Combinatorial Discrepancy for Boxes via the γ_2 Norm

Jiří Matoušek*[1] and Aleksandar Nikolov[2]

1 Department of Applied Mathematics
Charles University
Malostranské nám. 25 118 00 Praha 1, Czech Republic, and
Department of Computer Science
ETH Zurich
8092 Zurich, Switzerland
matousek@kam.mff.cuni.cz

2 Microsoft Research
Redmond, WA, USA
alenik@microsoft.com

Abstract

The γ_2 norm of a real $m \times n$ matrix A is the minimum number t such that the column vectors of A are contained in a 0-centered ellipsoid $E \subseteq \mathbb{R}^m$ that in turn is contained in the hypercube $[-t, t]^m$. This classical quantity is polynomial-time computable and was proved by the second author and Talwar to approximate the *hereditary discrepancy* $\operatorname{herdisc} A$ as follows: $\gamma_2(A)/O(\log m) \leq \operatorname{herdisc} A \leq \gamma_2(A) \cdot O(\sqrt{\log m})$. Here we provide a simplified proof of the first inequality and show that both inequalities are asymptotically tight in the worst case.

We then demonstrate on several examples the power of the γ_2 norm as a tool for proving lower and upper bounds in discrepancy theory. Most notably, we prove a new lower bound of $\Omega(\log^{d-1} n)$ for the *d-dimensional Tusnády problem*, asking for the combinatorial discrepancy of an n-point set in $\mathbb{R}^d$ with respect to axis-parallel boxes. For $d > 2$, this improves the previous best lower bound, which was of order approximately $\log^{(d-1)/2} n$, and it comes close to the best known upper bound of $O(\log^{d+1/2} n)$, for which we also obtain a new, very simple proof. Applications to lower bounds for dynamic range searching and lower bounds in differential privacy are given.

1998 ACM Subject Classification G.2.1 Combinatorics, F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases discrepancy theory, range counting, factorization norms

Digital Object Identifier 10.4230/LIPIcs.SOCG.2015.1

1 Introduction

Discrepancy and hereditary discrepancy. Let $V = [n] := \{1, 2, \ldots, n\}$ be a ground set and $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ be a system of subsets of V. The *discrepancy* of $\mathcal{F}$ is

$$\operatorname{disc} \mathcal{F} := \min_{x \in \{-1,1\}^n} \operatorname{disc}(\mathcal{F}, x),$$

where the minimum is over all choices of a vector $x \in \{-1, +1\}^n$ of signs for the points, and $\operatorname{disc}(\mathcal{F}, x) := \max_{i=1,2,\ldots,m} \bigl|\sum_{j \in F_i} x_j\bigr|$. (A vector $x \in \{-1, 1\}^n$ is usually called a *coloring* in this context.)

* Research supported by the ERC Advanced Grant No. 267165.

31st International Symposium on Computational Geometry (SoCG'15).
Editors: Lars Arge and János Pach; pp. 1–15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This combinatorial notion of discrepancy originated in the classical theory of *irregularities of distribution*, as treated, e.g., in [8, 20, 3], and more recently it has found remarkable applications in computer science and elsewhere (see [43, 15, 31] for general introductions and, e.g., [25] for a recent use).

For the subsequent discussion, we also need the notion of discrepancy for matrices: for an $m \times n$ real matrix A we set $\operatorname{disc} A := \min_{x\in\{-1,1\}^n} \|Ax\|_\infty$. If A is the incidence matrix of the set system $\mathcal{F}$ as above (with $a_{ij} = 1$ if $j \in F_i$ and $a_{ij} = 0$ otherwise), then the matrix definition coincides with the one for set systems.

A set system $\mathcal{F}$ with small, even zero, discrepancy may contain a set system with large discrepancy. This phenomenon was exploited in [14] for showing that, assuming $\mathrm{P} \neq \mathrm{NP}$, no polynomial-time algorithm can distinguish systems $\mathcal{F}$ with zero discrepancy from those with discrepancy of order $\sqrt{n}$ in the regime $m = O(n)$, which practically means that $\operatorname{disc} \mathcal{F}$ cannot be approximated at all in polynomial time.

A better behaved notion is the *hereditary discrepancy* of $\mathcal{F}$, given by

$$\operatorname{herdisc} \mathcal{F} := \max_{J\subseteq V} \operatorname{disc}(\mathcal{F}|_J),$$

were $\mathcal{F}|_J$ denotes the *restriction* of the set system $\mathcal{F}$ to the ground set J, i.e., $\{F \cap J : F \in \mathcal{F}\}$. Similarly, for a matrix A, $\operatorname{herdisc} A := \max_{J\subseteq[n]} \operatorname{disc} A_J$ where A_J is the submatrix of A consisting of the columns indexed by the set J.

At first sight, hereditary discrepancy may seem harder to deal with than discrepancy. For example, while $\operatorname{disc} \mathcal{F} \leq k$ has an obvious polynomial-time verifiable certificate, namely, a suitable coloring $x \in \{-1, 1\}^n$, it is not at all clear how one could certify either $\operatorname{herdisc} \mathcal{F} \leq k$ or $\operatorname{herdisc} \mathcal{F} > k$ in polynomial time.

However, hereditary discrepancy has turned out to have significant advantages over discrepancy. Most of the classical upper bounds for discrepancy of various set systems actually apply to hereditary discrepancy as well. A powerful tool, introduced by Lovász, Spencer and Vesztergombi [28] and called the *determinant lower bound*, works for hereditary discrepancy and *not* for discrepancy. The determinant lower bound for a matrix A is the following algebraically defined quantity:

$$\operatorname{detlb} A = \max_k \max_B |\det B|^{1/k},$$

where B ranges over all $k \times k$ submatrices of A. Lovász et al. proved that $\operatorname{herdisc} A \geq \frac{1}{2} \operatorname{detlb} A$ for all A. Later it was shown in [29] that $\operatorname{detlb} A$ also bounds $\operatorname{herdisc} A$ from above up to a polylogarithmic factor; namely, $\operatorname{herdisc} A = O(\operatorname{detlb}(A) \log(mn)\sqrt{\log n}\,)$.

While the quantity $\operatorname{detlb} A$ enjoys some pleasant properties, there is no known polynomial-time algorithm for computing it. Bansal [4] provided a polynomial-time algorithm that, given a system $\mathcal{F}$ with $\operatorname{herdisc} \mathcal{F} \leq D$, computes a coloring x witnessing $\operatorname{disc} \mathcal{F} = O(D \log(mn))$. However, this is not an approximation algorithm for the hereditary discrepancy in the usual sense, since it may find a low-discrepancy coloring even for $\mathcal{F}$ with large hereditary discrepancy.

The γ_2 factorization norm. The first polynomial-time approximation algorithm with a polylogarithmic approximation factor for hereditary discrepancy was found by the second author, Talwar, and Zhang [37]. Their result was further strengthened and streamlined by the second author and Talwar [35], who showed that hereditary discrepancy is approximated by geometrically defined quantity which turns out to be equivalent to the γ_2 factorization norm from Banach space theory.[1] This connection was implicit in [37].

[1] This equivalence was pointed out to us by Noga Alon and Assaf Naor.

Let the ℓ_∞ norm $\|E\|_\infty$ of an ellipsoid E be defined as the largest ℓ_∞ norm of any point in E. The geometric quantity studied in [35] is the minimum ℓ_∞ norm of a 0-centered ellipsoid E that contains all column vectors of A. As noticed by several experts, this quantity is equal to the γ_2 norm of A, taken as a linear operator from ℓ_1^n to ℓ_∞^m, which is also defined as

$$\gamma_2(A) := \min\{\|B\|_{2\to\infty}\|C\|_{1\to 2} : A = BC\}.$$

Above, $\|\cdot\|_{p\to q}$ stands for the $\ell_p \to \ell_q$ operator norm, and B, C range over linear operators. Treating B and C as matrices, it is easy to see that $\|B\|_{2\to\infty}$ is equal to the largest Euclidean norm of row vectors of B, and $\|C\|_{1\to 2}$ is equal to the largest Euclidean norm of column vectors of C. We will use both the definition in terms of ellipsoids and the one in terms of a factorization of A. We use the notation $\gamma_2(\mathcal{F})$ for a set system $\mathcal{F}$ to mean the γ_2 norm of the incidence matrix of $\mathcal{F}$.

In [35] it was shown that $\gamma_2(A)$ can be approximated to any desired accuracy in polynomial time, and the following two inequalities relating $\gamma_2(A)$ to herdisc A were proved: for every matrix A with m rows,

$$\operatorname{herdisc} A \geq \frac{\gamma_2(A)}{O(\log m)}, \text{ and} \tag{1}$$

$$\operatorname{herdisc} A \leq \gamma_2(A) \cdot O(\sqrt{\log m}\,) \tag{2}$$

These results together provide an $O(\log^{3/2} m)$-approximation algorithm for herdisc A. (As we will see in Section 4.1 below, (1) is actually valid with $\log\min\{m,n\}$ instead of $\log m$.)

The upper bounds guaranteed by inequality (2) are not constructive, in the sense that we do not know of a polynomial-time algorithm that computes a coloring achieving the upper bound. Nevertheless, the algorithms of Bansal [4] or Rothvoss [40] can be used to find colorings with discrepancy $\gamma_2(A) \cdot O(\log m)$ in polynomial time.

Results on the γ_2 norm. A number of useful properties of γ_2 are known, such as the non-obvious fact that it is indeed a norm [46] (we give an example of how the triangle inequality fails for detlb), and the fact that it is is multiplicative under the *Kronecker product* (or tensor product) of matrices [26]. We further prove a stronger form of the triangle inequality for matrices supported on disjoint subsets of the columns.

Linial, Mendelson, Schechtman and Shraibman [27] observed that for sign matrices A, $\gamma_2(A)$ can be formulated as the optimal value of a semidefinite program. Lee, Shraibman, and Špalek used generalized the semidefinite program to arbitrary real matrices, and used it to derive a dual characterization of γ_2. We use this characterization to give a simplified proof of inequality (1). We also prove that $\gamma_2(A)$ is between detlb A and $O(\operatorname{detlb}(A)\log m)$.

We show that both inequalities (1) and (2) are asymptotically tight in the worst case. For (1), the asymptotic tightness is demonstrated on the following simple example: for the system $\mathcal{I}_n$ of initial segments of $\{1, 2, \ldots, n\}$, whose incidence matrix is the lower triangular matrix T_n with 1s on the main diagonal and below it, we prove that the γ_2 norm is of order $\log n$, while the hereditary discrepancy is well known to be 1.

Applications in discrepancy theory. In the second part of the paper we apply the γ_2 norm to prove new results on combinatorial discrepancy, as well as to give simple new proofs of known results.

The most significant result is a new lower bound for the d-dimensional Tusnády's problem; before stating it, let us give some background.

The "great open problem." Discrepancy theory started with a result conjectured by Van der Corput [18, 19] and first proved by Van Aardenne-Ehrenfest [1, 2], stating that every infinite sequence $(u_1, u_2, \ldots)$ of real numbers in $[0,1]$ must have a significant deviation from a "perfectly uniform" distribution. Roth [39] found a simpler proof of a stronger bound, and he re-cast the problem in the following setting, dealing with finite point sets in the unit square $[0,1]^2$ instead of infinite sequences in $[0,1]$:

Given an n-point set $P \subset [0,1]^2$, the *discrepancy* of P is defined as

$$D(P, \mathcal{R}_2) := \sup\Big\{ \Big| |P \cap R| - n\lambda^2(R \cap [0,1]^d) \Big| : R \in \mathcal{R}_2 \Big\},$$

where $\mathcal{R}_2$ denotes the set of all 2-dimensional axis-parallel rectangles (or 2-dimensional intervals), of the form $R = [a_1, b_1] \times [a_2, b_2]$, and λ^2 is the area (2-dimensional Lebesgue measure). More precisely, $D(P, \mathcal{R}_2)$ is the *Lebesgue-measure discrepancy* of P w.r.t. axis-parallel rectangles. Further let $D(n, \mathcal{R}_2) = \inf_{P:|P|=n} D(P, \mathcal{R}_2)$ be the best possible discrepancy of an n-point set.

Roth proved that $D(n, \mathcal{R}_2) = \Omega(\sqrt{\log n})$, while earlier work of Van der Corput yields $D(n, \mathcal{R}_2) = O(\log n)$. Later Schmidt [41] improved the lower bound to $\Omega(\log n)$.

Roth's setting immediately raises the question about a higher-dimensional analog of the problem: letting $\mathcal{R}_d$ stand for the system of all axis-parallel boxes (or d-dimensional intervals) in $[0,1]^d$, what is the order of magnitude of $D(n, \mathcal{R}_d)$? There are many ways of showing an upper bound of $O(\log^{d-1} n)$, the first one being the Halton–Hammersley construction [24, 23], and Roth's lower bound method yields $D(n, \mathcal{R}_d) = \Omega(\log^{(d-1)/2} n)$. In these bounds, d is considered fixed and the implicit constants in the $O(.)$ and $\Omega(.)$ notation may depend on it.

Now, over 50 years later, the upper bound is still the best known, and Roth's lower bound has been improved only a little: first for $d = 3$ by Beck [7] and by Bilyk and Lacey [10], and then for all d by Bilyk, Lacey, and Vagharshakyan [11]. The lower bound from [11] has the form $\Omega((\log n)^{(d-1)/2+\eta(d)})$, where $\eta(d) > 0$ is a constant depending on d, with $\eta(d) \geq c/d^2$ for an absolute constant $c > 0$. Thus, the upper bound for $d \geq 3$ is still about the square of the lower bound, and closing this significant gap is called the "great open problem" in the book [8].

Tusnády's problem. Here we essentially solve a combinatorial analog of this problem. In the 1980s Tusnády raised a question which, in our terminology, can be stated as follows. Let $P \subset \mathbb{R}^2$ be an n-point set, and let $\mathcal{R}_2(P) := \{R \cap P : R \in \mathcal{R}_2\}$ be the system of all subsets of P induced by axis-parallel rectangles $R \in \mathcal{R}_2$. What can be said about the discrepancy of such a set system for the worst possible n-point P? In other words, what is

$$\operatorname{disc}(n, \mathcal{R}_2) = \max\{\operatorname{disc} \mathcal{R}_2(P) : |P| = n\}?$$

We stress that for the Lebesgue-measure discrepancy $D(n, \mathcal{R}_d)$ we ask for the best placement of n points so that each rectangle contains approximately the right number of points, while for $\operatorname{disc}(n, \mathcal{R}_2)$ the point set P is given by an adversary, and we seek a ±1 coloring so that the points in each rectangle are approximately balanced.

Tusnády actually asked if $\operatorname{disc}(n, \mathcal{R}_2)$ could be bounded by a constant independent of n. This was answered negatively by Beck [5], who also proved an upper bound of $O(\log^4 n)$. His lower bound argument uses a "transference principle," showing that the function $\operatorname{disc}(n, \mathcal{R}_2)$ in Tusnády's problem cannot be asymptotically smaller than the smallest achievable Lebesgue-measure discrepancy of n points with respect to axis-aligned boxes. (This principle is actually simple to prove and quite general; Simonovits attributes the idea to V. T. Sós.) The upper bound was improved to $O((\log n)^{3.5+\varepsilon})$ by Beck [6], to $O(\log^3 n)$ by Bohus [12], and to the current best bound of $O(\log^{2.5} n)$ by Srinivasan [44].

The obvious d-dimensional generalization of Tusnády's problem was attacked by similar methods. All known lower bounds so far relied on the transference principle mentioned above. The current best upper bound for $d \geq 3$ is $O(\log^{d+1/2} n)$ due to Larsen [25], which is a a slight strengthening of a previous bound of $O(\log^{d+1/2} n\sqrt{\log\log n}\,)$ from [30].

Here we improve on the lower bound for the d-dimensional Tusnády's problem significantly; while up until now the uncertainty in the exponent of $\log n$ was roughly between $(d-1)/2$ and $d+1/2$, we reduce it to $d-1$ versus $d+1/2$.

▶ **Theorem 1.** *For every fixed $d \geq 2$ and for infinitely many values of n, there exists an n-point set $P \subset \mathbb{R}^d$ with*

$$\operatorname{disc} \mathcal{R}_d(P) = \Omega(\log^{d-1} n),$$

where the constant of proportionality depends only on d.

From the point of view of the "great open problem," this result is perhaps somewhat disappointing, since it shows that, in order to determine the asymptotics of the Lebesgue-measure discrepancy $D(n, \mathcal{R}_d)$, one has to use some special properties of the Lebesgue measure—combinatorial discrepancy cannot help, at least for improving the upper bound.

Using the γ_2 norm as the main tool, our proof of Theorem 1 is surprisingly simple. In a nutshell, first we observe that, since the target bound is polylogarithmic in n, instead of estimating the discrepancy for some cleverly constructed n-point set P, we can bound from below the hereditary discrepancy of the regular d-dimensional grid $[n]^d$, where $[n] = \{1, 2, \ldots, n\}$. By a standard and well known reduction, instead of all d-dimensional intervals in $\mathcal{R}_d$, it suffices to consider only "anchored" intervals, of the form $[0, b_1] \times \cdots \times [0, b_d]$. Now the main observation is that the set system $\mathcal{G}_{d,n}$ induced on $[n]^d$ by anchored intervals is a d-fold product of the system $\mathcal{I}_n$ of one-dimensional intervals mentioned earlier, and its incidence matrix is the d-fold Kronecker product of the matrix T_n.

Thus, by the properties of the γ_2 norm established earlier, we get that $\gamma_2(\mathcal{G}_{d,n})$ is of order $\log^d n$, and inequality (1) finishes the proof of Theorem 1.

At the same time, using the other inequality (2), we obtain a new proof of the best known upper bound $\operatorname{disc}(n, \mathcal{R}_d) = O(\log^{d+1/2} n)$, with no extra effort. This proof is very different from the previously known ones and relatively simple.

The same method also gives a surprisingly precise upper bound on the discrepancy of the set system of all subcubes of the d-dimensional cube $\{0, 1\}^d$, where this time d is a variable parameter, not a constant as before. This discrepancy has previously been studied in [16, 17, 36], and it was known that it is between $2^{c_1 d}$ and $2^{c_2 d}$ for some constants $c_2 > c_1 > 0$. In Section 5.1 we show that it is $2^{(c_0+o(1))d}$, for $c_0 = \log_2(2/\sqrt{3}) \approx 0.2075$.

Immediate applications in computer science. Our lower bound for Tusnády's problem implies a lower bound of $\sqrt{t_u t_q} = \Omega(\log^d n)$ on the update time t_u and query time t_q of constant multiplicity oblivious data structures for orthogonal range searching in $\mathbb{R}^d$ in the group model. This lower bound is tight up to a constant. The relationship between hereditary discrepancy and differential privacy from [33] and the lower bound for Tusnády's problem imply that the necessary error for computing orthogonal range counting queries under differential privacy is $\Omega(\log^{d-1} n)$, which is best possible up to a factor of $\log n$.

Our lower and upper bounds on the discrepancy of subcubes of the Boolean cube $\{0,1\}^d$ and the results from [37] imply that the necessary and sufficient error for computing marginal queries on d-attribute databases under differential privacy is $(2/\sqrt{3})^{d+o(d)}$.

General theorems on discrepancy. Transferring the various properties of the γ_2 norm into the setting of hereditary discrepancy via inequalities (1), (2), we obtain general results about the behavior of discrepancy under operations on set systems. In particular, we get a sharper version of a result of [29] concerning the discrepancy of the union of several set systems, and a new bound on the discrepancy of a set system $\mathcal{F}$ in which every set $F \in \mathcal{F}$ is a disjoint union $F_1 \cup \cdots \cup F_t$, where $\mathcal{F}_1, \ldots, \mathcal{F}_t$ are given set systems and $F_i \in \mathcal{F}_i$, $i = 1, 2, \ldots, t$. These consequences are presented in the full version of the paper.

Other problems in combinatorial discrepancy: new simple proofs. In the full version we also we revisit two set systems for which discrepancy has been studied extensively: arithmetic progressions in $[n]$ and intervals in k permutations of $[n]$. In both of these cases, asymptotically tight bounds have been known. Using the γ_2 norm we recover almost tight upper bounds, up to a factor of $\sqrt{\log n}$, with very short proofs.

2 Properties of the γ_2 norm

2.1 Known properties of γ_2

The γ_2 norm has various favorable properties, which make it a very convenient and powerful tool in studying hereditary discrepancy, as we will illustrate later on. We begin by recalling some classical facts. It is clear that $\gamma_2(A)$ is monotone non-increasing under removing rows or columns of A. From the definition of $\gamma_2(A)$ in terms of factorization of matrices, we also see that $\gamma_2(A) = \gamma_2(A^T)$. Moreover, it is well-known (see e.g. [46]) that γ_2 is indeed a norm and therefore satisfies the triangle inequality, i.e. for any two $m \times n$ matrices A and B we have

$$\gamma_2(A+B) \le \gamma_2(A) + \gamma_2(B). \tag{3}$$

Remark on the determinant lower bound. Here is an example showing that the determinant lower bound of Lovász et al. does not satisfy the (exact) triangle inequality: for

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

we have $\operatorname{detlb} A = \operatorname{detlb} B = 1$, but $\operatorname{detlb}(A+B) = \sqrt{5}$.

It may still be that the determinant lower bound satisfies an approximate triangle inequality, say in the following sense: $\operatorname{detlb}(A_1 + \cdots + A_t) \overset{?}{\le} O(t) \cdot \max_i \operatorname{detlb} A_i$. However, at present we can only prove this kind of inequality with $O(t^{3/2})$ instead of $O(t)$.

On ellipsoids. An ellipsoid E in $\mathbb{R}^m$ is often defined as $\{x \in \mathbb{R}^m : x^T A x \le 1\}$, where A is a positive definite matrix. Here we will mostly work with the *dual matrix* $D = A^{-1}$. Using this dual matrix we have (see, e.g., [42])

$$E = E(D) = \{z \in \mathbb{R}^m : z^T x \le \sqrt{x^T D x} \text{ for all } x \in \mathbb{R}^m\}. \tag{4}$$

This definition can also be used for D only positive semidefinite; if D is singular, then $E(D)$ is a flat (lower-dimensional) ellipsoid.

2.2 Putting matrices side-by-side

▶ **Lemma 2.** *Let A, B be matrices, each with m rows, and let C be a matrix in which each column is a column of A or of B. Then*

$$\gamma_2(C)^2 \le \gamma_2(A)^2 + \gamma_2(B)^2.$$

Proof. After possibly reordering the columns of C, we can write $C = \tilde{A} + \tilde{B}$, where the first k columns of $\tilde{A}$ are among the columns of A and the remaining ℓ columns are zeros, and the last ℓ columns of $\tilde{B}$ are among the columns of B and the first k are zeros.

Since the γ_2 norm is, by definition, monotone under the removal of columns, we have $a := \gamma_2(\tilde{A}) \le \gamma_2(A)$, $b := \gamma_2(\tilde{B}) \le \gamma_2(B)$.

Let $E_1 = E(D_1)$ and $E_2 = E(D_2)$ be ellipsoids witnessing $\gamma_2(\tilde{A})$ and $\gamma_2(\tilde{B})$, respectively. We claim that the ellipsoid $E(D_1 + D_2)$ contains all columns of $\tilde{A}$ and also all columns of $\tilde{B}$. This is clear from the definition of the ellipsoid $E(D) = \{z : z^T x \le \sqrt{x^T D x} \text{ for all } x\}$, since for every x, we have $x^T(D_1 + D_2)x = x^T D_1 x + x^T D_2 x \ge x^T D_1 x$ by positive semidefiniteness of D_2. All the diagonal entries of D_1 are bounded above by a^2, those of D_2 are at most b^2, and hence $\|E\|_\infty \le \sqrt{a^2 + b^2}$. ◀

▶ **Lemma 3.** *If C is a block-diagonal matrix with blocks A and B on the diagonal, then $\gamma_2(C) = \max(\gamma_2(A), \gamma_2(B))$.*

Proof. If D_1 is the dual matrix of the ellipsoid witnessing $\gamma_2(A)$ and similarly for D_2 and B, then the block-diagonal matrix D with blocks D_1 and D_2 on the diagonal defines an ellipsoid containing all columns of C. This is easy to check using the formula (4) defining $E(D)$ and the fact that a sum of positive definite matrices is positive definite. ◀

2.3 Dual formulation

Let $\|A\|_*$ denote the *nuclear norm* of a matrix A, which is the sum of the singular values of A (other names for $\|A\|_*$ are *Schatten* 1*-norm*, *trace norm*, or *Ky Fan n-norm*; see the text by Bhatia [9] for general background on symmetric matrix norms). Using a semidefinite formulation of γ_2, and the duality theory for semidefinite programming, Lee, Shraibman and Špalek [26] derived a dual characterization of the γ_2 norm as a maximization problem.

▶ **Theorem 4** ([26, Thm. 9])**.** *We have*

$$\gamma_2(A) = \max\{\|P^{1/2} A Q^{1/2}\|_* : P, Q \textit{ diagonal, nonnegative}, \operatorname{Tr} P = \operatorname{Tr} Q = 1\}.$$

Several times we will use this theorem with A a square matrix and $P = Q = \frac{1}{n} I_n$, in which case it gives $\gamma_2(A) \ge \frac{1}{n}\|A\|_*$.

2.4 Kronecker product

Let A be an $m \times n$ matrix and B a $p \times q$ matrix. We recall that the *Kronecker product* $A \otimes B$ is the following $mp \times nq$ matrix, consisting of $m \times n$ blocks of size $p \times q$ each:

$$\begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}$$

In [26] it was shown that γ_2 is multiplicative with respect to the Kronecker product:

▶ **Theorem 5** ([26, Thm. 17])**.** *For every two matrices A, B we have*

$$\gamma_2(A \otimes B) = \gamma_2(A) \cdot \gamma_2(B).$$

3 The γ_2 norm for intervals

In this section we deal with a particular example: the system $\mathcal{I}_n$ of all initial segments $\{1, 2, \ldots, i\}$, $i = 1, 2, \ldots, n$, of $\{1, 2, \ldots, n\}$. Its incidence matrix is T_n, the $n \times n$ matrix with 0s above the main diagonal and 1s everywhere else.

It is well known, and easy to see, that $\operatorname{herdisc} T_n = 1$. We will prove that $\gamma_2(T_n)$ is of order $\log n$. This shows that the γ_2 norm can be $\log n$ times larger than the hereditary discrepancy, and thus the inequality (1) is asymptotically tight.

Moreover, this example is one of the key ingredients in the proof of the lower bound on the d-dimensional Tusnády problem.

▶ **Proposition 6.** *We have $\gamma_2(T_n) = \Theta(\log n)$.*

The upper bound follows from the observation $\operatorname{herdisc} T_n = 1$ and the inequality (1) relating γ_2 to herdisc. It can also be proved directly using, for example, a decomposition into dyadic intervals. In the next section we prove the lower bound.

3.1 Lower bound on $\gamma_2(T_n)$

Proof of the lower bound in Proposition 6. The nuclear norm $\|T_n\|_*$ can be computed exactly (we are indebted to Alan Edelman and Gil Strang for this fact); namely, the singular values of T_n are

$$\frac{1}{2 \sin \frac{(2j-1)\pi}{4n+2}}, \quad j = 1, 2, \ldots, n.$$

Using the inequality $\sin x \leq x$ for $x \geq 0$, we get

$$\gamma_2(T_n) \geq \frac{1}{n} \|T_n\|_* \geq \frac{2n+1}{\pi n} \sum_{j=1}^{n} \frac{1}{2j-1} = \Omega(\log n),$$

as needed.

The singular values of T_n can be obtained from the eigenvalues of the matrix $S_n := (T_n T_n^T)^{-1}$ which, as is not difficult to check, has the following simple tridiagonal form:

$$\begin{pmatrix} 2 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -1 & 1 \end{pmatrix}$$

(the 1 in the lower right corner is exceptional; the rest of the main diagonal are 2s). By general properties of eigenvalues and singular values, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of S_n, then the singular values of T_n are $\lambda_1^{-1/2}, \ldots, \lambda_n^{-1/2}$. The eigenvalues of S_n are computed, as a part of more general theory, in Strang and MacNamara [45, Sec. 9]; the calculation is not hard to verify since they also give the eigenvectors explicitly.

One can also calculate the characteristic polynomial $p_n(x)$ of S_n: it satisfies the recurrence $p_{n+1} = (2 - x)p_n - p_{n-1}$ with initial conditions $p_1 = 1 - x$ and $p_0 = 1$, from which one can check that $p_n(x) = U_n\left(\frac{2-x}{2}\right) - U_{n-1}\left(\frac{2-x}{2}\right)$, where U_n is the degree-n Chebyshev polynomial of the second kind. The claimed roots of p_n can then be verified using the trigonometric representation of U_n. ◀

4 Deviation of the γ_2 norm from the hereditary discrepancy

Here we consider the inequalities (1) and (2) relating γ_2 and herdisc. For the first one we provide a simplified and elementary proof, and for the second one we briefly recall the proof and prove asymptotic optimality.

We have already seen in Section 3 that (1) is asymptotically tight. Let us first mention a simple but perhaps useful observation, which gives a somewhat weaker result.

There are examples of set systems $\mathcal{F}_1, \mathcal{F}_2$ on an n-point set X such that $|\mathcal{F}_1|, |\mathcal{F}_2| = O(n)$, $\operatorname{herdisc} \mathcal{F}_1$ and $\operatorname{herdisc} \mathcal{F}_2$ are bounded by a constant (actually by 1), and $\operatorname{herdisc}(\mathcal{F}_1 \cup \mathcal{F}_2) = \Omega(\log n)$ [38, 34]. Therefore, no quantity obeying the triangle inequality (possibly up to a constant), such as the γ_2 norm, can approximate herdisc with a factor better than $\log n$.

4.1 The γ_2 norm is at most $\log m$ times the determinant lower bound

We establish the following inequalities relating the γ_2 norm to the determinant lower bound.

▶ **Theorem 7.** *For any $m \times n$ matrix A of rank r,*

$$\operatorname{detlb} A \le \gamma_2(A) \le O(\log r) \cdot \operatorname{detlb} A.$$

Inequality (1) is an immediate consequence of the second inequality in the theorem (and of $r \le \min\{m, n\}$):

$$\gamma_2(A) \le O(\log \min\{m, n\}) \cdot \operatorname{detlb} A \le O(\log \min\{m, n\}) \operatorname{herdisc} A,$$

where the last inequality uses the Lovász–Spencer–Vesztergombi bound $\operatorname{herdisc} A \ge \frac{1}{2} \operatorname{detlb} A$. In [35], inequality (1) was proved by using a sophisticated tool, the *restricted invertibility principle* of Bourgain and Tzafriri; see [13, 47]. Our proof of Theorem 7 is based only on elementary linear algebra and the determinant lower bound.

Before we prove Theorem 7, we need a lemma similar to an argument in [29].

▶ **Lemma 8.** *Let A be an $k \times n$ matrix, and let W be a nonnegative diagonal unit-trace $n \times n$ matrix. Then there exists a k-element set $J \subseteq [n]$ such that*

$$|\det A_J|^{1/k} \ge \sqrt{k/e} \cdot |\det AWA^T|^{1/2k}.$$

Proof of Theorem 7. For the inequality $\operatorname{detlb} A \le \gamma_2(A)$, we first observe that if B is a $k \times k$ matrix, then

$$|\det B|^{1/k} \le \frac{1}{k} \|B\|_* \tag{5}$$

Indeed, the left-hand side is the geometric mean of the singular values of B, while the right-hand side is the arithmetic mean.

Now let B be a $k \times k$ submatrix of A with $\operatorname{detlb} A = |\det B|^{1/k}$; then

$$\operatorname{detlb} A = |\det B|^{1/k} \le \frac{1}{k} \|B\|_* \le \gamma_2(B) \le \gamma_2(A).$$

For the second inequality $\gamma_2(A) \le O(\log m) \cdot \operatorname{detlb} A$, we compare $\det BB^T$ and the nuclear norm of B for a carefully chosen (rectangular) matrix B. First let P_0 and Q_0 be diagonal unit-trace matrices with $\gamma_2(A) = \|P_0^{1/2} A Q_0^{1/2}\|$ as in Theorem 4. For brevity, let us write $\tilde{A} := P_0^{1/2} A Q_0^{1/2}$, and let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ be the nonzero singular values of $\tilde{A}$.

By a standard bucketing argument (see, e.g., [29, Lemma 7]), there is some $t > 0$ such that if we set $K := \{i \in [m] : t \le \sigma_i < 2t\}$, then

$$\sum_{i\in K} \sigma_i \ge \Omega(\tfrac{1}{\log r}) \sum_{i=1}^{m} \sigma_i.$$

Let us set $k := |K|$.

Next, we define a suitable $k \times n$ matrix with singular values σ_i, $i \in K$. Let $\tilde{A} = U\Sigma V^T$ be the singular-value decomposition of $\tilde{A}$, with U and V orthogonal and Σ having $\sigma_1, \dots, \sigma_r$ on the main diagonal.

Let Π_K be the $k \times m$ matrix corresponding to the projection on the coordinates indexed by K; that is, Π_K has 1s in positions $(1, i_1), \dots, (k, i_k)$, where $i_1 < \dots < i_k$ are the elements of K. The matrix $\Pi_K\Sigma = \Pi_K U^T \tilde{A} V = U_K^T \tilde{A} V$ has singular values σ_i, $i \in K$, and so does the matrix $U_K^T \tilde{A}$, since right multiplication by the orthogonal matrix V^T does not change the singular values.

This $k \times m$ matrix $U_K^T \tilde{A}$ is going to be the matrix B alluded to in the sketch of the proof idea above. We have

$$|\det BB^T|^{1/2k} = \Bigl(\prod_{i\in K} \sigma_i\Bigr)^{1/k} \ge \frac{1}{2k} \sum_{i\in K} \sigma_i = \Omega\bigl(\tfrac{1}{k \log r}\bigr)\gamma_2(A).$$

It remains to relate $\det BB^T$ to the determinant of a square submatrix of A, and this is where Lemma 8 is applied—actually applied twice, once for columns, and once for rows.

First we set $C := U_K^T P_0^{1/2} A$; then $B = C Q_0^{1/2}$. Applying Lemma 8 with C in the role of A and Q_0 in the role of W, we obtain a k-element index set $J \subseteq [n]$ such that

$$|\det C_J|^{1/k} \ge \sqrt{k/e} \cdot |\det BB^T|^{1/2k}.$$

Next, we set $D := P_0^{1/2} A_J$, and we claim that $\det D^T D \ge (\det C_J)^2$. Indeed, we have $C_J = U_K^T D$, and, since U is an orthogonal transformation, $(U^T D)^T (U^T D) = D^T D$. Then, by the Binet–Cauchy formula,

$$\begin{aligned}\det D^T D = \det(U^T D)^T (U^T D) &= \sum_L (\det U_L^T D)^2 \\ &\ge (\det U_K^T D)^2 = (\det C_J)^2.\end{aligned}$$

The next (and last) step is analogous. We have $D^T = A_J^T P_0^{1/2}$, and so we apply Lemma 8 with A_J^T in the role of A and P_0 in the role of W, obtaining a k-element subset $I \subseteq [m]$ with $|\det A_{I,J}|^{1/k} \ge \sqrt{k/e} \cdot |\det D^T D|^{1/2k}$ (where $A_{I,J}$ is the submatrix of A with rows indexed by I and columns by J).

Following the chain of inequalities backwards, we have

$$\begin{aligned}\operatorname{detlb} A &\ge |\det A_{I,J}|^{1/k} \ge \sqrt{k/e} \cdot |\det D^T D|^{1/2k} \ge \sqrt{k/e} \cdot |\det C_J|^{1/k} \\ &\ge (k/e) |\det BB^T|^{1/2k} = \Omega\bigl(\tfrac{1}{\log r}\bigr)\gamma_2(A),\end{aligned}$$

and the theorem is proved. ◀

4.2 The hereditary discrepancy can be $\sqrt{\log m}$ times larger than γ_2

Next, we show that $\sqrt{\log m}$ in inequality (2) cannot be replaced by any asymptotically smaller factor.

▶ **Theorem 9.** *For all m, there are $m \times n$ matrices A, with $n = \Theta(\log m)$, such that*

$$\operatorname{herdisc} A \geq \Omega(\sqrt{\log m}\,) \cdot \gamma_2(A).$$

Proof. A very simple example is the incidence matrix A of the system of all subsets of $[n]$, with $m = 2^n$, whose discrepancy is $n/2 = \Theta(\log m)$. Indeed, the characteristic vectors of all sets fit into the ball of radius $\sqrt{n}$, and hence $\gamma_2(A) = \gamma_2(A^T) \leq \sqrt{n} = O(\sqrt{\log m})$, where we used the fact that γ_2 is invariant under transposition. ◀

5 On Tusnády's problem

Proof of Theorem 1. The proof was already sketched in the introduction, so here we just present it slightly more formally. Let $\mathcal{A}_d \subseteq \mathcal{R}_d$ be the set of all *anchored* axis-parallel boxes, of the form $[0, b_1] \times \cdots \times [0, b_d]$. Clearly $\operatorname{disc}(n, \mathcal{A}_d) \leq \operatorname{disc}(n, \mathcal{R}_d)$, and since every box $R \in \mathcal{R}_d$ can be expressed as a signed combination of at most 2^d anchored boxes, we have $\operatorname{disc}(n, \mathcal{R}_d) \leq 2^d \operatorname{disc}(n, \mathcal{A}_d)$.

Let us consider the d-dimensional grid $[n]^d \subset \mathbb{R}^d$ (with n^d points), and let $\mathcal{G}_{d,n} = \mathcal{A}_d([n]^d)$ be the subsets induced on it by anchored boxes. It suffices to prove that $\operatorname{herdisc} \mathcal{G}_{d,n} = \Omega(\log^{d-1} n)$, and for this, in view of inequality (1), it is enough to show that $\gamma_2(\mathcal{G}_{d,n}) = \Omega(\log^d n)$.

Now $\mathcal{G}_{d,n}$ is (isomorphic to) the d-fold product $\mathcal{I}_n^d$ of the system of initial segments in $\{1, 2, \ldots, n\}$, and so $\gamma_2(\mathcal{G}_{d,n}) = \gamma_2(T_n)^d = \Theta(\log^d n)$ (Theorem 5 and Proposition 6).

This finishes the proof of the lower bound. To prove the upper bound $\operatorname{disc}(n, \mathcal{R}_d) = O(\log^{d+1/2} n)$, we consider an arbitrary n-point set $P \subset \mathbb{R}^d$. Since the set system $\mathcal{A}_d(P)$ is not changed by a monotone transformation of each of the coordinates, we may assume $P \subseteq [n]^d$. Hence

$$\operatorname{disc}(\mathcal{A}_d(P)) \leq \operatorname{herdisc} \mathcal{G}_{d,n} \leq O(\gamma_2(\mathcal{G}_{d,n}) \sqrt{\log n^d}\,) = O(\log^{d+1/2} n).$$

◀

5.1 Discrepancy of boxes in high dimension

Chazelle and Lvov [16, 17] investigated the hereditary discrepancy of the set system $\mathcal{C}_d := \mathcal{R}_d(\{0,1\}^d)$, the set system induced by axis-parallel boxes on the d-dimensional Boolean cube $\{0,1\}^d$. In other words, the sets in $\mathcal{C}_d$ are subcubes of $\{0,1\}^d$. Unlike for Tusnády's problem where d was considered fixed, here one is interested in the asymptotic behavior as $d \to \infty$.

Chazelle and Lvov proved $\operatorname{herdisc} \mathcal{C}_d = \Omega(2^{cd})$ for an absolute constant $c \approx 0.0477$, which was later improved to $c = 0.0625$ in [36] (in relation to the hereditary discrepancy of homogeneous arithmetic progressions). Here we obtain an optimal value of the constant c:

▶ **Theorem 10.** *The system $\mathcal{C}_d$ of subcubes of the d-dimensional Boolean cube satisfies*

$$\operatorname{herdisc} \mathcal{C}_d = 2^{c_0 d + o(d)},$$

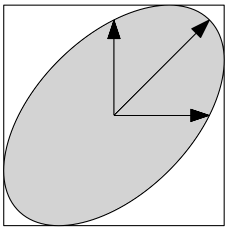
where $c_0 = \log_2(2/\sqrt{3}) \approx 0.2075$. The same bound holds for the system $\mathcal{A}_d(\{0,1\}^d)$ of all subsets of the cube induced by anchored boxes.

Proof. The number of sets in $\mathcal{C}_d$ is 3^d, and so in view of inequalities (1) and (2) it suffices to prove $\gamma_2(\mathcal{C}_d) = \gamma_2(\mathcal{A}_d(\{0,1\}^d)) = 2^{c_0 d}$.

The system $\mathcal{C}_d$ is the d-fold product $\mathcal{C}_1^d$, and so by Theorem 5, $\gamma_2(\mathcal{C}_d) = \gamma_2(\mathcal{C}_1)^d$. The incidence matrix of $\mathcal{C}_1$ is

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

To get an upper bound on $\gamma_2(A)$, we exhibit an appropriate ellipsoid; it is more convenient to do it for A^T, since this is a planar problem. The optimal ellipse containing the rows of A is $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 - x_1 x_2 \leq 1\}$; here are a picture and the dual matrix:



$$D = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{4}{3} \end{pmatrix}.$$

Hence $\gamma_2(A) \leq 2/\sqrt{3}$. The same ellipse also works for the incidence matrix of the system $\mathcal{A}_1(\{0,1\})$, which is the familiar lower triangular matrix T_2.

There are several ways of bounding $\gamma_2(T_2) \leq \gamma_2(A)$ from below. For example, we can use Theorem 4 with

$$P = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}, \quad Q = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}.$$

With some effort (or a computer algebra system) one can check that the singular values of $P^{1/2} T_2 Q^{1/2}$ are $\frac{1}{\sqrt{3}} \pm \frac{1}{3}$, and hence the nuclear norm is $2/\sqrt{3}$ as needed.

Alternatively, one can also check the optimality of the ellipse above by elementary geometry, or exhibit an optimal solution of the dual semidefinite program for $\gamma_2(T_2)$. ◀

Other set systems. In the full version of the paper we use the properties of γ_2 to give new simple proofs of other upper and lower bounds in discrepancy theory. In particular, we revisit two set systems that have been studied extensively: arithmetic progressions in $[n]$ and intervals of k permutations on $[n]$. While the bounds we get are slightly suboptimal, the proofs are very short.

6 Applications in Computer Science

Range searching in the oblivious group model. A range searching problem is defined by a system $\mathcal{F}$ of subsets of a set $P \subseteq \mathbb{R}^d$. The input is an assignment of weights to P, where each weight is an element of a commutative group; a query is specified by a range $F \in \mathcal{F}$ may ask, for example, whether for the sum of the weights of points in F or whether it is non-zero. The goal is to maintain a data structure that supports fast queries. One of the best studied special cases is orthogonal range searching, in which $\mathcal{F}$ is induced by axis-aligned boxes, i.e. $\mathcal{F} = \mathcal{R}_d(P)$.

Following Fredman [22] and Larsen [25], we define an oblivious data structure for a range searching problem given by $\mathcal{F}$ as a factorization $A = BC$, where $A \in \{0,1\}^{m \times n}$ is the incidence matrix of $\mathcal{F}$, and B, C are integer matrices. The update time t_u is defined as the maximum number of non-zero entries of a column of C, and the query time t_q is the maximum number of non-zero entries of a row of B. The multiplicity Δ is the maximum absolute value of an entry in B or C. The motivation is that the actual data structure kept in memory is $y = Cx$, where x are the weights assigned to P, and queries are answered by computing the appropriate entry of By. Then, updating a single weight requires updating at most t_u cells in the data structure, and answering a query requires reading at most t_q cells.

By the factorization definition of γ_2, we have that for any oblivious data structure for $\mathcal{F}$, $\gamma_2(\mathcal{F}) \le |\Delta|\sqrt{t_u t_q}$. In the proof of Theorem 1 we showed that for $\mathcal{G}_{d,n} = \mathcal{A}_d([n]^d)$ (recall $\mathcal{A}_d$ is the set of axis-aligned boxes anchored at 0), $\gamma_2(\mathcal{G}_{d,n}) = \Theta((\log n)^d)$. Therefore, for any oblivious data structure for orthogonal range searching on P with constant multiplicity, $t_u t_q = \Omega((\log n)^d)$. This lower bound is tight up to constants. The best previous lower bound was due to Larsen [25] and was on the order of $(\log n)^{(d-1)/2}$.

Differential Privacy. Differential privacy is a popular definition of privacy for data analysis algorithms. Informally, it states that an algorithm is private if its output distribution is almost the same when we add or remove one person's data from the input; see the book [21] for the formal definition. A class of problems of general interest in differential privacy are counting problems, in which a database is a multiset of elements of a universe U, and a family of queries is specified by a system $\mathcal{F}$ of subsets of U. A query given by a set $F \in \mathcal{F}$ asks for the number of elements of F that are in the database D (counted with multiplicity). In [37] it was shown that, up to factors logarithmic in $|\mathcal{F}|$, the optimal worst-case error for answering the queries specified by $\mathcal{F}$ is equal to $\gamma_2(\mathcal{F})$. A query set of special interest is the one given by the set system $\mathcal{C}_d$ of subcubes of the d-dimensional boolean cube, which corresponds to the set of marginal queries on a d-dimensional database. For these queries, Theorem 10 shows that the optimal worst-case error is on the order of $2^{c_0 d \pm o(d)}$, where $c_0 = \log_2(2/\sqrt{3})$. The best previous upper bound was $2^{d/2+o(d)}$.

Acknowledgments. We would like to thank Alan Edelman and Gil Strang for invaluable advice concerning the singular values of the matrix in Proposition 6, and Van Vu for recommending the right experts for this question. We would also like to thank Noga Alon and Assaf Naor for pointing out that the geometric quantity in [35, 32] is equivalent to the γ_2 norm. We also thank Imre Bárány and Vojtěch Tůma for useful discussions.

References

1 T. van Aardenne-Ehrenfest. Proof of the impossibility of a just distribution of an infinite sequence of points. *Nederl. Akad. Wet., Proc.*, 48:266–271, 1945. Also in *Indag. Math.* 7, 71-76 (1945).

2 T. van Aardenne-Ehrenfest. On the impossibility of a just distribution. *Nederl. Akad. Wet., Proc.*, 52:734–739, 1949. Also in *Indag. Math.* 11, 264-269 (1949).

3 J. R. Alexander, J. Beck, and W. W. L. Chen. Geometric discrepancy theory and uniform distribution. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, chapter 10, pages 185–207. CRC Press LLC, Boca Raton, FL, 1997.

4 N. Bansal. Constructive algorithms for discrepancy minimization. `http://arxiv.org/abs/1002.2259`, also in *FOCS'10: Proc. 51st IEEE Symposium on Foundations of Computer Science*, pages 3–10, 2010.

5 J. Beck. Balanced two-colorings of finite sets in the square. I. *Combinatorica*, 1:327–335, 1981.

6 J. Beck. Balanced two-colorings of finite sets in the cube. *Discrete Mathematics*, 73:13–25, 1989.

7 J. Beck. A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. *Compositio Math.*, 72:269–339, 1989.

8 J. Beck and W. W. L. Chen. *Irregularities of Distribution.* Cambridge University Press, Cambridge, 1987.

9 Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1997.

10 D. Bilyk and M. T. Lacey. On the small ball inequality in three dimensions. *Duke Math. J.*, 143(1):81–115, 2008.

11 D. Bilyk, M. T. Lacey, and A. Vagharshakyan. On the small ball inequality in all dimensions. *J. Funct. Anal.*, 254(9):2470–2502, 2008.

12 G. Bohus. On the discrepancy of 3 permutations. *Random Struct. Algo.*, 1:215–220, 1990.

13 J. Bourgain and L. Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.

14 M. Charikar, A. Newman, and A. Nikolov. Tight hardness results for minimizing discrepancy. In *Proc. 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, California, USA*, pages 1607–1614, 2011.

15 B. Chazelle. *The Discrepancy Method.* Cambridge University Press, Cambridge, 2000.

16 B. Chazelle and A. Lvov. A trace bound for the hereditary discrepancy. *Discrete Comput. Geom.*, 26(2):221–231, 2001.

17 B. Chazelle and A. Lvov. The discrepancy of boxes in higher dimension. *Discrete Comput. Geom.*, 25(4):519–524, 2001.

18 J. G. van der Corput. Verteilungsfunktionen I. *Akad. Wetensch. Amsterdam, Proc.*, 38:813–821, 1935.

19 J. G. van der Corput. Verteilungsfunktionen II. *Akad. Wetensch. Amsterdam, Proc.*, 38:1058–1066, 1935.

20 M. Drmota and R. F. Tichy. *Sequences, discrepancies and applications (Lecture Notes in Mathematics 1651).* Springer-Verlag, Berlin etc., 1997.

21 Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

22 Michael L. Fredman. The complexity of maintaining an array and computing its partial sums. *J. ACM*, 29(1):250–260, 1982.

23 J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2:84–90, 1960.

24 J. M. Hammersley. Monte Carlo methods for solving multivariable problems. *Ann. New York Acad. Sci.*, 86:844–874, 1960.

25 K. G. Larsen. On range searching in the group model and combinatorial discrepancy. *SIAM Journal on Computing*, 43(2):673–686, 2014.

26 Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC 2008, 23-26 June 2008, College Park, Maryland, USA*, pages 71–80. IEEE Computer Society, 2008.

27 Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

28 L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European J. Combin.*, 7:151–160, 1986.

29 J. Matoušek. The determinant bound for discrepancy is almost tight. *Proc. Amer. Math. Soc.*, 141(2):451–460, 2013.

30 J. Matoušek. On the discrepancy for boxes and polytopes. *Monatsh. Math.*, 127(4):325–336, 1999.

31 J. Matoušek. *Geometric Discrepancy (An Illustrated Guide), 2nd printing.* Springer-Verlag, Berlin, 2010.

32 Jiří Matoušek and Aleksandar Nikolov. Combinatorial discrepancy for boxes via the ellipsoid-infinity norm. To appear in SoCG 15., 2014.

33 S. Muthukrishnan and A. Nikolov. Optimal private halfspace counting via discrepancy. In *STOC '12: Proceedings of the 44th symposium on Theory of Computing*, pages 1285–1292, New York, NY, USA, 2012. ACM.

34 A. Newman, O. Neiman, and A. Nikolov. Beck's three permutations conjecture: A counterexample and some consequences. In *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 253–262, 2012.

35 A. Nikolov and K. Talwar. Approximating hereditary discrepancy via small width ellipsoids. Preprint arXiv:1311.6204, 2013.

36 A. Nikolov and K. Talwar. On the hereditary discrepancy of homogeneous arithmetic progressions. *To Appear in Proceedings of AMS*, 2013.

37 A. Nikolov, K. Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Proc. 45th ACM Symposium on Theory of Computing (STOC), Palo Alto, California, USA*, pages 351–360, 2013. Full version to appear in SIAM Journal on Computing as The Geometry of Differential Privacy: the Small Database and Approximate Cases.

38 D. Pálvölgyi. Indecomposable coverings with concave polygons. *Discrete Comput. Geom.*, 44(3):577–588, 2010.

39 K. F. Roth. On irregularities of distribution. *Mathematika*, 1:73–79, 1954.

40 Thomas Rothvoß. Constructive discrepancy minimization for convex sets. *CoRR*, abs/1404.0339, 2014. To Appear in FOCS 2014.

41 W. M. Schmidt. On irregularities of distribution VII. *Acta Arith.*, 21:45–50, 1972.

42 A. Seeger. Calculus rules for combinations of ellipsoids and applications. *Bull. Australian Math. Soc.*, 47(01):1–12, 1993.

43 J. Spencer. *Ten Lectures on the Probabilistic Method.* CBMS-NSF. SIAM, Philadelphia, PA, 1987.

44 A. Srinivasan. Improving the discrepancy bound for sparse matrices: better approximations for sparse lattice approximation problems. In *Proc. 8th ACM-SIAM Symposium on Discrete Algorithms*, pages 692–701, 1997.

45 G. Strang and S. MacNamara. Functions of difference matrices are Toeplitz plus Hankel. *SIAM Review*, 2014. To appear.

46 Nicole Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics.* Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.

47 R. Vershynin. John's decompositions: Selecting a large part. *Israel Journal of Mathematics*, 122(1):253–277, 2001.