Constructive Discrepancy Minimization with Hereditary L2 Guarantees

Kasper Green Larsen
Department of Computer Science, Aarhus University, Denmark
larsen@cs.au.dk

Abstract

In discrepancy minimization problems, we are given a family of sets $\mathcal{S} = \{S_1, \dots, S_m\}$, with each $S_i \in \mathcal{S}$ a subset of some universe $U = \{u_1, \dots, u_n\}$ of n elements. The goal is to find a coloring $\chi : U \to \{-1, +1\}$ of the elements of U such that each set $S \in \mathcal{S}$ is colored as evenly as possible. Two classic measures of discrepancy are ℓ_∞-discrepancy defined as $\text{disc}_\infty(\mathcal{S}, \chi) := \max_{S \in \mathcal{S}} |\sum_{u_i \in S} \chi(u_i)|$ and ℓ_2-discrepancy defined as $\text{disc}_2(\mathcal{S}, \chi) := \sqrt{(1/|\mathcal{S}|) \sum_{S \in \mathcal{S}} \left(\sum_{u_i \in S} \chi(u_i)\right)^2}$. Breakthrough work by Bansal [FOCS'10] gave a polynomial time algorithm, based on rounding an SDP, for finding a coloring χ such that $\text{disc}_\infty(\mathcal{S}, \chi) = O(\lg n \cdot \text{herdisc}_\infty(\mathcal{S}))$ where $\text{herdisc}_\infty(\mathcal{S})$ is the hereditary ℓ_∞-discrepancy of $\mathcal{S}$. We complement his work by giving a clean and simple $O((m + n)n^2)$ time algorithm for finding a coloring χ such $\text{disc}_2(\mathcal{S}, \chi) = O(\sqrt{\lg n} \cdot \text{herdisc}_2(\mathcal{S}))$ where $\text{herdisc}_2(\mathcal{S})$ is the hereditary ℓ_2-discrepancy of $\mathcal{S}$. Interestingly, our algorithm avoids solving an SDP and instead relies simply on computing eigendecompositions of matrices. To prove that our algorithm has the claimed guarantees, we also prove new inequalities relating both herdisc_∞ and herdisc_2 to the eigenvalues of the incidence matrix corresponding to $\mathcal{S}$. Our inequalities improve over previous work by Chazelle and Lvov [SCG'00] and by Matousek, Nikolov and Talwar [SODA'15+SCG'15]. We believe these inequalities are of independent interest as powerful tools for proving hereditary discrepancy lower bounds. Finally, we also implement our algorithm and show that it far outperforms random sampling of colorings in practice. Moreover, the algorithm finishes in a reasonable amount of time on matrices of sizes up to 10000×10000.

2012 ACM Subject Classification Theory of computation

Keywords and phrases Discrepancy, Hereditary Discrepancy, Combinatorics, Computational Geometry

Digital Object Identifier 10.4230/LIPIcs.STACS.2019.45

Funding This work is supported by a Villum Young Investigator Grant and an AUFF Starting Grant.

1 Introduction

Combinatorial discrepancy minimization is an important field with numerous applications in theoretical computer science, see e.g. the excellent books by Chazelle [9] and Matousek [16]. In discrepancy minimization problems, we are typically given a family of sets $\mathcal{S} = \{S_1, \dots, S_m\}$, with each $S_i \in \mathcal{S}$ a subset of some universe $U = \{u_1, \dots, u_n\}$ of n elements. The goal is to find a red-blue coloring of the elements of U such that each set $S \in \mathcal{S}$ is colored as evenly as possible. More formally, if we define the $m \times n$ incidence matrix A with $a_{i,j} = 1$ if $u_j \in S_i$ and $a_{i,j} = 0$ otherwise, then we seek a coloring $x \in \{-1, +1\}^n$ minimizing either the ℓ_∞-discrepancy $\text{disc}_\infty(A, x) := \|Ax\|_\infty$ or the ℓ_2-discrepancy $\text{disc}_2(A, x) = (1/\sqrt{m})\|Ax\|_2$. We say that the ℓ_∞-discrepancy of A is $\text{disc}_\infty(A) := \min_{x \in \{-1,+1\}^n} \text{disc}_\infty(A, x)$ and the ℓ_2-discrepancy of A is $\text{disc}_2(A) := \min_{x \in \{-1,+1\}^n} \text{disc}_2(A, x)$. With this matrix view, it is clear that discrepancy minimization makes sense also for general matrices and not just ones arising from set systems.

36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019).
Editors: Rolf Niedermeier and Christophe Paul; Article No. 45; pp. 45:1–45:12
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Much research has been devoted to understanding both the ℓ_∞- and ℓ_2-discrepancy of various families of set systems and matrices. In particular set systems corresponding to incidences between geometric objects such as axis-aligned rectangles and points have been studied extensively, see e.g. [17, 15, 1, 11]. Another fruitful line of research has focused on general matrices, including the celebrated "Six Standard Devitations Suffice" result by Spencer [21], showing that any $n \times n$ matrix with $|a_{i,j}| \leq 1$ admits a coloring $x \in \{-1, +1\}^n$ such that $\mathrm{disc}_\infty(A, x) = O(\sqrt{n})$. Finding low discrepancy colorings for set systems where each element appears in at most t sets (the matrix A has at most t non-zeroes per column, all bounded by 1 in absolute value) has also received much attention. Beck and Fiala [7] gave a deterministic algorithm that finds a coloring x with $\mathrm{disc}_\infty(A, x) = O(t)$. Banaszczyk [2] improved this to $O(\sqrt{t \lg n})$ when $t \geq \lg n$. Determining whether a discrepancy of $O(\sqrt{t})$ can be achieved remains one of the biggest open problems in discrepancy minimization.

Constructive Discrepancy Minimization. Many of the original results, like Spencer's [21] and Banaszczyk's [2] were purely existential and it was not clear whether polynomial time algorithms finding such colorings were possible. In fact, Charikar et al. [8] presented very strong negative results in this direction. More concretely, they proved that it is NP-hard to even distinguish whether the ℓ_∞- or ℓ_2-discrepancy of an $n \times n$ set system is 0 or $\Omega(\sqrt{n})$. The first major breakthrough on the upper bound side was due to Bansal [3], who amongst others gave a polynomial time algorithm for finding a coloring matching the bounds by Spencer. Brilliant follow-up work by Lovett and Meka [14] gave simpler randomized algorithms achieving the same. A deterministic algorithm for Spencer's result was later given by Levy et al. [12]. A number of constructive algorithms were also given for the "sparse" set system case, finally resulting in polynomial time algorithms [4, 6, 5] matching the existential results by Banaszczyk.

Another very surprising result in Bansal's seminal paper [3] shows that, given a matrix A, one can find in polynomial time a coloring x achieving an ℓ_∞-discrepancy roughly bounded by the *hereditary* discrepancy of A. Hereditary discrepancy is a notion introduced by Lovász et al. [13] in order to prove discrepancy lower bounds. The hereditary ℓ_∞-discrepancy of a matrix A is defined $\mathrm{herdisc}_\infty(A) := \max_B \mathrm{disc}_\infty(B)$, where B ranges over all matrices obtained by removing a subset of the columns in A. In the terminology of set systems, the hereditary discrepancy is the maximum discrepancy over all set systems obtained by removing a subset of the elements in the universe. We also have an analogous definition for hereditary ℓ_2-discrepancy: $\mathrm{herdisc}_2(A) := \max_B \mathrm{disc}_2(B)$. Based on rounding an SDP, Bansal gave a polynomial time algorithm for finding a coloring x achieving $\mathrm{disc}_\infty(A, x) = O(\lg n \, \mathrm{herdisc}_\infty(A))$. This is quite surprising in light of the strong negative results by Charikar et al. [8], since it shows that is is in fact possible to find a low discrepancy coloring of an arbitrary matrix as long as all its submatrices have low discrepancy.

Our Results Overview. Our main algorithmic result is an ℓ_2 equivalent of Bansal's algorithm with hereditary guarantees. More concretely, we give a polynomial time algorithm for finding a coloring x such that $\mathrm{disc}_2(A, x) = O(\sqrt{\lg n} \cdot \mathrm{herdisc}_2(A))$. We note that neither our result nor Bansal's approximately imply the other: In one direction, the coloring x we find might have very low ℓ_2 discrepancy, but a very large value of $\|Ax\|_\infty$. In the other direction, $\mathrm{herdisc}_\infty(A)$ may be much larger than $\mathrm{herdisc}_2(A)$, thus Bansal's algorithm does not give any guarantees wrt. $\mathrm{herdisc}_2(A)$.

Our algorithm takes a very different approach than Bansal's in the sense that we completely avoid solving an SDP. Instead, we first prove a number of new inequalities relating $\mathrm{herdisc}_2(A)$ and $\mathrm{herdisc}_\infty(A)$ to the eigenvalues of $A^T A$. Relating hereditary discrepancy to the eigenvalues of $A^T A$ was also done by Chazelle and Lvov [10] and by Matoušek et al. [18].

However the result by Chazelle and Lvov is too weak for our applications as it degenerates exponentially fast in the ratio between m and n. The result of Matoušek et al. could be used, but can only show that we find a coloring such that $\text{disc}_2(A, x) = O(\lg^{3/2} n \cdot \text{herdisc}_2(A))$. We believe our new inequalities are of independent interest as strong tools for proving discrepancy lower bounds.

With these inequalities established, we design a simple and efficient deterministic algorithm, inspired by Beck and Fiala's [7] algorithm for sparse set systems. Our key idea is to find a coloring x that is almost orthogonal to all the eigenvectors of $A^T A$ corresponding to large eigenvalues. This in turn means that $\|Ax\|_2$ becomes bounded by $\text{herdisc}_2(A)$.

We now proceed to present the previous results for proving lower bounds on the hereditary discrepancy of matrices in order to set the stage for presenting our new results.

Previous Hereditary Discrepancy Bounds. One of the most useful tools in proving lower bounds for hereditary discrepancy is the determinant lower bound proved in the original paper introducing hereditary discrepancy:

▶ **Theorem 1** (Determinant Lower Bound (Lovász et al. [13])). *For an $m \times n$ real matrix A it holds that*

$$\text{herdisc}_\infty(A) \geq \max_k \max_B \frac{1}{2} |\det(B)|^{1/k},$$

where k ranges over all positive integers up to $\min\{n, m\}$ and B ranges over all $k \times k$ submatrices of A.

While it is easier to bound the max determinant of a submatrix B than it is to bound the discrepancy of a matrix directly, it still requires one to argue that we can find some B where all eigenvalues are non-zero. Chazelle and Lvov demonstrated how it suffices to bound the k'th largest eigenvalue of a matrix in order to derive hereditary discrepancy lower bounds:

▶ **Theorem 2** (Chazelle and Lvov [10]). *For an $m \times n$ real matrix A with $m \leq n$, let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For any integer $k \leq m$, it holds that*

$$\text{herdisc}_\infty(A) \geq \frac{1}{2} 18^{-n/k} \sqrt{\lambda_k}.$$

The result of Chazelle and Lvov has two substantial caveats. First, it requires $m \leq n$. Since we will be using the *partial coloring* framework, we will end up with matrices having very few columns but many rows. This completely rules out using the above result for analysing our new algorithm. Since $k \leq m$, the lower bound also goes down exponentially fast in the gap between m and n (we note that Chazelle and Lvov didn't explicitly state that one needs $k \leq m$, but since $\text{rank}(A) \leq m$, we have $\lambda_k = 0$ whenever $k > m$).

Chazelle and Lvov used their eigenvalue bound to prove the following trace bound which has been very useful in the study of set systems corresponding to incidences between geometric objects:

▶ **Theorem 3** (Trace Bound (Chazelle and Lvov [10])). *For an $m \times n$ real matrix A with $m \leq n$, let $M = A^T A$. Then:*

$$\text{herdisc}_\infty(A) \geq \frac{1}{4} 324^{-n \operatorname{tr} M^2 / \operatorname{tr}^2 M} \sqrt{\operatorname{tr} M / n}.$$

Matoušek et al. [18] presented an alternative to the result of Chazelle and Lvov, relating $\text{herdisc}_\infty(A)$ and $\text{herdisc}_2(A)$ to the sum of singular values of A, i.e. they proved:

▶ **Theorem 4** (Matoušek et al. [18]). *For an $m \times n$ real matrix A, let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. Then*

$$\mathrm{herdisc}_\infty(A) \geq \mathrm{herdisc}_2(A) = \Omega\left(\frac{1}{\lg n}\sum_{k=1}^{n}\sqrt{\frac{\lambda_k}{mn}}\right).$$

which for all positive integers $k \leq \min\{m, n\}$ implies:

$$\mathrm{herdisc}_\infty(A) \geq \mathrm{herdisc}_2(A) = \Omega\left(\frac{k}{\lg n}\sqrt{\frac{\lambda_k}{mn}}\right).$$

Comparing the bound to the result of Chazelle and Lvov, we see that the loss in terms of the ratio between k and n is much better. However for k, m and n all within a constant factor of each other, Chazelle and Lvov's bound implies $\mathrm{herdisc}_\infty(A) = \Omega(\sqrt{\lambda_k})$ whereas the bound of Matoušek et al. loses a $\lg n$ factor and gives $\mathrm{herdisc}_\infty(A) \geq \mathrm{herdisc}_2(A) = \Omega(\sqrt{\lambda_k}/\lg n)$ (strictly speaking, the bound in terms of the sum of $\sqrt{\lambda_k}$'s is incomparable, but the bound only in terms of the k'th largest eigenvalue does lose this factor).

Our Results. We first give a new inequality relating $\mathrm{herdisc}_\infty(A)$ to the eigenvalues of $A^T A$, simultaneously improving over the previous bounds by Chazelle and Lvov, and by Matoušek et al.:

▶ **Theorem 5.** *For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq \min\{n, m\}$, we have*

$$\mathrm{herdisc}_\infty(A) \geq \frac{k}{2e}\sqrt{\frac{\lambda_k}{mn}}.$$

Notice that our lower bound goes down as $k/\sqrt{mn}$ whereas Chazelle and Lvov's goes down as $18^{-n/k}$ and requires $m \leq n$. Thus our loss is exponentially better than theirs. Compared to the bound by Matoušek et al., we avoid the $\lg n$ loss (at least compared to the bound of Matoušek et al. that is only in terms of the k'th largest eigenvalue and not the sum of eigenvalues).

Re-executing Chazelle and Lvov's proof of the trace bound with the above lemma in place of theirs immediately gives a stronger version of the trace bound as well:

▶ **Corollary 6.** *For an $m \times n$ real matrix A, let $M = A^T A$. Then:*

$$\mathrm{herdisc}_\infty(A) \geq \frac{\mathrm{tr}^2 M}{8e\min\{n, m\}\,\mathrm{tr}\, M^2}\sqrt{\frac{\mathrm{tr}\, M}{\max\{m, n\}}}.$$

In establishing lower bounds on $\mathrm{herdisc}_2(A)$ in terms of eigenvalues, we need to first prove an equivalent of the determinant lower bound for non-square matrices (and for ℓ_2-discrepancy rather than ℓ_∞):

▶ **Theorem 7.** *For an $m \times n$ real matrix A, we have*

$$\mathrm{herdisc}_\infty(A) \geq \mathrm{herdisc}_2(A) \geq \sqrt{\frac{n}{8\pi e m}}\det(A^T A)^{1/2n}.$$

We remark that proving Theorem 7 for the ℓ_∞-case appears as an exercise in [16] and we make no claim that the proof of Theorem 7 requires any new or deep insights (we suspect that it is folklore, but have not been able to find a mentioning of the above theorem in the literature). We finally arrive at our main result for lower bounding hereditary ℓ_2-discrepancy:

▶ **Corollary 8.** *For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq \min\{n, m\}$, we have*

$$\text{herdisc}_2(A) \geq \frac{k}{e}\sqrt{\frac{\lambda_k}{8\pi mn}}.$$

We note that Theorem 5 actually follows (up to constant factors) from Corollary 8 using the fact that $\text{herdisc}_\infty(A) \geq \text{herdisc}_2(A)$, but we will present separate proofs of the two theorems since the direct proof of Theorem 5 is very short and crisp.

The exciting part in having established Corollary 8, is that it hints the direction for giving an efficient algorithm for obtaining colorings x with $\text{disc}_2(A, x)$ being bounded by some function of $\text{herdisc}_2(A)$. More concretely, we give an algorithm that is based on computing an eigendecomposition of $A^T A$ and using this to perform partial coloring that is orthogonal to the eigenvectors corresponding to the largest eigenvalues. Via Corollary 8, this gives a coloring with hereditary ℓ_2 guarantees. The precise guarantees of our algorithm are given in the following:

▶ **Theorem 9.** *There is an $O((m + n)n^2)$ time algorithm that given an $m \times n$ matrix A, computes a coloring $x \in \{-1, +1\}^n$ satisfying $\text{disc}_2(A, x) = O(\sqrt{\lg n} \cdot \text{herdisc}_2(A))$.*

We implemented our algorithm and performed various experiments to examine its practical performance. Section 4 shows that the algorithm far outperforms random sampling a coloring $x \in \{-1, +1\}^n$. In fact, it far outperforms random sampling, even if we repeatedly sample vectors for as long time as our algorithm runs and use the best one sampled. Moreover, the algorithm is efficient enough that it can be run on 1000×1000 matrices in less than 10 seconds and on matrices of sizes up to 10000×10000 in about 4 hours on a standard laptop. While it is conceivable that Bansal's SDP based approach can be modified to give ℓ_2 guarantees with a polynomial running time, it seems highly unlikely that it can process such large matrices in a reasonable amount of time. Moreover, our algorithm is much simpler to analyse and implement.

2 Eigenvalue Bounds for Hereditary Discrepancy

In this section, we prove new results relating the hereditary discrepancy of a matrix A to the eigenvalues of $A^T A$. The section is split in two parts, one studying hereditary ℓ_∞-discrepancy and one studying hereditary ℓ_2-discrepancy.

2.1 Hereditary ℓ_∞-discrepancy

Our first result concerns hereditary ℓ_∞-discrepancy and is a strengthening of the previous bound due to Chazelle and Lvov [10] (see Section 1). The simplest formulation is the following:

▷ Restatement of Theorem 5. For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq \min\{n, m\}$, we have

$$\text{herdisc}_\infty(A) \geq \frac{k}{2e}\sqrt{\frac{\lambda_k}{mn}}.$$

Theorem 5 is an immediate corollary of the following slightly more general result:

▶ **Theorem 10.** *For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq \min\{n, m\}$, we have*

$$\mathrm{herdisc}_\infty(A) \geq \frac{1}{2} \left(\frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}\binom{m}{k}} \right)^{1/2k}$$

Theorem 5 follows from Theorem 10 by using that $\binom{n}{k} \leq (en/k)^k$ and that $\prod_{i=1}^k \lambda_i \geq \lambda_k^k$. Thus our goal is to prove Theorem 10. The first step of our proof uses the following linear algebraic fact:

▶ **Lemma 11.** *For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq n$, there exists an $m \times k$ submatrix C of A such that $\det(C^T C) \geq (\prod_{i=1}^k \lambda_i)/\binom{n}{k}$.*

Proof. The k'th symmetric function of $\lambda_1, \ldots, \lambda_n$ is defined as (see e.g. the textbook [19] p. 494): $s_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \lambda_{i_1} \cdots \lambda_{i_k}$. Since all λ_i are non-negative, we have $s_k \geq \prod_{i=1}^k \lambda_i$. If we let $\mathcal{S}_k(A^T A)$ denote the set of all $k \times k$ principal submatrices of $A^T A$, then it also holds that (see e.g. the textbook [19] p. 494): $s_k = \sum_{B \in \mathcal{S}_k(A^T A)} \det(B)$. Since $|\mathcal{S}_k(A^T A)| = \binom{n}{k}$ there must be a $B \in \mathcal{S}_k(A^T A)$ for which $\det(B) \geq \left(\prod_{i=1}^k \lambda_i\right)/\binom{n}{k}$. Since B is a $k \times k$ principal submatrix of $A^T A$, it follows that there exists an $m \times k$ submatrix C of A such that $B = C^T C$ and thus $\det(C^T C) \geq \left(\prod_{i=1}^k \lambda_i\right)/\binom{n}{k}$. ◀

With Lemma 11 established, we are ready to present the proof of Theorem 10:

Proof of Theorem 10. Let A be a real $m \times n$ matrix and let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. From Lemma 11, it follows that for every $k \leq n$, there is an $m \times k$ submatrix C of A such that $\det(C^T C) \geq (\prod_{i=1}^k \lambda_i)/\binom{n}{k}$. If we also have $k \leq m$, we can let $\mathcal{S}_k(C)$ denote the set of all $k \times k$ principal submatrices of C and use the Cauchy-Binet formula to conclude that: $\det(C^T C) = \sum_{D \in \mathcal{S}_k(C)} \det(D)^2$. But $\mathcal{S}_k(C) \subseteq \mathcal{S}_k(A)$ hence there must exist a $k \times k$ matrix $D \in \mathcal{S}_k(A)$ such that

$$\det(D)^2 \geq \frac{\det(C^T C)}{|\mathcal{S}_k(C)|} \geq \frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}\binom{m}{k}} \Rightarrow |\det(D)| \geq \sqrt{\frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}\binom{m}{k}}}.$$

It follows from the determinant lower bound for hereditary discrepancy (Theorem 1) that

$$\mathrm{herdisc}_\infty(A) \geq \frac{1}{2}|\det(D)|^{1/k} \geq \frac{1}{2} \left(\frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}\binom{m}{k}} \right)^{1/2k}.$$

◀

Having established a stronger connection between eigenvalues and hereditary discrepancy than the one given by Chazelle and Lvov [10], we can also re-execute their proof of the trace bound and obtain the following strengthening:

▷ Restatement of Corollary 6. For an $m \times n$ real matrix A, let $M = A^T A$. Then:

$$\mathrm{herdisc}_\infty(A) \geq \frac{\mathrm{tr}^2 M}{8e \min\{n, m\} \,\mathrm{tr}\, M^2} \sqrt{\frac{\mathrm{tr}\, M}{\max\{m, n\}}}.$$

Proof. Let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of M. Chazelle and Lvov [10] proved that if we choose $k = \operatorname{tr}^2 M/(2 \operatorname{tr} M^2)$ then $\lambda_k \geq \operatorname{tr} M/(4n)$. Examining their proof, one can in fact strengthen it slightly to $\lambda_k \geq \operatorname{tr} M/(4 \min\{m,n\})$ (their proof of ([10] Lemma 2.4) considers a uniform random eigenvalue λ amongst $\lambda_1, \ldots, \lambda_n$ and uses that $\operatorname{tr} M = n\mathbb{E}[\lambda]$. However, one needs only λ to be uniform random amongst the non-zero eigenvalues and there are at most $\min\{m,n\}$ such eigenvalues yielding $\operatorname{tr} M = \min\{n,m\}\mathbb{E}[\lambda]$). Inserting these bounds in Theorem 5 gives us

$$\operatorname{herdisc}_\infty(A) \geq \frac{\operatorname{tr}^2 M}{8e \operatorname{tr} M^2}\sqrt{\frac{\operatorname{tr} M}{mn \min\{m,n\}}} = \frac{\operatorname{tr}^2 M}{8e \min\{n,m\} \operatorname{tr} M^2}\sqrt{\frac{\operatorname{tr} M}{\max\{m,n\}}}.$$

◀

2.2 Hereditary ℓ_2-discrepancy

This section proves the following determinant result for hereditary ℓ_2-discrepancy of $m \times n$ matrices:

▷ Restatement of Theorem 7. For an $m \times n$ real matrix A with $\det(A^T A) \neq 0$, we have

$$\operatorname{herdisc}_\infty(A) \geq \operatorname{herdisc}_2(A) \geq \sqrt{\frac{nm}{8\pi e}} \det(A^T A)^{1/2n}.$$

The fact $\operatorname{herdisc}_\infty(A) \geq \operatorname{herdisc}_2(A)$ is true for all A, thus the difficulty in proving Theorem 7 lies in establishing that $\operatorname{herdisc}_2(A) \geq \sqrt{nm/(8\pi e)} \det(A^T A)^{1/2n}$. Our proof uses many of the ideas from the proof of the determinant lower bound (Theorem 1) in [13]. We start by introducing the linear discrepancy in the ℓ_2 setting and summarize known relations between linear discrepancy and hereditary discrepancy.

▶ **Definition 12.** *Let A be an $m \times n$ real matrix. Then its linear ℓ_2-discrepancy is defined as:*

$$\operatorname{lindisc}_2(A) := \max_{c \in [-1,+1]} \min_{x \in \{-1,+1\}^n} \frac{1}{\sqrt{m}} \|A(x-c)\|_2.$$

The linear ℓ_2-discrepancy has a clean geometric interpretation (this is a direct translation of the similar interpretation of linear ℓ_∞-discrepancy given e.g. in [13, 16]). For an $m \times n$ real matrix A, let: $U_A := \{x : \|Ax\|_2 \leq \sqrt{m}\}$. For $t > 0$, place 2^n translated copies $U_1, \ldots, U_{2^n}$ of tU_A such that there is one copy centered at each point in $\{-1,+1\}^n$. Then $\operatorname{lindisc}_2(A)$ is the least number t for which the sets U_j cover all of $[-1,+1]^n$.

We will need the following relationship between the hereditary and linear discrepancy:

▶ **Lemma 13** (Lovász et al. [13]). *For all $m \times n$ real matrices A, it holds that* $\operatorname{lindisc}_2(A) \leq 2 \operatorname{herdisc}_2(A)$.

We remark that [13] proved Lemma 13 only for the ℓ_∞-discrepancy, but their proof only uses the fact that $\{x : \|Ax\|_\infty \leq 1\}$ is centrally symmetric and convex (see [13] Lemma 1). The same is true for the U_A defined above.

In light of Lemma 13, we set out to lower bound the linear discrepancy of an $m \times n$ matrix A in terms of $\det(A^T A)$. We will prove the following lemma using an adaptation of the ideas in [13] (we have not been able to find a proof of this result elsewhere, but remark that the case of $m = n$ should follow by adapting the proof in [13]):

▶ **Lemma 14.** *Let A be an $m \times n$ real matrix with $\det(A^T A) \neq 0$. Then* $\operatorname{lindisc}_2(A) \geq \sqrt{n/(2\pi e m)} \det(A^T A)^{1/2n}$.

Proof. From the geometric interpretation given earlier, we know that if we place a copy of $\text{lindisc}_2(A)U_A$ on each point in $\{-1,+1\}^n$, then they cover all of $[-1,1]^n$ hence $\text{vol}(\text{lindisc}_2(A)U_A) \geq \text{vol}([-1,1]^n)/2^n = 1$. But

$$\begin{aligned}\text{vol}(\text{lindisc}_2(A)U_A) &= (\text{lindisc}_2(A))^n \,\text{vol}(U_A)\\ &= (\text{lindisc}_2(A))^n \,\text{vol}(\{x : \|Ax\|_2 \leq \sqrt{m}\})\\ &= (\text{lindisc}_2(A))^n \,\text{vol}(\{x : x^T A^T A x \leq m\}).\end{aligned}$$

Observe now that $\{x : x^T A^T A x \leq m\} = \{x : x^T(m^{-1}A^T A)x \leq 1\}$ is an ellipsoid. It is well-known that the volume of such an ellipsoid equals $v_n/\sqrt{\det(m^{-1}A^T A)} = v_n/\sqrt{m^{-n}\det(A^T A)}$ where v_n is the volume of the n-dimensional ℓ_2 unit ball. Since $v_n = \pi^{n/2}/\Gamma(n/2+1) \leq (2\pi e/n)^{n/2}$, we conclude:

$$\begin{aligned}1 &\leq \frac{(\text{lindisc}_2(A))^n v_n}{\sqrt{m^{-n}\det(A^T A)}} \Rightarrow\\ 1 &\leq (\text{lindisc}_2(A))^n \left(\frac{2\pi e m}{n}\right)^{n/2} \frac{1}{\sqrt{\det(A^T A)}} \Rightarrow\\ \text{lindisc}_2(A) &\geq \sqrt{\frac{n}{2\pi e m}}\det(A^T A)^{1/2n}.\end{aligned}$$

◀

Combining Lemma 13 and Lemma 14 proves Theorem 7.

Having establishes Theorem 7, we are ready to prove our last result on hereditary ℓ_2-discrepancy:

▷ Restatement of Corollary 8. For an $m \times n$ real matrix A, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \leq \min\{n, m\}$, we have $\text{herdisc}_2(A) \geq (k/e)\sqrt{\lambda_k/(8\pi m n)}$.

Proof. Let A be an $m \times n$ real matrix and let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of $A^T A$. From Lemma 11, we know that for all $k \leq n$, there is an $m \times k$ submatrix C of A such that $\det(C^T C) \geq (\prod_{i=1}^k \lambda_i)/\binom{n}{k} \geq (k\lambda_k/(en))^k$. From Theorem 7, we get that $\text{herdisc}_2(C) \geq \sqrt{k/(8\pi e m)}\det(C^T C)^{1/2k} \geq (k/e)\sqrt{\lambda_k/(8\pi m n)}$. Since C is obtained from A by deleting a subset of the columns, it follows that $\text{herdisc}_2(A) \geq \text{herdisc}_2(C)$, completing the proof. ◀

3 Discrepancy Minimization with Hereditary ℓ_2 Guarantees

This section gives our new algorithm for discrepancy minimization. The goal is to prove the following:

▷ Restatement of Theorem 9. There is an $O((m+n)n^2)$ time algorithm that given an $m \times n$ matrix A, computes a coloring $x \in \{-1,+1\}^n$ satisfying $\text{disc}_2(A, x) = O(\sqrt{\lg n} \cdot \text{herdisc}_2(A))$.

Our algorithm follows the same overall approach as several previous algorithms. The general setup is that we first give a procedure for partial coloring. This procedure takes a matrix A and a partial coloring $x \in [-1,+1]^n$. We say that coordinates i of x such that $|x_i| < 1$ are *live*. If there are k live coordinates prior to calling the partial coloring method, then upon termination we get a new vector γ such that the number of live coordinates in $\hat{x} = x + \gamma$ is no more than $k/2$. At the same time, all coordinates of $\hat{x}$ are bounded by 1 in absolute value and $\|A\hat{x}\|_2$ is not much larger than $\|Ax\|_2$.

We start by presenting the partial coloring algorithm and then show how to use it to get the final coloring.

3.1 Partial Coloring

In this section, we present our partial coloring algorithm. The algorithm takes as input an $m \times n$ matrix A and a vector $x \in [-1,+1]^n$. We think of the vector x as a partial coloring. We call a coordinate x_i of x *live* if $|x_i| < 1$ and we let k denote the number of live coordinates in x. For ease of notation, we let $\text{live}_x(i)$ denote the index of the i'th live coordinate in x and we define $\oplus_x : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ as the function such that $a \oplus_x b$ for $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^k$, is the vector obtained from a by adding the i'th coordinate of b to the coordinate of index $\text{live}_x(i)$ in a (where $\text{live}_x(i)$ refers to the i'th live coordinate in x).

Upon termination, the algorithm returns another vector $\gamma \in \mathbb{R}^k$. If we let $\hat{x} = x \oplus_x \gamma$ be the vector in $\mathbb{R}^n$ obtained from x by adding γ_i to $x_{\text{live}_x(i)}$, then the partial coloring algorithm guarantees the following:

1. There are at most $k/2$ live coordinates in $\hat{x}$.
2. For all i, we have $|\hat{x}_i| \le 1$.
3. $\|A\hat{x}\|_2^2 - \|Ax\|_2^2 = O(m(\text{herdisc}_2(A))^2)$.

Thus upon termination, the new vector $\hat{x}$ has half as many live coordinates, and the discrepancy did not increase by much. In particular the change is related to the hereditary ℓ_2-discrepancy of A.

The main idea in our algorithm is to use the connection between eigenvalues and hereditary ℓ_2-discrepancy that we proved in Corollary 8. Our algorithm proceeds in iterations, where in each step it finds a vector v and adds it to γ. The way we choose v is roughly to find the eigenvectors of $A^T A$ and then pick v orthogonal to the eigenvectors corresponding to the largest eigenvalues. This bounds the difference $\|A(x \oplus_x (\gamma + v))\|_2 - \|A(x \oplus_x \gamma)\|_2$ in terms of the eigenvalues and thus hereditary ℓ_2-discrepancy. At the same time, we use the ideas by Beck and Fiala (and many later papers) where we include constraints forcing v orthogonal to e_i for every coordinate i that is not live. The algorithm is as follows:

PartialColor(A, x):

1. Let k denote the number of live coordinates in x and let C denote the $m \times k$ matrix obtained from A by deleting all columns corresponding to coordinates that are not live.
2. Initialize $\gamma = \mathbf{0} \in \mathbb{R}^k$.
3. Compute an eigendecomposition of $C^T C$ to obtain the eigenvalues $\lambda_1 \ge \cdots \ge \lambda_k \ge 0$ and corresponding eigenvectors $\mu_1, \ldots, \mu_k$.
4. While **True**:
   a. Compute the set S of coordinates i such that $|\gamma_i + x_{\text{live}_x(i)}| = 1$. If $|S| \ge k/2$, **return** γ.
   b. Find a unit vector v orthogonal to all e_j with $j \in S$ and to all μ_i with $i \le k/4$.
   c. Let $\sigma = -\text{sign}(\langle Ax, A(\mathbf{0} \oplus_x v)\rangle)$. Compute the largest $\beta > 0$ such that all coordinates of $x \oplus_x (\gamma + \sigma\beta v)$ are less than or equal to 1 in absolute value. Update $\gamma \leftarrow \gamma + \sigma\beta v$.

Correctness. We prove that the vector γ returned by the above **PartialColor** algorithm satisfies the three claimed properties. First observe that in every iteration of the while loop, we find a vector v that is orthogonal to e_i whenever $|\gamma_i + x_{\text{live}_x(i)}| = 1$. Hence if $|\gamma_i + x_{\text{live}_x(i)}|$ becomes 1, it never changes again. Moreover, by maximizing β in each iteration, we guarantee that at least one more coordinate satisfies $|\gamma_i + x_{\text{live}_x(i)}| = 1$ after every iteration. Thus the algorithm terminates after at most $k/2$ iterations of the while loop and no coordinate of $x \oplus_x \gamma$ is larger than 1 in absolute value. What remains is to bound $\|A(x \oplus_x \gamma)\|_2^2 - \|Ax\|_2^2$.

Let $v^{(i)}$ denote the vector v found during the i'th iteration of the while loop. Upon termination, we have that $\gamma = \sigma_1\beta_1 v^{(1)} + \cdots + \sigma_r\beta_r v^{(r)}$ where $\sigma_i = -\operatorname{sign}(\langle Ax, v^{(i)}\rangle)$ and each $v^{(i)}$ is orthogonal to $\mu_1, \ldots, \mu_{k/4}$. Thus γ is also orthogonal to $\mu_1, \ldots, \mu_{k/4}$. We therefore have:

$$
\begin{aligned}
\|A(x \oplus_x \gamma)\|_2^2 &= \|A(x + (\mathbf{0} \oplus_x \gamma))\|_2^2 \\
&\leq \|Ax\|_2^2 + \|A(\mathbf{0} \oplus_x \gamma)\|_2^2 + 2\langle Ax, A(\mathbf{0} \oplus_x \gamma)\rangle \\
&= \|Ax\|_2^2 + \|C\gamma\|_2^2 + 2\sum_{i=1}^{r}\langle Ax, A(\mathbf{0} \oplus_x \sigma_i\beta_i v^{(i)})\rangle \\
&\leq \|Ax\|_2^2 + \lambda_{k/4}\|\gamma\|_2^2 - 2\sum_{i=1}^{r}\operatorname{sign}(\langle Ax, A(\mathbf{0} \oplus_x v^{(i)})\rangle)\langle Ax, A(\mathbf{0} \oplus_x \beta_i v^{(i)})\rangle \\
&= \|Ax\|_2^2 + \lambda_{k/4}\|\gamma\|_2^2 - 2\sum_{i=1}^{r}\operatorname{sign}(\langle Ax, A(\mathbf{0} \oplus_x v^{(i)})\rangle)^2|\langle Ax, A(\mathbf{0} \oplus_x \beta_i v^{(i)})\rangle| \\
&\leq \|Ax\|_2^2 + \|\gamma\|_\infty^2 k\lambda_{k/4} - 0 \\
&\leq \|Ax\|_2^2 + 4k\lambda_{k/4}.
\end{aligned}
$$

We would like to use Corollary 8 to relate $k\lambda_{k/4}$ to the hereditary discrepancy of A. Since C is an $m \times k$ submatrix of A, we have $\operatorname{herdisc}_2(A) \geq \operatorname{herdisc}_2(C)$. Using Corollary 8 we have $\operatorname{herdisc}_2(C) \geq (k/4e)\sqrt{\lambda_{k/4}/mk} = (1/4e)\sqrt{k\lambda_{k/4}/(8\pi)m}$. Hence we conclude that

$$\|A\hat{x}\|_2^2 - \|Ax\|_2^2 \leq 128e^2\pi m(\operatorname{herdisc}_2(A))^2 = O(m(\operatorname{herdisc}_2(A))^2).$$

Running Time. Step 1. of **PartialColor** takes $O(mk)$ time and step 2. takes $O(k)$. Step 3. takes $O(mk^2)$ time to compute C^TC (can be improved via fast matrix multiplication) and $O(k^3)$ time to compute the eigendecomposition. As argued above, each iteration of the while loop increases the size of S by at least one. Hence there are no more than $k/2$ iterations of the loop. Computing S in step (a) takes $O(k)$ time. Finding the unit vector v in step (b) can be done in $O(k^2)$ time as follows: Whenever adding a coordinate i to S, use Gram-Schmidt to compute the normalized (unit-norm) projection $\hat{e}_i$ of e_i onto the orthogonal complement of $\mu_1, \ldots, \mu_{k/4}$ and all previous vectors $\hat{e}_j$. This takes $O(k^2)$ time per i. To find v, sample a uniform random unit vector in $\mathbb{R}^k$ and run Gram-Schmidt to compute its projection onto the orthogonal complement of $\hat{e}_j$ for $j \in S$ and $\mu_1, \ldots, \mu_{k/4}$. The expected length of the projection is $\Omega(1)$ and we can scale it to unit length afterwards. This gives the desired vector. The Gram-Schmidt step takes $O(k^2)$ time. Computing $A(\mathbf{0} \oplus_x v)$ in step (c) takes $O(mk)$ time and computing Ax can be done outside the while loop in $O(mn)$ time. The inner product takes $O(m)$ time to compute. Computing β and adding $\sigma\beta v$ to γ takes $O(k)$ time. Overall, the **PartialColor** algorithm takes $O(mn + mk^2 + k^3)$ time. If Ax is given as argument to the algorithm, the time is further reduced to $O((m + k)k^2)$.

3.2 The Final Algorithm

Now that we have the **PartialColor** algorithm, getting to a low discrepancy coloring is straight forward. Given an $m \times n$ matrix A, we initialize $x \leftarrow \mathbf{0}$. We then repeatedly invoke **PartialColor**(A, x). Each call returns a vector γ. We update $x \leftarrow x + \gamma$ and continue. We stop once there are no live coordinates in x, i.e. all coordinates satisfy $|x_i| = 1$.

In each iteration, the number of live coordinates of i decreases by at least a factor two, and thus we are done after at most $\lg n$ iterations. This means that the final vector x satisfies

$$\begin{aligned} \|Ax\|_2^2 &\leq \lg n \cdot O(m(\text{herdisc}_2(A))^2) \Rightarrow \\ \|Ax\|_2 &= O(\sqrt{m \lg n} \cdot \text{herdisc}_2(A)) \Rightarrow \\ \text{disc}_2(A, x) &= O(\sqrt{\lg n} \cdot \text{herdisc}_2(A)). \end{aligned}$$

For the running time, observe that after each call to **PartialColor**, we can compute $A(x+\gamma)$ from Ax in $O(mk)$ time. Thus we can provide Ax as argument to **PartialColor** and thereby reduce its running time to $O((m+k)k^2)$. Since k halves in each iteration, we get a running time of

$$O\left(\sum_{i=1}^{\lg n}(m+n/2^i)(n/2^i)^2\right) = O((m+n)n^2).$$

This concludes the proof of Theorem 9.

4 Experiments

In this section, we present a number of experiments to test the practical performance of our discrepancy minimization algorithm. We denote the algorithm by L2Minimize in the following. We compare it to two base line algorithms Sample and SampleMany. Sample simply picks a uniform random $\{-1,+1\}$ vector as its coloring. SampleMany repeatedly samples a uniform random $\{-1,+1\}$ vector and runs for the same amount of time as L2Minimize. It returns the best vector found within the time limit.

The algorithms were implemented in Python, using NumPy and SciPy for linear algebra operations. All tests were run on a MacBook Pro (15-inch, Late 2013) running macOS Sierra 10.13.3. The machine has a 2 GHz Intel Core i7 and 8GB DDR3 RAM.

We tested the algorithms on three different classes of matrices:

- **Uniform** matrices: Each coordinate is uniform random and independently chosen among -1 and $+1$.
- **2D Corner** matrices: Obtained by sampling two sets $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_m\}$ of n and m points in the plane, respectively. The points are sampled uniformly in the $[0,1] \times [0,1]$ unit square. The resulting matrix has one column per point $p_j \in P$ and one row per point $q_i \in Q$. The entry (i,j) is 1 if p_j is *dominated* by q_i, i.e. $q_i.x > p_j.x$ and $q_i.y > p_j.y$ and it is 0 otherwise. Such matrices are known to have hereditary ℓ_2-discrepancy $O(\lg^{1.5} n)$ [20].
- **2D Halfspace** matrices: Obtained by sampling a set $P = \{p_1, \dots, p_n\}$ of n points in the unit square $[0,1] \times [0,1]$, and a set Q of m halfspace. Each halfspace in Q is sampled by picking one point a uniformly on either the left boundary of the unit square or on the top boundary, and another point b uniformly on either the right boundary or the bottom boundary of the unit square. The halfspace is then chosen uniformly to be either everything above the line through a, b or everything below it. The resulting matrix has one column per point $p_j \in P$ and one row per halfspace $h_i \in Q$. The entry (i,j) is 1 if p_j is in the halfspace h_i and it is 0 otherwise. Such matrices are known to have hereditary ℓ_2-discrepancy $O(n^{1/4})$ [15].

Each test is run 10 times and the average ℓ_2 discrepancy and average runtime is reported. The running times of the algorithms varied exclusively with the matrix size and not the type of matrix, thus we only show one time column which is representative of all types of matrices. The results are shown in Table 1.

Algorithm	Matrix Size	Disc Uniform	Disc 2D Corner	Disc 2D Halfspace	Time (s)
L2Minimize	200×200	7.2	1.8	1.6	< 1
Sample	200×200	13.8	7.6	11.0	< 1
SampleMany	200×200	11.6	2.3	2.7	< 1
L2Minimize	1000×1000	15.7	1.9	2.3	9
Sample	1000×1000	31.6	16.0	18.3	< 1
SampleMany	1000×1000	28.9	4.9	5.5	9
L2Minimize	4000×4000	31.0	2.1	2.6	717
Sample	4000×4000	63.1	21.0	34.0	< 1
SampleMany	4000×4000	60.3	9.5	10.7	717
L2Minimize	10000×10000	48.3	2.1	3.1	15260
Sample	10000×10000	99.9	51.4	96.8	< 1
SampleMany	10000×10000	96.8	14.2	15.6	15260
L2Minimize	10000×2000	35.9	2.1	2.7	535
Sample	10000×2000	44.7	20.6	24.1	< 1
SampleMany	10000×2000	43.4	6.7	8.0	535
L2Minimize	2000×10000	21.4	1.8	2.0	5809
Sample	2000×10000	99.9	40.8	70.8	< 1
SampleMany	2000×10000	92.2	13.8	16.4	5809

Table 1 Results of experiments with our L2Minimize algorithm. The Matrix Size column gives the size $m \times n$ of the input matrix. The Disc columns shows $\text{disc}_2(A, x) = \|Ax\|_2/\sqrt{m}$ for the coloring x found by the algorithm on the given type of matrix. Time is measured in seconds. Each entry is the average of 10 executions.

The table clearly shows that L2Minimize gives superior colorings for all types of matrices and all sizes. The tendency is particularly clear on the structured matrices **2D Corner** and **2D Halfspace** where the coloring found by L2Minimize on 10000×10000 matrices is a factor 25-30 smaller than a single round of random sampling (Sample) and a factor 5-7 better than random sampling for as long time as L2Minimize runs (SampleMany).

The $O((m+n)n^2)$ running time makes the algorithm practical up to matrices of size about 10000×10000, at which point the algorithm runs for 15260 seconds $\approx$ 4 hours and 15 minutes.

5 Acknowledgment

The author wishes to thank Nikhil Bansal for useful discussions and pointers to relevant literature. The author also thanks an anonymous STOC reviewer for comments that simplified the **PartialColor** algorithm.

References

1 R. Alexander. Geometric methods in the study of irregularities of distribution. *Combinatorica*, 10(2):115–136, 1990.

2 Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12:351–360, July 1998.

3 Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Proc. 51th Annual IEEE Symposium on Foundations of Computer Science (FOCS'10)*, pages 3–10, 2010.

4 Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An algorithm for komlós conjecture matching banaszczyk's bound. In *Proc. 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS'16)*, pages 788–799, 2016.

5 Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The gram-schmidt walk: A cure for the banaszczyk blues. *CoRR*, abs/1708.01079, 2017. URL: `http://arxiv.org/abs/1708.01079`.

6 Nikhil Bansal and Shashwat Garg. Algorithmic discrepancy beyond partial coloring. In *Proc. 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, STOC 2017, pages 914–926, 2017.

7 J. Beck and T. Fiala. Integer-making theorems. *Discrete Applied Mathematics*, 3:1–8, February 1981.

8 Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In *Proc. 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 1607–1614, 2011.

9 Bernard Chazelle. *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2000.

10 Bernard Chazelle and Alexey Lvov. A trace bound for the hereditary discrepancy. In *Proc. 16th Annual Symposium on Computational Geometry*, SCG '00, pages 64–69, 2000.

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

12 Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. Deterministic discrepancy minimization via the multiplicative weight update method. In *Integer Programming and Combinatorial Optimization - 19th International Conference, IPCO 2017, Waterloo, ON, Canada, June 26-28, 2017, Proceedings*, pages 380–391, 2017.

13 L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151 – 160, 1986. `doi:https://doi.org/10.1016/S0195-6698(86)80041-5`.

14 Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.

15 J. Matoušek. Tight upper bounds for the discrepancy of half-spaces. *Discrete and Computational Geometry*, 13:593–601, 1995.

16 J. Matousek. *Geometric Discrepancy: An Illustrated Guide*. Algorithms and Combinatorics. Springer Berlin Heidelberg, 1999.

17 Jirí Matoušek and Aleksandar Nikolov. Combinatorial Discrepancy for Boxes via the gamma_2 Norm. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34, pages 1–15, 2015.

18 Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *CoRR*, abs/1408.1376, 2014. URL: `http://arxiv.org/abs/1408.1376`.

19 Carl D. Meyer, editor. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

20 Aleksandar Nikolov. Tighter bounds for the discrepancy of boxes and polytopes. *Mathematika*, 63:1091–1113, 2017.

21 Joel Spencer. Six standard deviations suffice. *Trans. Amer. Math. Soc.*, 289:679–706, 1985.