A tight lower bound on the minimal dispersion

M. Trödler[*][†]   J. Volec[*][‡]   and J. Vybíral[*][§]

Abstract

We give a new lower bound for the minimal dispersion of a point set in the unit cube and its inverse function in the high dimension regime. This is done by considering only a very small class of test boxes, which allows us to reduce bounding the dispersion to a problem in extremal set theory. Specifically, we translate a lower bound on the size of r-cover-free families to a lower bound on the inverse of the minimal dispersion of a point set. The lower bound we obtain matches the recently obtained upper bound on the minimal dispersion up to logarithmic terms.

1 Introduction and the Main Result

For a given subset of the d-dimensional unit cube $X \subset [0,1]^d$, there are various different ways how to measure whether X is well spread over the unit cube. Based on the previous work of Hlawka [7] and Niederreiter [14], Rote and Tichy [15] introduced in 1996 a concept called the dispersion of X. The *dispersion of X* is the volume of the largest axis-parallel box in $[0,1]^d$, which does not intersect X, i.e.,

$$\mathrm{disp}(X) = \sup_{B : B \cap X = \emptyset} |B|,$$

where the supremum runs over all the boxes $B = \prod_{i=1}^{d}(a_i, b_i)$, where $0 \leq a_i < b_i \leq 1$ for all $i \in \{1, \ldots, d\}$, and $|B|$ stands for the volume of B.

Intuitively, the smaller the dispersion of a point set is, the better spread over the unit cube the points have. We are therefore interested in point sets that have simultaneously small cardinality and small dispersion.

Given $n \in \mathbb{N}$, we denote the minimal dispersion of a point set that has cardinality n by

$$\mathrm{disp}^*(n, d) = \inf_{\substack{X \subset [0,1]^d \\ \#X = n}} \mathrm{disp}(X).$$

For a fixed $d \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, we define the inverse function of the minimal dispersion in $[0, 1]^d$ by

$$N(\varepsilon, d) = \min\{n \in \mathbb{N} : \mathrm{disp}^*(n, d) \leq \varepsilon\}$$
$$= \min\{n \in \mathbb{N} : \exists X \subset [0, 1]^d \text{ with } \#X = n \text{ and } \mathrm{disp}(X) \leq \varepsilon\}.$$

The study of the minimal dispersion $\mathrm{disp}^*(n, d)$ as well as its inverse function $N(\varepsilon, d)$ recently attracted a remarkable attention in the mathematical community. The pigeonhole principle implies the trivial lower bound $\mathrm{disp}^*(n, d) \geq \frac{1}{n+1}$, which translates into

$$N(\varepsilon, d) \geq \frac{1}{\varepsilon} - 1.$$

This bound was complemented in [15] and then further improved by Bukh and Chao [4] to

$$N(\varepsilon, d) \leq C \cdot \frac{d^2 \log d}{\varepsilon}. \tag{1}$$

Let us note that further upper and lower bounds for different regimes of ε and d were obtained also in [5, 11, 12, 13, 16], and the dispersion of certain specific sets was studied in [3, 6, 8, 9, 10, 19, 20].

The first non-trivial lower bound on $N(\varepsilon, d)$ that exhibits a growth if d increases was achieved by Aistleitner, Hinrichs and Rudolf [1]. Specifically, they showed that

$$N(\varepsilon, d) \geq \frac{\log_2 d}{8\varepsilon} \tag{2}$$

for every $d \geq 2$ and every $0 < \varepsilon < 1/4$. The results mentioned so far suggest that $N(\varepsilon, d)$ depends linearly on $1/\varepsilon$ and it only remains to determine the d-dependence.

In another line of research, Sosnovec [18] showed that $N(\varepsilon, d) \leq C_\varepsilon \log d$. The dependence of C_ε was not optimized in [18], and a brief inspection of the proof reveals that it is actually super-exponential in $1/\varepsilon$. Nevertheless, it matches (2) when it comes to the dependence on d. Let us mention that

2

the point set with small dispersion was generated in [18] by sampling the coordinates of the points independently at random from a uniform grid in $[0, 1]$. An improved analysis of the same method was given in [21], which together with a further improvement due to Litvak [11] yields

$$N(\varepsilon, d) \leq C \cdot \frac{\log d \cdot \log(\frac{1}{\varepsilon})}{\varepsilon^2}. \tag{3}$$

The comparison between (1) and (3) is not easy. Depending on the relation between d and $1/\varepsilon$, one or the other might perform better. Although the analysis in [11, 21] regarding the dependence on $1/\varepsilon$ was much finer than the one in [18], it was quite natural to assume that the second power of $1/\varepsilon$ in (3) is an artefact of the proof method and that the bound could be potentially further improved even when ε is moderately large.

The main result of this work is to show that, surprisingly, this is not the case and the second power of $1/\varepsilon$ in (3) is optimal when ε is sufficiently large.

Theorem 1. *There is an absolute constant $c > 0$, such that the following statement is true. For any integer $d \geq 2$ and any real ε satisfying $\frac{1}{4} \geq \varepsilon \geq \frac{1}{4\sqrt{d}}$, it holds that*

$$N(\varepsilon, d) > \frac{c \log d}{\varepsilon^2 \cdot \log \frac{1}{\varepsilon}}. \tag{4}$$

The proof of this result is motivated by the insight gained by a detailed inspection of the proof given in [21]. For a fixed box B, the sampling from a finite grid, which is bounded away of zero and one, naturally splits the individual coordinates into two groups — those, where the corresponding interval (a_i, b_i) is large enough and covers the whole grid, and the coordinates, where this is not the case. It turned out in [21] that the boxes with a large number of the coordinates inside the second group are the most difficult to hit by a randomly generated point. Therefore, we choose a very small class of test boxes when compared to the number of all the axis-parallel boxes, and consider any point set that hits all these test boxes. This naturally reduces the question under study to a well-known combinatorial problem called *r-cover-free systems*, see Section 2 for the definition. The key tool in our proof is a lower bound of Alon and Asodi [2] on the minimum size of r-cover-free system.

Reformulated in the language of minimal dispersion, Theorem 1 states the following.

3

Corollary 2. *There are two absolute constants $c_1, c_2 > 0$ such that for every positive integers d and n with $d \geq 2$ and $2 \log d \leq n \leq c_1 d$ it holds that*

$$\mathrm{disp}^*(n, d) \geq c_2 \left(\frac{\log d}{n}\right)^{1/2} \cdot \left(\log\left(\frac{n}{\log d}\right)\right)^{-1/2}.$$

Before we present our proof of Theorem 1, let us add a few remarks.

Remark.   *1. It follows from the proof of Theorem 1 that one can actually choose $c = 1/1920$, but we did not attempt to optimize this constant.*

2. The essentially tight factor $1/\varepsilon^2$ in the lower bound for $N(\varepsilon, d)$ in (4) is rather surprising. It seems that a logarithmic dependence on d and a linear dependence on $1/\varepsilon$ exclude each other. What might be even more surprising, this lower bound is obtained by considering only a very small class of axis-parallel test boxes.

3. We leave it as an open problem whether the proof method can be further improved, possibly extending (4) to smaller values of ε.

2   Proof of Theorem 1

We start by fixing some notation that will be used in the proof. For a finite set X, we denote its size by $\#X$. For a positive integer d, we denote by $[d]$ the set $\{1, 2, \ldots, d\}$. Additionally, for a non-negative integer k, we write $\binom{[d]}{k}$ to denote the collection of all the k-element subsets of $[d]$. Given a point $x \in [0,1]^d$ and an integer $i \in [d]$, we denote by $(x)_i$ the i-th coordinate of x.

By focusing on a very specific set of axis-parallel boxes of volume at least ε, we are able to translate bounding $N(\varepsilon, d)$ from below to a question in extremal set theory. We say that a family $\mathcal{F}$ of subsets of a ground set X is *r-cover-free* if no $F \in \mathcal{F}$ is contained in the union of any r sets from $\mathcal{F} \setminus \{F\}$. Based on a previous work of Ruszinkó [17], Alon and Asodi [2] proved the following lower-bound on the size of the ground set of an r-cover-free family.

Theorem 3 ([2, Lemma 2.8]). *Let $\mathcal{F}$ be a family of d subsets of a ground set X. If $\mathcal{F}$ is r-cover-free, where $2 \leq r \leq 2\sqrt{d}$, then*

$$\#X > \frac{1}{10} \cdot \frac{r^2 \log\left(d - \frac{r}{2}\right)}{\log r}.$$

We now proceed with the proof of Theorem 1. Fix $d \geq 2$ and ε such that $\frac{1}{4} \geq \varepsilon \geq \frac{1}{4\sqrt{d}}$, and let k be the unique integer satisfying $2^{-(k+1)} < \varepsilon \leq 2^{-k}$. Note that $k \geq 2$.

Our plan is to define a collection $\mathcal{B}$ of $\binom{d}{2^{k-2}}\left(d - 2^{k-2}\right)$ boxes such that each box has one coordinate where all its points are close to zero, while in some other 2^{k-2} coordinates all the points of the box are bounded away from zero. Next, we will sort out the points of X with at least one coordinate close to zero into (not necessarily disjoint) sets $F_1, F_2, \ldots, F_d$. All of this is done in such a way that the sets $F_1, F_2, \ldots, F_d$ are 2^{k-2}-cover-free, thus Theorem 3 yields the desired lower bound on X.

We now give all the details. For a given set $A \subseteq [d]$ and $j \in [d] \setminus A$, let $B^{A,j} \subseteq [0,1]^d$ be the axes-parallel box defined as

$$B^{A,j} := I_1 \times I_2 \times \cdots \times I_d, \quad \text{where} \quad \begin{cases} I_j = (0, 2^{1-k}), \\ I_i = (2^{1-k}, 1) & \text{for } i \in A, \\ I_i = (0, 1) & \text{for } i \in [d] \setminus (A \cup \{j\}). \end{cases}$$

Let $\mathcal{B} := \left\{ B^{A,j} : A \in \binom{[d]}{2^{k-2}} \text{ and } j \in [d] \setminus A \right\}$. The following claim yields that any set of points from $[0,1]^d$ with dispersion at most ε must have at least one point inside every $B \in \mathcal{B}$.

Claim 1. $\left| B^{A,j} \right| \geq \varepsilon$ for any $A \subseteq [d]$ of size at most 2^{k-2} and any $j \in [d] \setminus A$.

Proof. Since 2^{-2x} is convex, it holds that $(1-x) \geq 2^{-2x}$ for every $x \in [0, 1/2]$. Therefore,

$$\left| B^{A,j} \right| = \left(1 - 2^{1-k} \right)^{\#A} \cdot 2^{1-k} \geq \left(2^{-2^{2-k}} \right)^{\#A} \cdot 2^{1-k} \geq 2^{-k} \geq \varepsilon,$$

which finishes the proof of the claim. $\qquad\square$

Fix a finite set of points $X \subseteq [0,1]^d$ that intersects every $B \in \mathcal{B}$. For every $j \in [d]$, we define an auxiliary set F_j of the points with a small j-th coordinate. Specifically, let $F_j := \{x \in X : (x)_j < 2^{1-k}\}$. In order to lower bound $\#X$ using Theorem 3, we establish the following claim.

Claim 2. The family $\{F_1, F_2, \ldots, F_d\}$ is 2^{k-2}-cover-free.

Proof. Suppose there is a set $A \in \binom{[d]}{2^{k-2}}$ and an integer $j \in [d] \setminus A$ such that

$$F_j \subseteq \bigcup_{i \in A} F_i.$$

5

However, there must exist a point $x \in X \cap B^{A,j}$. Since $(x)_j < 2^{1-k}$ and $(x)_i > 2^{1-k}$ for all $i \in A$, we conclude that x is an element of $F_j \setminus \bigcup_{i \in A} F_i$, which is a contradiction. $\square$

If $k = 2$, then $\varepsilon \in (1/8, 1/4]$ and $\#X \geq \log_2(d) > \frac{3}{64} \cdot \frac{\log d}{\varepsilon^2 \cdot \log(1/\varepsilon)}$ by the 1-cover-freeness. For $k \geq 3$, we have that $2^{k-2} \leq \frac{1}{4\varepsilon} \leq \sqrt{d}$, thus Theorem 3 applied to the family $\{F_1, \ldots, F_d\}$ and $r = 2^{k-2}$ readily yields that

$$\#X > \frac{1}{10} \cdot \frac{2^{2k-4} \cdot \log\left(d - 2^{k-3}\right)}{\log\left(2^{k-2}\right)} \geq \frac{1}{10} \cdot \frac{2^{2k-4} \cdot \log\left(d - \sqrt{d}/2\right)}{\log\left(2^{k-2}\right)}.$$

Since $\frac{1}{8\varepsilon} < 2^{k-2} < \frac{1}{\varepsilon}$ and $\log(d - \sqrt{d}/2) \geq \frac{\log d}{3}$ for all $d \geq 2$, we conclude that

$$\#X > \frac{1}{1920} \cdot \frac{\log d}{\varepsilon^2 \cdot \log \frac{1}{\varepsilon}}.$$

This finishes the proof of Theorem 1.

References

[1] C. Aistleitner, A. Hinrichs and D. Rudolf, *On the size of the largest empty box amidst a point set*, Discrete Appl. Math. 230 (2017), 146–150

[2] N. Alon and V. Asodi, *Learning a hidden subgraph*, SIAM J. Discrete Math. 18 (2005), 697–712

[3] S. Breneis and A. Hinrichs, *Fibonacci lattices have minimal dispersion on the two-dimensional torus*, in: Dmitriy Bilyk, Josef Dick, Friedrich Pillichshammer (Eds.), Discrepancy Theory, De Gruyter, 2020, pp. 117–132

[4] B. Bukh and T. Chao, *Empty axis-parallel boxes*, Int. Math. Res. Notices 18 (2022), 13811–13828

[5] A. Dumitrescu and M. Jiang, *On the largest empty axis-parallel box amidst n points*, Algorithmica 66 (2013), 225–248

[6] A. Hinrichs, D. Krieg, R. J. Kunsch, and D. Rudolf, *Expected dispersion of uniformly distributed points*, J. Compl. 61 (2020), 101483

[7] E. Hlawka, *Abschätzung von trigonometrischen Summen mittels diophantischer Approximationen*, Österreich. Akad. Wiss. Math.- Naturwiss. Kl. S.-B. II, 185 (1976), 43–50

[8] D. Krieg, *On the dispersion of sparse grids*, J. Compl. 45 (2018), 115–119

[9] R. Kritzinger, *Dispersion of digital $(0, m, 2)$-nets*, Monatshefte Math. 195 (1) (2021), 155–171

[10] T. Lachmann and J. Wiart, *The area of empty axis-parallel boxes amidst 2-dimensional lattice points*, J. Compl. 76 (2023), 101724

[11] A. E. Litvak, *A remark on the minimal dispersion*, Commun. Contemp. Math. 23 (2021), 2050060

[12] A. E. Litvak and G. V. Livshyts, *New bounds on the minimal dispersion*, J. Compl. 72 (2022), 101648

[13] K. MacKay, *Minimal dispersion of large volume boxes in the cube*, J. Compl. 72 (2022), 101650

[14] H. Niederreiter, *A quasi-Monte Carlo method for the approximate computation of the extreme values of a function*, In: P. Erdös, L. Alpár, G. Halász, A. Sárközy, (eds.), Birkhäuser Basel, Basel (1983), 523-529

[15] G. Rote and R. F. Tichy, *Quasi-Monte Carlo methods and the dispersion of point sequences*, Math. Comput. Modelling 23 (1996), 9–23

[16] D. Rudolf, *An upper bound of the minimal dispersion via delta covers*, Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan, Springer-Verlag (2018), 1099–1108

[17] M. Ruszinkó, *On the upper bound of the size of the r-cover-free families*, J. Combin. Theory Ser. A 66 (1994), 302–310

[18] J. Sosnovec, *A note on the minimal dispersion of point sets in the unit cube*, Eur. J. Comb. 69 (2018), 255–259

[19] V. N. Temlyakov, *Dispersion of the Fibonacci and the Frolov point sets*, preprint, 2017, arXiv:1709.08158

[20] M. Ullrich, *A note on the dispersion of admissible lattices*, Discrete Appl. Math. 257 (2019), 385–387

[21] M. Ullrich and J. Vybíral, *An upper bound on the minimal dispersion*, J. Compl. 45 (2018), 120–126