\right)
\]

\[
\leq \sqrt{\frac{2}{\pi d}} e^{(1+\xi + \rho)d} \exp \left(-2c_0^2 \tau_\mu (k - 1) d\right)
\]

17
for \( \mu + 1 \leq k \leq K \), where \( \zeta = \log(1 + 2^{-\mu - 1}) + \log(2) + \frac{\log(2)}{d} + \frac{\log(\delta_k)}{d^3} \). Using \( P(E) = 1 - P(E^c) \geq 1 - \sum_{k=\mu}^{K} P(E_k^c) \) it follows that

\[
P(E) \geq 1 - \sqrt{\frac{2}{\pi d}} e^{-(2e_0^2 - \rho - \mu + \sigma)d} \left( 1 + e^{(1 + \zeta + \rho)d} \sum_{j=0}^{K-\mu-1} e^{-(2e_0^2\tau_\mu - \log(2))jKd} \right) \geq 1 - \sqrt{\frac{2}{\pi d}} e^{-(2e_0^2 - \rho - \mu + \sigma)d} \left( 1 + \frac{e^{-(2e_0^2(\mu_\tau - 1) + (1 - \log(2))\mu - 1 - \zeta - \sigma)d}}{1 - e^{-(2e_0^2\tau_\mu - \log(2))d}} \right).
\]

If \( c_0 \geq \sqrt{(\mu + \rho - \sigma)/2} \), then we obtain

\[
P(E) \geq 1 - \sqrt{\frac{2}{\pi d}} e^{-(2e_0^2 - \rho - \mu + \sigma)d} \left( 1 + \frac{e^{-(\mu + \rho - \sigma)(\mu_\tau - 1) + (1 - \log(2))\mu - 1 - \zeta - \sigma)d}}{1 - e^{-(\mu + \rho - \sigma)\tau_\mu - \log(2))d}} \right).
\]

We find that

\[
1 + \frac{e^{-(\mu + \rho - \sigma)(\mu_\tau - 1) + (1 - \log(2))\mu - 1 - \zeta - \sigma)d}}{1 - e^{-(\mu + \rho - \sigma)\tau_\mu - \log(2))d}} \leq \sqrt{\frac{\pi d}{2}} \tag{22}
\]

holds for \( \mu = 12 \) and \( \tau_\mu = 0.0871 \). This can be checked for \( d = 2 \) and \( \rho = 0 \) by a computer calculation. As the left hand side of (22) is monotonic decreasing in \( d \) and in \( \rho \) and the right hand side is monotonic increasing in \( d \), the general case follows. By inserting these explicit values into the formulas for \( c_\mu \) and \( c_1 \) and relate \( c \) and \( c_0 \) via

\[
c = c_0 \left( 1 + c_1 c_\mu \sqrt{\frac{e}{2}} \right),
\]

we obtain the explicit lower bound for the probability. The minimal possible value for \( c \) is taken if \( c_0 = \sqrt{(\mu - \sigma)/2} \) and hence estimate (20) holds only if \( c > 2.4968 \). Estimate (21) follows by putting \( q := 1 - e^{-(1.67681c^2 - 10.45292 - \rho)d} \) and solving for \( c \).

**Remark 3.6.** Note that also the probability \( 1 - e^{-(1.67681c^2 - 10.45292 - \rho)d} \) is positive if \( c > 2.4968 \).

In the special case of Monte Carlo points \( X_1, X_2, \ldots, X_N \) we have \( \rho = 0 \) and our bound (21) improves the result from [AH14], Theorem 1, and [GH19], Theorem 4.4. Furthermore, the smaller constant in (23) (in comparison to 2.5287 from [GH19], Corollary 4.5) can immediately be deduced from Theorem 3.5. Corollary 3.7 improves on the result from [AH14], table at the bottom of p. 1374, and [GH19], Corollary 4.4.
Corollary 3.7. Let $d, N \in \mathbb{N}$. Then there exists a point set $P \subset [0,1)^d$ of $X$ such that

$$D_N^*(P) \leq 2.4968 \sqrt{\frac{d}{N}}. \quad (23)$$

Let $d \geq 2$ and let $X = (X_n)_{n=1}^N$ be a Monte Carlo point set uniformly distributed in $[0,1)^d$. The probability that a realization $P$ of $X$ satisfies

$$D_N^*(P) \leq 2.5 \sqrt{\frac{d}{N}}, \quad \text{and} \quad D_N^*(P) \leq 3 \sqrt{\frac{d}{N}}$$

is at least 0.0528 and 0.9999, respectively.

Proof. In the case $d = 1$ the set $P := \{1/2N, 3/2N, \ldots, (2N - 1)/2N\}$ satisfies $D_N^*(P) = 1/2N$, which is even a stronger result than (23). The case $d \geq 2$ can be solved with Theorem 3.5. We know that

$$D_N^*(X) \leq c \sqrt{\frac{d}{N}}$$

holds with probability $q$ which is at least

$$1 - e^{-(1.67681c^2 - 10.45292) \cdot d},$$

and we want to show that the last term is strictly positive. Since the exponential is decreasing in $d$, it suffices to consider the case $d = 2$. Solving the inequality for $c$ yields the first estimate, namely

$$c > \sqrt{\frac{10.45292}{1.67681}} \geq 2.49676 \ldots.$$ 

The remaining probabilities can be found if we use $c = 2.5$ and $c = 3$ respectively, i.e.

$$q \geq 1 - e^{-(1.67681 \cdot 2.5^2 - 10.45292) \cdot 2} \geq 0.0528$$

and

$$q \geq 1 - e^{-(1.67681 \cdot 3^2 - 10.45292) \cdot 2} \geq 0.9999.$$

In fact, Corollary 3.7 indicates that the realization of a Monte Carlo point set typically has star-discrepancy $D_N^*(P) \leq c \sqrt{\frac{d}{N}}$. Next we show that this is not only the typical case in terms of probability but also in expectation. We split the corresponding statement into two separate parts.

Lemma 3.8. Let $d, N \in \mathbb{N}$ and let $X = (X_n)_{n=1}^N$ be a sampling scheme in $[0,1]^d$. Then

$$E[D_N^*(X)] = \left( \int_0^\infty \sqrt{\frac{N/d}{d}} \mathbb{P} \left( D_N^*(X) \geq \tau \sqrt{d/N} \right) d\tau \right) \sqrt{d/N}.$$
Proof. The equation
\[ \mathbb{E}[D^*_N(X)] = \int_0^1 P(D^*_N(X) \geq t)dt \]
is a standard result from probability theory and an immediate consequence of Fubini’s Theorem. The result then follows from transformation of variables. \( \square \)

**Proposition 3.9.** Let \( X = (X^*_n)_{n=1}^N \) be a sampling scheme in \([0,1]^d\) and \( \alpha, \beta \geq 0 \) such that for all \( c > 0 \)
\[ P \left( D^*_N(X) \leq c\sqrt{\frac{d}{N}} \right) \geq 1 - e^{-\left(\alpha c^2 - \beta\right)d}. \]
Then
\[ \mathbb{E}[D^*_N(X)] \leq \sqrt{\frac{\beta}{\alpha}} \left( 1 + \frac{\pi}{\beta d} e^{\beta d} \left( \Phi \left( \sqrt{2\alpha N} \right) - \Phi \left( \sqrt{2\beta d} \right) \right) \right) \sqrt{\frac{d}{N}} \]
\[ \leq \sqrt{\frac{\beta}{\alpha}} \left( 1 + \frac{1}{2\beta d} \right) \sqrt{\frac{d}{N},} \]
where \( \Phi(\cdot) \) denotes the cumulative distribution function of the standard normal distribution.

Proof. Lemma 3.8 implies
\[ \mathbb{E}[D^*_N(X)] = \int_0^{\sqrt{N/d}} P \left( D^*_N(X) \geq \tau \sqrt{d/N} \right) d\tau \sqrt{d/N}. \]
By assumption it holds that
\[ P \left( D^*_N(X) \geq \tau \sqrt{d/N} \right) \leq \min \left( 1, e^{-\left(\alpha \tau^2 - \beta\right)d} \right) \]
and \( e^{-\left(\alpha \tau^2 - \beta\right)d} < 1 \) if \( \tau > \sqrt{\beta/\alpha} \). Thus
\[ \mathbb{E}[D^*_N(X)] < \left( \int_0^{\sqrt{\beta/\alpha}} d\tau + \int_{\sqrt{\beta/\alpha}}^{\sqrt{N/d}} e^{-\left(\alpha \tau^2 - \beta\right)d} d\tau \right) \sqrt{d/N}. \]
The first integral is equal to \( \sqrt{\beta/\alpha} \). Applying the transformation \( x^2 = 2\alpha \tau^2d \), the second integral can be evaluated as follows:
\[ \int_{\sqrt{\beta/\alpha}}^{\sqrt{N/d}} e^{-\left(\alpha \tau^2 - \beta\right)d} d\tau = e^{\beta d} \int_{\sqrt{\beta/\alpha}}^{\sqrt{N/d}} e^{-\alpha \tau^2 d} d\tau \]
\[ = e^{\beta d} \sqrt{2\alpha d} \int_{\sqrt{2\alpha N}}^{\sqrt{2\beta d}} e^{-\frac{x^2}{2\alpha d}} dx \]
\[ = \sqrt{\frac{\pi}{\alpha d}} e^{\beta d} \left( \Phi \left( \sqrt{2\alpha N} \right) - \Phi \left( \sqrt{2\beta d} \right) \right). \]
Because
\[
\Phi\left(\sqrt{2\alpha N}\right) - \Phi\left(\sqrt{2\beta d}\right) \leq \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\beta d}}^{\infty} e^{-\frac{x^2}{2}} \, dx \leq \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\beta d}}^{\infty} \left(1 + \frac{1}{x^2}\right) e^{-\frac{x^2}{2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \left(-\frac{1}{x}\right) e^{-\frac{1}{2x^2}} \bigg|_{\sqrt{2\beta d}}^{\infty} = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\beta d}} e^{-\beta d}
\]
also the second inequality follows. \[\square\]

In explicit numerical terms, we finally get the following inequalities for the expected value:

**Corollary 3.10.** Let \(d, N \in \mathbb{N}\) and let \(X = (X_n)_{n=1}^{N}\) be a Monte Carlo point set uniformly distributed in \([0, 1]^d\). Then

\[
\mathbb{E}[D_N^* (X)] \leq 2.55648 \sqrt{\frac{d}{N}}
\]

for \(d = 2\) and

\[
\mathbb{E}[D_N^* (X)] \leq 2.53657 \sqrt{\frac{d}{N}}
\]

for all \(d \geq 3\).

**Remark 3.11.** Notice that \(\sqrt{\beta/\alpha} = 2.49676\ldots\) and thus the quantity \(c_d\) in the inequality \(\mathbb{E}[D_N^* (X)] \leq c_d \sqrt{d/N}\) in Proposition 3.9 converges to the same constant \(c\) as in Corollary 3.7 when \(d\) goes to \(\infty\).

From Corollary 3.10, Theorem 3.5 and from the estimates (3) and (4) established in [Doe14] we immediately obtain the following corollary.

**Corollary 3.12.** Let \(d, N \in \mathbb{N}\) and let \(X = (X_n)_{n=1}^{N}\) be a Monte Carlo point set uniformly distributed in \([0, 1]^d\). Then there exist constants \(K_1, K_2 > 0\), independent of \(d\) and \(N\), such that

\[
K_1 \sqrt{\frac{d}{N}} \leq \mathbb{E}[D_N^* (X)] \leq K_2 \sqrt{\frac{d}{N}}
\]

and the estimate

\[
K_1 \sqrt{\frac{d}{N}} \leq D_N^* (X) \leq K_2 \sqrt{\frac{d}{N}}
\]

holds with probability at least \(1 - 2e^{-\Theta(d)}\).

**Remark 3.13.** For \(d \geq 2\) the result of Corollary 3.12 holds also for Latin hypercube samples instead of Monte Carlo samples. This follows from Proposition 3.9 and the results proved in [GH19] and [DDG18].
Remark 3.14. Theorem 3.5, Proposition 3.9, and Corollary 3.12 show that the discrepancy of an $N$-point sampling scheme (consisting of i.i.d. or negatively dependent points) in $[0,1]^d$ has typically (on average and with very high probability) a discrepancy of order at most $\sqrt{d/N}$. Now in practice it is desirable to have sampling schemes that are extensible in the number of points $N$ or even in the dimension $d$. The results presented in this paper do not guarantee that the proved discrepancy bounds still hold with high probability if we keep adding random points to our sampling scheme or additional random components to our points to obtain higher-dimensional point sets. Results that take care about these issues were first presented by Dick [Dic07], who proved probabilistic discrepancy bounds for infinite-dimensional infinite sequences of independent and uniformly distributed random points; these results were slightly improved in [DGKP08]. Later, Aistleitner [Ais13] and Aistleitner and Weimar [AW13] substantially improved these bounds and extended them. To provide a specific example, it was proved in [AW13] that with positive probability a random matrix $X \in [0,1]^{N \times N}$ satisfies for all $d \in \mathbb{N}$ and all $N \geq 2$ the discrepancy bound
\[
D^*_N(P_{N,d}) \leq \sqrt{2130 + 308 \log \log(N) \frac{\log(N)}{d}} \sqrt{\frac{d}{N}},
\tag{24}
\]
where $P_{N,d} \subset ([0,1]^d)^N$ denotes the projection of $X$ onto its first $d$ rows and first $N$ columns. Note that compared to (20), an additional term $\sqrt{\log \log(N)/d}$ appears in (24). This cannot be avoided: although the typical discrepancy of random points is $\sqrt{d/N}$, the law of the iterated logarithm states that for fixed $d$ we have
\[
\limsup_{N \to \infty} \frac{\sqrt{N} D^*_N(P_{N,d})}{\log \log(N)} \geq \frac{1}{\sqrt{2}}.
\]
Note also that the constants in (24) are rather large. We are convinced that we are able to improve the constants in the bounds from [AW13] (and probably also from [Ais13]) reasonably by using our improved bounds for $\delta$-covers from Theorem 2.5 and some of the tweaks used in the proofs in [GH19]. But since this would require additional effort and this paper is already rather long, we leave this as a potential topic of future research (cf. also Section 4).

Since proving that a given sampling scheme is $\mathcal{D}_0^d-e^{\rho d}$-negatively dependent is in some cases a difficult task, we now consider the relaxed condition of $\mathcal{C}_0^d-e^{\rho d}$-negative dependence. With the help of Theorem 2.5 we can also improve the existing bounds in that case ([WGH20], Theorem 3.1).

Corollary 3.15. Let $d, N \in \mathbb{N}$ and $\rho \in [0,\infty)$. Let $X = (X_j)_{j=1}^N$ be a $\mathcal{C}_0^d-e^{\rho d}$-negatively dependent sampling scheme in $[0,1]^d$. Then for every $c > 0$
\[
D^*_N(X) \leq c \sqrt{\frac{d}{N} \max \left(1, \log \left(\frac{N}{d}\right)\right)} \tag{25}
\]
holds with probability at least $1 - 2\exp \left( \left( -\frac{1}{2}(c^2 - 1)\xi + \rho + \log(e(c + 1)) + \frac{\log(b_d)}{d} \right) \right)$, where $\xi := \max \left( 1, \log \left( \frac{N}{d} \right) \right)$ and $b_d := \max(1, \frac{1}{d^{d-101}})$. Moreover, for every $\theta \in (0, 1)$,

$$P \left( D_N^*(X) \leq \sqrt{\frac{2}{N}} \sqrt{d\log(\eta) + \rho d + \log \left( \frac{2}{1 - \theta} \right)} \right) \geq \theta,$$

(26)

where

$$\eta := \eta(N, d) = 3.3e \left( \max \left( 1, \frac{N}{2d\log(3.3e)} \right) \right)^{\frac{1}{2}}.$$

Proof. The proof relies on an intermediate result of Theorem 3.1 in [WGH20], namely the inequality

$$P(D_N^*(X) < 2\delta) \geq 1 - 2e^{\rho \delta |\Gamma|} e^{-2N\delta^2}$$

(27)

that holds for all $\delta > 0$; here $\Gamma$ is the cardinality of an arbitrary $\delta$-cover. Due to Theorem 2.5 and Sterling’s formula we have

$$|\Gamma| \leq N(d, \delta) \leq 2N[1](d, \delta) < \max(1, \frac{1}{d^{d-101}}, 1) e^d(\delta^{-1} + 1)^d,$$

since

$$\frac{2d^d}{d!} \leq 2 \cdot \frac{e^d}{\sqrt{2\pi d}} < e^d$$

for all $d \in \mathbb{N}$. Now, we deduce (26) from (27) and follow exactly the lines of the proof of Theorem 3.1 in [WGH20]; the only exception is that we end up with the condition

$$\left( \frac{\eta}{b_d e} - 1 \right) \log(\eta) \frac{1}{2} > \left( \frac{2N}{d} \right)^{\frac{1}{2}},$$

which is fulfilled by the $\eta$ in the statement of the result. Then, we apply (27) with $2\delta = c\sqrt{\frac{\eta}{b_d e}}$ to prove (26).

3.3 Bounds for the Weighted Star-Discrepancy

Finally, we also treat the case of the weighted star-discrepancy and improve the quantitative bounds from [WGH20], Theorem 3.4, and [Ais14], Theorem 1. For that purpose, we shortly recall the definition of weighted star-discrepancy as introduced in [SW98]: The weights are a collection of non-negative numbers $\gamma = (\gamma_u)_{\emptyset \neq u \subseteq [d]}$, where $\gamma_u$ is interpreted as the weight of the coordinates from the subset $u$. Moreover, recall from our presentation of the Koksma-Hlawka inequality that $(x_u, 1)$ denotes the point in $[0, 1]^d$ whose $i$-th component is $x_i$ if $i \in u$ and 1 otherwise. The weighted star-discrepancy of a $N$-point set $P$ and weights $\gamma$ is defined by

$$D_{N, \gamma}^*(P) := \sup_{z \in [0, 1]^d} \max_{\emptyset \neq u \subseteq [d]} \gamma_u |D_N(P, (z_u, 1))|.$$

More information about the weighted star discrepancy can, e.g., be found in [Ais14, HPS08]. The following proposition leads to an improvement of [Ais14], Theorem 1, see Corollary 3.17.
Proposition 3.16. Let $d, N \in \mathbb{N}$ and $\alpha, \beta \in [0, \infty)$. Let $X = (X_j)_{j=1}^N \subset [0, 1)^d$ be a family of random points such that for every $\emptyset \neq u \subseteq [d]$ its projection $X^u$ onto the coordinates in $u$ satisfies for every $q \in (0, 1)$

$$D_N^*(X^u) \leq \alpha \sqrt{\beta + \frac{\log((1 - q)^{-1})}{|u|}} \sqrt{\frac{|u|}{N}} \tag{28}$$

with probability at least $q$. Then there exists a realization $P \subset [0, 1]^d$ of $X$ such that

$$D_N^{*\gamma}(X) \leq \frac{\alpha}{\sqrt{N}} \max_{\emptyset \neq u \subseteq [d]} \gamma_u \sqrt{\beta + 1 + \log \left(1 + \frac{1}{\sqrt{2\pi}}\right) + \log(d) - \left(1 + \frac{1}{2|u|}\right) \log(|u|) \sqrt{|u|}}. \tag{29}$$

Proof. For $\emptyset \neq u \subseteq d$ we put

$$q := 1 - e^{-|u| \left( \log \left( \frac{ed}{|u|} \right) + \log \left(1 + \frac{1}{\sqrt{2\pi}}\right)\right) + \frac{1}{2} \log(|u|)}$$

and

$$A_u := \left\{ D_N^*(X^u) > \alpha \sqrt{\beta + \frac{\log((1 - q)^{-1})}{|u|}} \sqrt{\frac{|u|}{N}} \right\}.$$

To prove Proposition 3.16 it suffices to show that $P \left( \bigcup_{\emptyset \neq u \subseteq [d]} A_u \right) < 1$. Indeed, using

$$\left( \frac{d}{r} \right) \leq \frac{1}{\sqrt{2\pi r}} \left( \frac{ed}{r} \right)^r$$

for every $1 \leq r \leq d$,

we obtain due to (28)

$$\mathbb{P} \left( \bigcup_{\emptyset \neq u \subseteq [d]} A_u \right) \leq \sum_{r=1}^d \sum_{|u|=r} \mathbb{P}(A_u) \leq \sum_{r=1}^d \left( \frac{d}{r} \right) e^{-r \left( \log \left( \frac{ed}{r} \right) + \log \left(1 + \frac{1}{\sqrt{2\pi}}\right)\right) + \frac{1}{2} \log(r)} \leq \frac{1}{\sqrt{2\pi}} \sum_{r=1}^d e^{-r \log \left(1 + \frac{1}{\sqrt{2\pi}}\right)} < \frac{1}{\sqrt{2\pi}} \left( \frac{1}{1 - \left(1 + \frac{1}{\sqrt{2\pi}}\right)^{-1}} - 1 \right) = 1.$$

We apply Proposition 3.16 to Monte Carlo point sets. Then Theorem 3.5 with $\rho = 0$ gives us immediately the following corollary.

Corollary 3.17. There exists a point set $P \subset [0, 1)^d$ such that

$$D_N^{*\gamma}(X) \leq \frac{0.7723}{\sqrt{N}} \max_{\emptyset \neq u \subseteq [d]} \gamma_u \sqrt{11.78864 + \log(d) - \left(1 + \frac{1}{2|u|}\right) \log(|u|) \sqrt{|u|}}. \tag{30}$$
Finally, we improve \cite{WGH20}, Theorem 3.4.

**Corollary 3.18.** Let $d, N \in \mathbb{N}$ and let $X = (X_j)_{j=1}^N \subset [0, 1)^d$ be a sampling scheme such that for every $\emptyset \neq u \subseteq [d]$ its projection onto the coordinates in $u$ is $D_{0}^{[u]} - e^{\rho |u|}$-negatively dependent. Then for any weights $(\gamma_u)_{\emptyset \neq u \subseteq [d]}$ and any $c > 0$ we obtain

$$D_{N, \gamma}^*(X) \leq c \max_{\emptyset \neq u \subseteq [d]} \gamma_u \sqrt{|u|}$$

with probability at least $2 - (1 + e^{-(1.67681c^2 - 10.45292-\rho)})^d$. For $\theta \in (0, 1)$ we have

$$P \left( D_{N, \gamma}^*(X) \leq \sqrt{\rho + 10.45292 - \log((2 - \theta)^{1/d} - 1)} \cdot \max_{\emptyset \neq u \subseteq [d]} \gamma_u \sqrt{|u|} \right) \geq \theta. \quad (31)$$

The proof can be done exactly along the lines of the one of \cite{WGH20}, Theorem 3.4.

**4 Future Research**

In this last short section we would like to mention potential future research. It would be interesting to further improve our upper bounds on the bracketing number and to provide explicit bracketing and $\delta$-covers that satisfy those bounds and that can be efficiently computed.

Clearly, with our bracketing result Theorem \ref{thm:bracketing} or even additional refinements one may, similar as in Theorem \ref{thm:disp_err} further improve other probabilistic discrepancy or dispersion bounds and extend them from Monte Carlo point sets to the larger class of $\gamma$-negatively-dependent random point sets. For instance, by using Theorem \ref{thm:bracketing} and the dyadic chaining method from \cite{Ais11} one may substantially improve the results for the extreme discrepancy from \cite[Theorem 2.2]{Gne08a}. Similarly, using Theorem \ref{thm:bracketing} and some additional tweaks from the proofs in \cite{GH19} should lead to much smaller constants in the discrepancy bounds for the (infinite-dimensional infinite) sequences of points presented in \cite{Ais14, AW13}. Such an improvement would be valuable in practice, since for numerical applications it is favorable to have sequences of sample points instead of sheer sample sets that are not extensible in the number of points, cf. Remark \ref{rem:multilevel} and the remarks on multilevel algorithms below.

Moreover, new discrepancy bounds and better $\delta$-covers may be used to improve algorithms for calculating discrepancies and for constructing point sets and sequences with small discrepancies, cf. \cite{DGW09, DGW10, DGW14, GGP20, Thi01}. It would also be interesting to investigate whether good estimates for the dispersion in combination with better $\delta$-covers lead to competitive algorithms for calculating the dispersion of given point sets or to construct point sets with low dispersion (cf. also \cite{UV19}). Note that algorithms
that compute the discrepancy (or dispersion) of given point sets may, in particular, be used for a “semi-construction” of low-discrepancy (or low-dispersion) sets: One may randomly generate a point set that has low discrepancy (or dispersion) with high probability and compute afterwards its discrepancy (or dispersion). If it is small enough, one accepts the point set, otherwise one rejects the point set and generates a new one, and so on.

A more ambitious goal is to verify that certain structure-preserving randomizations of low-discrepancy point sets are $\gamma$-negatively dependent and therefore satisfy the probabilistic pre-asymptotic discrepancy bounds from Theorem 3.5 with high probability. Due to the structure-preserving randomization, these points then have essentially optimal asymptotic discrepancy bounds for sure. Such point sets would be extremely useful for numerical integration in high or infinite-dimensions. A particularly thrilling perspective would be to use them as integration nodes of randomized quasi-Monte Carlo algorithms that serve as building blocks for sophisticated adaptive multilevel algorithms (cf., e.g., [BG14, GW09, HMNR10]): the resulting building block algorithms would from a probabilistic point of view never be worse than pure Monte Carlo algorithms (even for levels of high dimension where only a small number of samples is used), but on levels of lower dimension, where a large number of samples is used, their superior asymptotic behavior would automatically kick in. Note that we are referring to adaptive algorithms, i.e. the meaning of low and high dimension as well as of small and large numbers of samples are relative. They may change while the algorithm is running and investing more and more resources and introducing more and more levels. That means, we cannot say beforehand whether the building block algorithms for a specific level should rather exhibit a good asymptotic behavior or preferably a good pre-asymptotic behavior. That is why building block algorithms that satisfy good asymptotic as well as pre-asymptotic bounds are highly desirable.

Acknowledgment The authors would like to thank Friedrich Pillichshammer for his question which motivated Corollary 3.10 and the anonymous referees for their valuable comments. Furthermore, the first-named and the third-named author would like to thank RICAM in Linz for their hospitality where they got to know each other during the special semester on Multivariate Algorithms and their Foundations in Number Theory.

References

[Ais11] C. Aistleitner. Covering numbers, dyadic chaining and discrepancy. J. Complexity 27 (2011), 531–540.

[Ais13] C. Aistleitner. On the inverse of the discrepancy for infinite-dimensional infinite sequences. J. Complexity 29 (2013), 182–194.

[Ais14] C. Aistleitner. Tractability results for the weighted star discrepancy. J. Complexity 30 (2014), 381–391.
[DGW10] B. Doerr, M. Gnewuch, M. Wahlstr¨ om. Algorithmic construction of low-discrepancy point sets via dependent randomized rounding, J. Complexity 26 (2010), 490–507.

[DGW14] C. Doerr, M. Gnewuch, M. Wahlstr¨ om. Calculation of Discrepancy Measures and Applications. In: W. W. L. Chen, A. Srivastav, G. Travaglini (Eds.), Panorama of Discrepancy Theory, Lecture Notes in Mathematics 2107, 621-678, Springer-Verlag, 2014.

[GW09] M. B. Giles, B. J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. Radon Ser. Comput. Appl. Math. 8 (2009), 165–181.

[Gla03] P. Glasserman. Monte Carlo Methods in Financial Engineering, Springer (2003).

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

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

[Gne12] M. Gnewuch. Entropy, Randomization, Derandomization, and Discrepancy. In: L. Plaskota, H. Woźniakowski (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing 2010, Springer-Verlag, 2012.

[GH19] M. Gnewuch, N. Hebbinghaus. Discrepancy bounds for a class of negatively dependent random points including Latin hypercube samples. Preprint 2019. (To appear in Annals of Applied Probability; arXiv:2102.04451)

[GGP20] A. Gómez, D. Gómez-Pérez, F. Pillichshammer, Secure pseudorandom bit generators and point sets with low-discrepancy, arXiv:2004.14158 (2020).

[Heb12] N. Hebbinghaus, Mixed sequences and application to multilevel algorithms, Master’s thesis, Christ Church, University of Oxford (2012).

[HNUW01] S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski. The inverse of the star-discrepancy depends linearly on the dimension, Acta Arith. 96 (2001).

[HMNR10] F. J. Hickernell, T. Müller-Gronbach, B. Niu, K. Ritter, Multi-level Monte Carlo algorithms for infinite-dimensional integration on $\mathbb{R}^n$. J. Complexity 26 (2010), 229–254.

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

[HKKR20] A. Hinrichs, D. Krieg, R. Kunsch, D. Rudolf. Expected dispersion of uniformly distributed points, J. Complexity 61 (2020), 101483, 9 pp.

[HPS08] A. Hinrichs, F. Pillichshammer, W. C. Schmid. Tractability properties of the weighted star discrepancy, J. Complexity 24 (2008), 134–143.

[Knu93] D. E. Knuth. Johann Faulhaber and sums of powers, Mathematics of Computation, 61 (203), (1993), 277–294.

[KN74] L. Kuipers, H. Niederreiter. Uniform distribution of sequences, John Wiley & Sons, New York (1974).

[L’Ec17] P. L’Ecuyer. Randomized Quasi-Monte Carlo: An Introduction for Practitioners. 12th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2016), Aug 2016, Stanford, United States
[Len18] C. Lemieux. Negative dependence, scrambled nets, and variance bounds, Mathematics of Operations Research, 43 (2018), 228–251.

[Lev19] C. Levaillant, Wilson’s theorem modulo $p^2$ derived from Faulhaber polynomials, arXiv: 1912.06652 (2019).

[Mat10] J. Matoušek. Geometric Discrepancy. Springer-Verlag Berlin Heidelberg, revised second printing, 2010.

[NP18] M. Neumüller, F. Pillichshammer. Metrical star discrepancy bounds for lacunary subsequences of digital Kronecker-sequences and polynomial tractability, Unif. Distrib. Theory, 13 (2018), 65–86.

[Nie92] H. Niederreiter. Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics 63, SIAM, Philadelphia, 1992.

[Owe95] A. B. Owen. Randomly permuted $(t, m, s)$-nets and $(t, s)$-sequences. In: H. Niederreiter, and P. J.-S. Shiue (eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, 299–317, Springer, New York, 1995.

[PW20] H. Pasing, C. Weiß. Improving a constant in high-dimensional discrepancy estimates., Publications de l’Institut Mathématique, 107, 67–74 (2020).

[Pas94] S. H. Paskov. Computing high dimensional integrals with applications to finance, Columbia University. Technical Report CUCS-023–94 (1994).

[Rud18] D. Rudolf. An upper bound of the minimal dispersion via delta covers, in: Dick, J., Kuo, F. Y., Woźniakowski, H. (Eds.), Contemporary Computational Mathematics – A Celebration of the 80th Birthday of Ian Sloan, 1099-1108, Springer, 2018.

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

[SW98] I. Sloan, H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, J. Complexity, 14 (1998), 1–33.

[Thi01] E. Thiémard. An algorithm to compute bounds for the star discrepancy, J. Complexity, 17 (2001), 850–880.

[UV19] M. Ullrich, J. Vybíral. Deterministic constructions of high-dimensional sets with small dispersion, preprint (2019), arXiv:1901.06702

[vdVW96] A. W. van der Vaart, J. A. Wellner. Weak Convergence and Empirical Process Theory. Springer Series in Statistics, Springer, New York, (1996).

[WN19] C. Weiß, Z. Nikolic. An Aspect of Optimal Regression Design for LSMC. Monte Carlo Methods and Applications, 4 (2019), 283–290.

[WLD19] J. Wiart, C. Lemieux, G. Dong. On the dependence structure of scrambled $(t, m, s)$-nets, Preprint (2019), arXiv:1903.09877

[WGH20] M. Wnuk, M. Gnewuch, N. Hebbinghaus. On Negatively Dependent Sampling Schemes, Variance Reduction, and Probabilistic Upper Discrepancy Bounds. In: D. Bylik, J. Dick, F. Pillichshammer (Eds.), Discrepancy Theory, Radon Series on Computational and Applied Mathematics 26, DeGruyter, 2020, pp. 43–68. (arXiv:1904.10796)

[WG20] M. Wnuk, M. Gnewuch. Note on pairwise negative dependence of randomly shifted and jittered rank-1 lattices, Operations Research Letters 48 (2020), 410–414. (arXiv:1903.02261)