Constructive Discrepancy Minimization with Hereditary L2 Guarantees

Kasper Green Larsen*

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 $\mathrm{disc}_\infty(\mathcal{S}, \chi) := \max_{S \in \mathcal{S}} |\sum_{u_i \in S} \chi(u_i)|$ and ℓ_2-discrepancy defined as $\mathrm{disc}_2(\mathcal{S}, \chi) := \sqrt{(1/|\mathcal{S}|)\sum_{S \in \mathcal{S}} \left(\sum_{u_i \in S} \chi(u_i)\right)^2}$. Breakthrough work by Bansal [FOCS'10] gave a polynomial time algorithm, based on rounding an SDP, for finding a coloring χ such that $\mathrm{disc}_\infty(\mathcal{S}, \chi) = O(\lg n \cdot \mathrm{herdisc}_\infty(\mathcal{S}))$ where $\mathrm{herdisc}_\infty(\mathcal{S})$ is the hereditary ℓ_∞-discrepancy of $\mathcal{S}$. We complement his work by giving a polynomial time algorithm for finding a coloring χ such $\mathrm{disc}_2(\mathcal{S}, \chi) = O(\sqrt{\lg n} \cdot \mathrm{herdisc}_2(\mathcal{S}))$ where $\mathrm{herdisc}_2(\mathcal{S})$ 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 $\mathrm{herdisc}_\infty$ and $\mathrm{herdisc}_2$ 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+SoCG'15]. We believe these inequalities are of independent interest as powerful tools for proving hereditary discrepancy lower bounds.

*MADALGO. Aarhus University. `larsen@cs.au.dk`. 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 [15]. 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 $\mathrm{disc}_\infty(A, x) := \|Ax\|_\infty$ or the ℓ_2-discrepancy $\mathrm{disc}_2(A, x) = (1/\sqrt{m})\|Ax\|_2$. We say that the ℓ_∞-discrepancy of A is $\mathrm{disc}_\infty(A) := \min_{x \in \{-1,+1\}^n} \mathrm{disc}_\infty(A, x)$ and the ℓ_2-discrepancy of A is $\mathrm{disc}_2(A) := \min_{x \in \{-1,+1\}^n} \mathrm{disc}_2(A, x)$. With this matrix view, it is clear that discrepancy minimization makes sense also for general matrices and not just ones arising from set systems.

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. [16, 14, 1, 11]. Another fruitful line of research has focused on general matrices, including the celebrated "Six Standard Devitations Suffice" result by Spencer [19], 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 an 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 [19] 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 [13] gave simpler algorithms achieving the same. 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. [12] 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. [17]. 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 $\mathrm{disc}_2(A, x) = O(\lg^{3/2} n \cdot \mathrm{herdisc}_2(A))$. We believe our new inequalities are of independent interest as strong tools for proving discrepancy lower bounds.

With these inequalities established, we build on the ideas in the Edge-Walk algorithm by Lovett and Meka [13] 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 $\mathrm{herdisc}_2(A)$.

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. [12])). *For an $m \times n$ real matrix A it holds that*

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

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*

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

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 $\mathrm{rank}(A) \leq m$, 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:*

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

Matoušek et al. [17] presented an alternative to the result of Chazelle and Lvov, relating $\mathrm{herdisc}_\infty(A)$ and $\mathrm{herdisc}_2(A)$ to the sum of singular values of A, i.e. they proved:

Theorem 4 (Matoušek et al. [17]). *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 $k \le \min\{m, n\}$ implies:

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

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 $\mathrm{herdisc}_\infty(A) \ge \mathrm{herdisc}_2(A) = \Omega(\sqrt{\lambda_k}/\lg n)$ (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 $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \le \min\{n, m\}$, we have*

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

Notice that our lower bound goes down as $k/\sqrt{mn}$ whereas Chazelle and Lvov's goes down as $18^{-n/k}$ and requires $m \le n$. 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 1. *For an $m \times n$ real matrix A, let $M = A^T A$. Then:*

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

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 6. *For an $m \times n$ real matrix A, we have*

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

We remark that proving Theorem 6 for the ℓ_∞-case appears as an exercise in [15] and we make no claim that the proof of Theorem 6 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 2. *For an $m \times n$ real matrix A, let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$ denote the eigenvalues of $A^T A$. For all positive integers $k \le \min\{n, m\}$, we have*

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

We note that Theorem 5 actually follows (up to constant factors) from Corollary 2 using the fact that $\mathrm{herdisc}_\infty(A) \ge \mathrm{herdisc}_2(A)$, but we will present separate proofs of the two theorems since the direct proof of Theorem 5 is very short and instructive.

The exciting part in having established Corollary 2, is that it hints the direction for giving a polynomial time algorithm for obtaining colorings x with $\mathrm{disc}_2(A, x)$ being bounded by some function of $\mathrm{herdisc}_2(A)$. More concretely, we give an algorithm that is based on computing an eigendecomposition of $A^T A$ and using this to perform partial coloring via a "random walk" that is orthogonal to the eigenvectors corresponding to the largest eigenvalues. Via Corollary 2, this gives a coloring with hereditary ℓ_2 guarantees. The precise guarantees of our algorithm are given in the following:

Theorem 7. *There is an $\tilde{O}(mn^2 + n^5)$ time algorithm that given an $m \times n$ matrix A, computes a coloring $x \in \{-1, +1\}^n$ satisfying $\mathrm{disc}_2(A, x) = O(\sqrt{\lg n} \cdot \mathrm{herdisc}_2(A))$.*

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*

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

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

Theorem 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*

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

Theorem 5 follows from Theorem 8 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 8. The first step of our proof uses the following linear algebraic fact:

Lemma 1. *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*

$$\det(C^T C) \geq \frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}}.$$

Proof. The k'th symmetric function of $\lambda_1, \ldots, \lambda_n$ is defined as (see e.g. the textbook [18] 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* (the subset of column indices is the same as the subset of row indices) of $A^T A$, then it also holds that (see e.g. the textbook [18] p. 494):

$$s_k = \sum_{B \in \mathcal{S}_k(A^T A)} \det(B).$$

Since $|\mathcal{S}_k(A^TA)| = \binom{n}{k}$ there must be a $B \in \mathcal{S}_k(A^TA)$ 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^TA, it follows that there exists an $m \times k$ submatrix C of A such that $B = C^TC$ and thus $\det(C^TC) \geq \left(\prod_{i=1}^k \lambda_i\right)/\binom{n}{k}$. ☐

With Lemma 1 established, we are ready to present the proof of Theorem 8:

Proof of Theorem 8. 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^TA. From Lemma 1, it follows that for every $k \leq n$, there is an $m \times k$ submatrix C of A such that

$$\det(C^TC) \geq \frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}}.$$

If we also have $k \leq m$, we can let $\mathcal{T}_k(C)$ denote the set of all $k \times k$ submatrices of C and use the Cauchy-Binet formula to conclude that:

$$\det(C^TC) = \sum_{D \in \mathcal{T}_k(C)} \det(D)^2.$$

But $\mathcal{T}_k(C) \subseteq \mathcal{T}_k(A)$ hence there must exist a $k \times k$ matrix $D \in \mathcal{T}_k(A)$ such that

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

$$|\det(D)| \geq \sqrt{\frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}\binom{m}{k}}}.$$

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

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

☐

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 1. *For an $m \times n$ real matrix A, let $M = A^TA$. Then:*

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

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

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

☐

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 6. *For an $m \times n$ real matrix A with $\det(A^T A) \neq 0$, we have*

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

The fact $\text{herdisc}_\infty(A) \geq \text{herdisc}_2(A)$ is true for all A, thus the difficulty in proving Theorem 6 lies in establishing that

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

Our proof uses many of the ideas from the proof of the determinant lower bound (Theorem 1) in [12]. We start by introducing the linear discrepancy in the ℓ_2 setting and summarize known relations between linear discrepancy and hereditary discrepancy.

Definition 1. *Let A be an $m \times n$ real matrix. Then its linear ℓ_2-discrepancy is defined as:*

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

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 [12, 15]). 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 $\text{lindisc}_2(A)$ 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 2 (Lovász et al. [12])**.** *For all $m \times n$ real matrices A, it holds that* $\text{lindisc}_2(A) \leq 2\,\text{herdisc}_2(A)$.

We remark that [12] proved Lemma 2 only for the ℓ_∞-discrepancy, but their proof only uses the fact that $\{x : \|Ax\|_\infty \leq 1\}$ is centrally symmetric and convex (see [12] Lemma 1). The same is true for the U_A defined above.

In light of Lemma 2, 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 [12] (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 [12]):

Lemma 3. *Let A be an $m \times n$ real matrix with $\det(A^T A) \neq 0$. Then*

$$\text{lindisc}_2(A) \geq \sqrt{\frac{n}{2\pi e m}} \det(A^T A)^{1/2n}.$$

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

$$v_n = \pi^{n/2}/\Gamma(n/2+1) \leq \left(\frac{2\pi e}{n}\right)^{n/2},$$

we conclude:

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

□

Combining Lemma 2 and Lemma 3 proves Theorem 6.

Having establishes Theorem 6, we are ready to prove our last result on hereditary ℓ_2-discrepancy:

Restatement of Corollary 2. *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*

$$\mathrm{herdisc}_2(A) \geq \frac{k}{e}\sqrt{\frac{\lambda_k}{8\pi m n}}.$$

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 1, we know that for all $k \leq n$, there is an $m \times k$ submatrix C of A such that

$$\det(C^T C) \geq \frac{\prod_{i=1}^k \lambda_i}{\binom{n}{k}} \geq \left(\frac{k\lambda_k}{en}\right)^k.$$

From Theorem 6, we get that

$$\begin{aligned}
\mathrm{herdisc}_2(C) &\geq \sqrt{\frac{k}{8\pi e m}} \det(C^T C)^{1/2k} \\
&\geq \frac{k}{e}\sqrt{\frac{\lambda_k}{8\pi m n}}.
\end{aligned}$$

Since C is obtained from A by deleting a subset of the columns, it follows that $\mathrm{herdisc}_2(A) \geq \mathrm{herdisc}_2(C)$, 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 7. *There is an $\tilde{O}(mn^2 + n^5)$ time algorithm that given an $m \times n$ matrix A, computes a coloring $x \in \{-1, +1\}^n$ satisfying $\mathrm{disc}_2(A, x) = O(\sqrt{\lg n} \cdot \mathrm{herdisc}_2(A))$.*

Our algorithm is inspired mainly by the Edge-Walk algorithm by Lovett and Meka [13]. 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 $|x_i| < 1 - \delta$ 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, a parameter $\delta > 0$ and a vector $x \in [-1, +1]^n$. We think of the vector x as a partial coloring and let k denote the number of live coordinates. Upon termination, it returns another vector $\gamma \in \mathbb{R}^n$. If we let $\hat{x} = x + \gamma$, then we say that the algorithm *succeeds* if the following holds:

- There are at most $k/2$ live coordinates in $\hat{x}$.
- For all i, we have $|\hat{x}_i| \leq 1$.
- $\|A\hat{x}\|_2^2 - \|Ax\|_2^2 \leq 2^{21} m (\text{herdisc}_2(A))^2$.

Thus if the partial coloring algorithm succeeds, then $\|A\hat{x}\|_2$ doesn't change 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 2. Our algorithm proceeds in iterations, where in each step it samples a random vector v that it adds to a result vector γ. The way we sample v is roughly to find the eigenvectors of $A^T A$ and then sample v orthogonal to the eigenvectors corresponding to the largest eigenvalues. This bounds the growth of $\|A(x + \gamma)\|_2$ when going to $\|A(x + \gamma + v)\|_2$. At the same time, we use the ideas by Lovett and Meka where we include constraints forcing v orthogonal to e_i for every coordinate i that is not live. Finally, if the vectors we sample are "short" enough, the coordinates of $x + \gamma$ will never exceed 1 with good probability. The full details of the algorithm are as follows:

PartialColor(A, x, δ):

1. Let k denote the number of live coordinates in x.
2. Initialize $\gamma \leftarrow \mathbf{0}$, $\varepsilon \leftarrow \delta/\sqrt{c_0 \lg(k/\delta)}$ for a big enough constant c_0, $T \leftarrow 160\varepsilon^{-2}$.
3. Let C denote the $m \times n$ matrix obtained from A by zeroing all columns corresponding to coordinates that are not live.
4. Compute an eigendecomposition of $C^T C$ to obtain the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and corresponding eigenvectors $\mu_1, \ldots, \mu_n$.
5. For $j = 1, \ldots, T$:
   (a) Compute the set S of coordinates i such that $|(x + \gamma)_i| \geq 1 - \delta$. If $|S| \geq n - k/2$, return γ.
   (b) Form an $n \times (|S| + k/4)$ matrix B having $\mu_1, \ldots, \mu_{k/4}$ as columns, as well as e_j for every $j \in S$.
   (c) Compute a QR-factorization $B = QR$ of B. The last $k/4$ columns $q_{n-k/4+1}, \ldots, q_n$ of Q form an orthonormal basis of a $k/4$-dimensional subspace V orthogonal to the span of $\mu_1, \ldots, \mu_{k/4}$ and all e_j with $j \in S$.
   (d) Let $g_{n-k/4+1}, \ldots, g_n$ be independent $\mathcal{N}(0, 1)$ random variables and set $v := (\sum_{i=n-k/4+1}^{n} g_i q_i)$.
   (e) Update $\gamma \leftarrow \gamma + \varepsilon v$.
6. Abort.

If the algorithm terminates in the abort step, we also say that it does not succeed. Our goal is to prove the following about our partial coloring algorithm:

Lemma 4. *Given a parameter $0 < \delta < 1$, an $m \times n$ matrix A and a vector $x \in [-1, +1]^n$, the **PartialColor** algorithm runs in time $O(mn^2 + n^3 \delta^{-2} \lg(n/\delta))$ and succeeds with probability at least $1/8$.*

The bound on the running time is shown as follows: Computing C^TC takes $O(mn^2)$ time. Computing an eigendecomposition of C^TC takes time $O(n^3)$. The running time of an iteration of the for-loop is dominated by the time to compute a QR-factorization, which takes $O(n^3)$ time. Since the loop runs $T = O(\varepsilon^{-2}) = O(\delta^{-2}\lg(n/\delta))$ times, the running time follows.

What remains is to prove the claimed success probability. We split the proof into three steps, arguing that each of the three requirements for succeeding are satisfied with high probability. At the end, we can simply apply a union bound.

No Large Coordinates. The first thing we bound is the probability that there is some coordinate i of $\hat{x}$ with $|\hat{x}_i| > 1$. The main idea in the proof follows previous work quite closely. More concretely, we observe that as soon as a coordinate exceeds $1-\delta$ in absolute value, it will never change again. Thus for a coordinate to exceed 1, it must be the case that adding a single vector εv to γ caused a coordinate to increase by at least δ in absolute value. Since ε is sufficiently smaller than δ and our vectors v are sampled from a fairly high-dimensional subspace, this will happen with tiny probability. We are ready for the formal proof.

We start by defining a few random variables that we will need throughout our proofs. Let $Q_{j,n-k/4+1}, \dots, Q_{j,n}$ be the random variables giving the vectors $q_{n-k/4+1}, \dots, q_n$ computed in step (c) of the above algorithm during the j'th iteration of the for-loop. If the algorithm already terminated in step (a) in the j'th iteration or earlier, we let each $Q_{j,i}$ equal $\mathbf{0}$. Let $G_{j,n-k/4+1}, \dots, G_{j,n}$ be the random variables giving the values $g_{n-k/4+1}, \dots, g_n$ sampled in step (d) during the j'th iteration of the for-loop (taking values 0 if we don't reach the j'th iteration). Finally, let V_j be the random variable giving the vector v computed in the j'th iteration of the for-loop, i.e. $V_j = \sum_{i=n-k/4+1}^{n} G_{j,i}Q_{j,i}$.

Since we are dealing with $\mathcal{N}(0,1)$ distributed random variables, the following tail bound will be useful:

Fact 1. *If X is $\mathcal{N}(0,1)$ distributed, then* $\Pr[|X| \geq \lambda] \leq 2\exp(-\lambda^2/2)$.

We will also need the following facts about normal distributed random variables:

Fact 2. *If $X \sim \mathcal{N}(0,\sigma_x^2)$ and $Y \sim \mathcal{N}(0,\sigma_y^2)$ with X and Y independent, then $aX + bY \sim \mathcal{N}(0, a^2\sigma_x^2 + b^2\sigma_y^2)$ if a and b are constants. We also have $\mathbb{E}[X^2] = \sigma_x^2$.*

We start by proving that no coordinate of γ changes by much when adding εv to γ in step (e) of the algorithm. More formally, we prove:

Lemma 5. *For any parameter $\lambda > 0$, we have:*

$$\Pr\left[\max_j \max_i |\langle V_j, e_i\rangle| > \lambda\right] \leq 2kT\exp(-\lambda^2/2).$$

Proof. Consider any e_i and any V_j. If e_i is not live in the j'th iteration of the for-loop, then $\langle V_j, e_i\rangle = 0$ since we choose V_j in a subspace orthogonal to e_i. Thus the rest of the analyzis assumes i is live during the j'th iteration. Note that there are at most k such indices i. We have:

$$\begin{aligned}\langle V_j, e_i\rangle &= \langle \sum_{h=n-k/4+1}^{n} G_{j,h}Q_{j,h}, e_i\rangle \\ &= \sum_{h=n-k/4+1}^{n} G_{j,h}\langle Q_{j,h}, e_i\rangle.\end{aligned}$$

If the algorithm already terminated in (a) in the j'th iteration or earlier, we have $G_{j,h} = 0$ for all h and thus

$$\Pr\left[|\max_i \langle V_j, e_i\rangle| > \lambda\right] = 0 < 2\exp(-\lambda^2/2).$$

Otherwise, each $G_{j,h}$ is $\mathcal{N}(0,1)$ distributed and we see that:

$$\langle V_j, e_i \rangle \quad \sim \quad \mathcal{N}\left(0, \sum_{h=n-k/4+1}^{n} \langle Q_{j,h}, e_i \rangle^2 \right).$$

But $\sum_{h=n-k/4+1}^{n} \langle Q_{j,h}, e_i \rangle^2$ is the squared ℓ_2-norm of the length of the projection of e_i onto $\mathrm{span}(Q_{j,n-k/4+1}, \dots, Q_n)$. Since this is no more than $\|e_i\|_2^2 = 1$, we have that

$$\Pr[|\langle V_j, e_i \rangle| \geq \lambda] \leq 2 \exp(-\lambda^2/2).$$

By a union bound over all e_i that were live when we started the partial coloring and over all j, we get that

$$\Pr[\max_j \max_i |\langle V_j, e_i \rangle| \geq \lambda] \leq 2kT \exp(-\lambda^2/2).$$

$\square$

Corollary 3. *For any parameter $\lambda > 0$, we have:*

$$\Pr\left[\max_i |\hat{x}_i| > 1 - \delta + \lambda\right] \leq 2kT \exp(-\varepsilon^{-2}\lambda^2/2).$$

Proof. Assume $\max_j \max_i |\langle V_j, e_i \rangle| \leq \lambda$ for some $\lambda > 0$ and consider a coordinate i. Since our algorithm always chooses a v such that v is orthogonal to all e_j with $j \in S$, it follows that $\langle V_j, e_i \rangle$ is non-zero only if $|(x+\gamma)_i| < 1 - \delta$ when reaching iteration j. This implies that, conditioned on $\max_j \max_i |\langle V_j, e_i \rangle| \leq \lambda$, we always have $|(x+\gamma)_i| < 1 - \delta + \varepsilon\lambda$ since we add εv to γ in step (e). Substituting λ by $\varepsilon^{-1}\lambda$ in Lemma 5 completes the proof. $\square$

It follows immediately from Corollary 3 that

$$\begin{aligned}
\Pr[\max_i |\hat{x}_i| > 1] &\leq 2kT \exp(-\varepsilon^{-2}\delta^2/2) \\
&= 2kT \exp(-c_0(\ln(k/\delta)/\delta^2)\delta^2/2) \\
&= 2kTk^{-c_0/2}\delta^{c_0} \\
&= k^{-\Omega(1)}.
\end{aligned}$$

Thus we have shown:

Corollary 4. *The **PartialColor** algorithm satisfies that:*

$$\Pr[\max_i |\hat{x}_i| > 1] = k^{-\Omega(1)}.$$

Rarely Aborts. In this paragraph, we prove that **PartialColor** most often terminates in step (a) rather than aborting in step 6. When it terminates in (a), we have that there are no more than $k/2$ live coordinates in $\hat{x} = x + \gamma$, which is one of the requirements for succeeding. The main idea in our proof is to show that $\mathbb{E}[\| \sum_{j=1}^{T} V_j \|_2^2]$ is very large if the probability of returning in step (a) is small. Combining this with the fact that all coordinates of $\hat{x}$ are bounded by 1 with high probability gives a contradiction and thus we must have that we often terminate in step (a). We start by bounding $\mathbb{E}[\| \sum_{j=1}^{T} V_j \|_2^2]$:

$$\begin{aligned}
\mathbb{E}\left[\| \sum_{j=1}^{T} V_j \|_2^2\right] &= \mathbb{E}\left[\left(\sum_{j=1}^{T} V_j\right)^T \left(\sum_{j=1}^{T} V_j\right)\right] \\
&= \sum_{j=1}^{T} \mathbb{E}\left[\|V_j\|_2^2\right] + \sum_{j=1}^{T} \sum_{h \neq j} \mathbb{E}\left[\langle V_j, V_h \rangle\right] \\
&= \sum_{j=1}^{T} \sum_{i=n-k/4+1}^{n} \mathbb{E}\left[G_{j,i}^2\right] + \sum_{j=1}^{T} \sum_{h \neq j} \mathbb{E}\left[\langle V_j, V_h \rangle\right]
\end{aligned}$$

To analyze $\mathbb{E}\left[G_{j,i}^2\right]$, let E_j denote the event that the algorithm already terminated in step (a) in some iteration $i \leq j$ of the for-loop. Then

$$\begin{aligned}
\mathbb{E}\left[G_{j,i}^2\right] &= \Pr[E_j]\mathbb{E}\left[G_{j,i}^2 \mid E_j\right] + (1-\Pr[E_j])\mathbb{E}\left[G_{j,i}^2 \mid \neg E_j\right] \\
&= \Pr[E_j] \cdot 0 + (1-\Pr[E_j]) \cdot 1 \\
&= (1-\Pr[E_j]).
\end{aligned}$$

We thus have:

$$\begin{aligned}
\mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2\right] &= \sum_{j=1}^{T} k(1-\Pr[E_j])/4 + \sum_{j=1}^{T}\sum_{h\neq j} \mathbb{E}\left[\langle V_j, V_h\rangle\right] \\
&= (k/4)\sum_{j=1}^{T}(1-\Pr[E_j]) + \sum_{j=1}^{T}\sum_{h\neq j}\sum_{i=n-k/4+1}^{n}\sum_{\ell=n-k/4+1}^{n} \mathbb{E}\left[G_{j,i}G_{h,\ell}\langle Q_{j,i}, Q_{h,\ell}\rangle\right]
\end{aligned}$$

Consider $\mathbb{E}\left[G_{j,i}G_{h,\ell}\langle Q_{j,i}, Q_{h,\ell}\rangle\right]$ and assume wlog. that $j > h$. Then for any values $(g, q_1, q_2) \in \mathrm{supp}(G_{h,\ell}, Q_{j,i}, Q_{h,\ell})$, the random variable $G_{j,i}$ is still symmetric around zero when conditioning on $G_{h,\ell} = g \wedge Q_{j,i} = q_1 \wedge Q_{h,\ell} = q_2$. Hence $\mathbb{E}[G_{j,i}G_{h,\ell}\langle Q_{j,i}, Q_{h,\ell}\rangle] = 0$ and we conclude:

$$\mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2\right] = (k/4)\sum_{j=1}^{T}(1-\Pr[E_j]).$$

Now define M_r as the event $\max_i |\hat{x}_i| \in (r, r+1]$ for every integer $r > 0$. Define M_0 as the event $\max_i |\hat{x}_i| \leq 1$. We have:

$$\mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2\right] = \Pr[M_0] \cdot \mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2 \mid M_0\right] + \sum_{r=1}^{\infty}\Pr[M_r] \cdot \mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2 \mid M_r\right].$$

We now use that conditioned on M_r, we have all coordinates of $\sum_{j=1}^{T} V_j$ bounded by $\varepsilon^{-1}(1-\delta+r+1) < \varepsilon^{-1}(r+2)$. To see this, note that we start out with $|x_i| < 1-\delta$ for all i and M_r says that $\max_i |\hat{x}_i| < r+1$. But $\hat{x}_i = x_i + (\sum_j \varepsilon V_j)_i$ and thus all coordinates are bounded as claimed. Moreover, there are at most k non-zero coordinates in $\sum_j V_j$ (since there are only k live coordinates in x when we start out). This implies that $\|\sum_{j=1}^{T} V_j\|_2^2 \leq k\varepsilon^{-2}(r+2)^2$. Using Corollary 3 to bound $\Pr[M_r]$ for $r > 0$, we get that

$$\begin{aligned}
\mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2\right] &\leq \\
\Pr[M_0] \cdot 4\varepsilon^{-2}k + \sum_{r=1}^{\infty} 2kT\exp(-\varepsilon^{-2}(r+1)^2/2)k\varepsilon^{-2}(r+2)^2 &= \\
\Pr[M_0] \cdot 4\varepsilon^{-2}k + \sum_{r=1}^{\infty} 240k^2\exp(-\varepsilon^{-2}(r+1)^2/2)\varepsilon^{-4}(r+2)^2.&
\end{aligned}$$

Since $\varepsilon^{-2} \geq c_0 \ln k$ for a big enough constant c_0, we get

$$\mathbb{E}\left[\|\sum_{j=1}^{T} V_j\|_2^2\right] \leq \Pr[M_0] \cdot 4\varepsilon^{-2}k + 1 \leq 4\varepsilon^{-2}k + 1.$$

We have thus shown that

$$4\varepsilon^{-2}k + 1 \geq (k/4)\sum_{j=1}^{T}(1 - \Pr[E_j]).$$

Since $\Pr[E_j] \geq \Pr[E_h]$ for $j > h$, we get that

$$\begin{aligned} 4\varepsilon^{-2}k + 1 \geq (k/4)T(1 - \Pr[E_T]) \Rightarrow \\ (1 - \Pr[E_T]) \leq \frac{16\varepsilon^{-2}k + 4}{kT} \Rightarrow \\ \Pr[E_T] \geq 1 - \frac{16\varepsilon^{-2}k + 4}{kT} \geq 1 - \frac{20\varepsilon^{-2}k}{kT}. \end{aligned}$$

Since $T = 160\varepsilon^{-2}$, we get that $\Pr[E_T] \geq 7/8$. We have thus shown:

Lemma 6. *The **PartialColor** algorithm satisfies that there are no more than $k/2$ live coordinates in $\hat{x} = x + \gamma$ with probability at least $7/8$.*

Growth in Discrepancy. In this paragraph, we bound the increase in discrepancy when updating x to $\hat{x} = x + \gamma$. For this, we first rewrite $\|A(\gamma + x)\|_2^2 - \|Ax\|_2^2$. We see that:

$$\begin{aligned} \|A(\gamma + x)\|_2^2 - \|Ax\|_2^2 &= \\ \|A\gamma\|_2^2 + \|Ax\|_2^2 + 2\gamma^T A^T Ax - \|Ax\|_2^2 &= \\ \|A\gamma\|_2^2 + 2\gamma^T A^T Ax. \end{aligned}$$

We start by examining $\|A\gamma\|_2^2$ and see that:

$$\mathbb{E}\left[\|A\gamma\|_2^2\right] = \varepsilon^2 \mathbb{E}\left[\|\sum_j AV_j\|_2^2\right] = \varepsilon^2 \sum_j \sum_h \mathbb{E}\left[V_h^T A^T A V_j\right].$$

Consider first a pair $j \neq h$ and assume wlog. $h > j$. Then $V_h^T A^T A V_j = \langle V_h, A^T A V_j\rangle$. Now for every value $v \in \mathrm{supp}(V_j)$, we see that the distribution of $\langle V_h, A^T A V_j\rangle$ is symmetric around 0 conditioned on $V_j = v$. Thus $\mathbb{E}\left[V_h^T A^T A V_j\right] = 0$. Therefore

$$\mathbb{E}\left[\|A\gamma\|_2^2\right] \quad = \quad \varepsilon^2 \sum_j \mathbb{E}\left[V_j^T A^T A V_j\right].$$

By construction, we know that V_j is orthogonal to the eigenvectors $\mu_1, \dots, \mu_{k/4}$ of $C^T C$. Furthermore, V_j is orthogonal to every e_i where i was not live when we started the partial coloring. Thus $V_j^T A^T A V_j = V_j^T C^T C V_j$ and hence $V_j^T A^T A V_j \leq \lambda_{k/4}\|V_j\|_2^2$. We continue:

$$\begin{aligned} \mathbb{E}\left[\|A\gamma\|_2^2\right] &\leq \varepsilon^2 \lambda_{k/4} \sum_j \mathbb{E}\left[\|V_j\|_2^2\right] \\ &= \varepsilon^2 \lambda_{k/4} \sum_j \sum_{i=n-k/4+1}^{n} \mathbb{E}\left[G_{j,i}^2\right] \\ &\leq \varepsilon^2 \lambda_{k/4} T k/4 \\ &= 40k\lambda_{k/4}. \end{aligned}$$

Thus by Markov's inequality, $\Pr[\|A\gamma\|_2^2 > 320k\lambda_{k/4}] \leq 1/8$. Consider now $\mathbb{E}[\gamma^T A^T Ax]$:

$$\mathbb{E}[\gamma^T A^T Ax] \quad = \quad \sum_{j=1}^{T} \mathbb{E}[\langle \varepsilon V_j, A^T Ax\rangle].$$

Since $A^T Ax$ is just a fixed vector, we have that the distribution of $\langle \varepsilon V_j, A^T Ax \rangle$ is symmetric around 0 and hence $\Pr[\gamma^T A^T Ax \leq 0] \geq 1/2$. By Markov's inequality and a union bound, we simultaneously have that $\|A\gamma\|_2^2 \leq 320k\lambda_{k/4}$ and $\gamma^T A^T Ax \leq 0$ with probability at least $1 - 1/8 - 1/2 = 3/8$. Since

$$\|A(\gamma + x)\|_2^2 - \|Ax\|_2^2 = \|A\gamma\|_2^2 + 2\gamma^T A^T Ax$$

we conclude that with probability at least 3/8, we have

$$\|A(\gamma + x)\|_2^2 - \|Ax\|_2^2 \leq 320k\lambda_{k/4}.$$

We would like to use Corollary 2 to relate $k\lambda_{k/4}$ to the hereditary discrepancy of A. Unfortunately, C is $m \times n$ and we would lose something like a $\sqrt{n/k}$ factor, which is quite terrible as k decreases. Our key idea is to relate $\lambda_{k/4}$ to the eigenvalues of an $m \times k$ submatrix of A and then use that the hereditary discrepancy of such a submatrix is less than or equal to the hereditary discrepancy of A.

So consider the $m \times k$ matrix D obtained from C by deleting the columns that were zeroed out (i.e. D is the submatrix obtained from A by deleting all columns corresponding to coordinates that were not live when we started the partial coloring). Let $\lambda'_1 \geq \cdots \geq \lambda'_k \geq 0$ denote the eigenvalues of $D^T D$. We claim $\lambda'_j = \lambda_j$ for all $j = 1, \ldots, k$. To see this, observe that if $\lambda_j > 0$, then we must have that μ_j has all coordinates i corresponding to zeroed columns in C equal to 0 as otherwise we cannot have $\mu_j = \lambda_j C^T C \mu_j$ (if the i'th column of C is zeroed, so is the i'th row of $C^T C$). This means that if we take μ_j and delete all coordinates corresponding to zeroed out columns of C, then the resulting vector μ'_j satisfies $\mu'_j = \lambda_j D^T D \mu'_j$, thus all μ'_j are also distinct eigenvectors of $D^T D$ having the corresponding eigenvalues λ_j. In the opposite direction, if μ is an eigenvector of $D^T D$ with non-zero eigenvalue $\lambda > 0$, then if we pad with zeroes corresponding to the columns deleted when going from C to D, we would also get an eigenvector for $C^T C$ with the same non-zero eigenvalue. Hence $D^T D$ and $C^T C$ have exactly the same non-zero eigenvalues. It therefore holds with probability at least 3/8 that:

$$\|A(\gamma + x)\|_2^2 - \|Ax\|_2^2 \leq 320k\lambda'_{k/4}.$$

Now D is an $m \times k$ matrix (instead of $m \times n$ as C was). Thus it follows from Corollary 2 that

$$\begin{aligned} 320k\lambda'_{k/4} &= 320 \cdot 2^7 e^2 \pi m \cdot \frac{(k/4)^2}{e^2} \frac{\lambda'_{k/4}}{8\pi m k} \\ &\leq 2^{21} m (\mathrm{herdisc}_2(D))^2. \end{aligned}$$

But D was obtained by deleting columns of A and thus $\mathrm{herdisc}_2(D) \leq \mathrm{herdisc}_2(A)$ by definition, and we get that with probability at least 3/8, we have:

$$\|A(\gamma + x)\|_2^2 - \|Ax\|_2^2 \leq 2^{21} m (\mathrm{herdisc}_2(A))^2.$$

Using a union bound together with Corollary 4 and Lemma 6 completes the proof of Lemma 4.

3.2 The Final Algorithm

Now that we have the **PartialColor** algorithm, getting to a lower 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, δ) for some parameter $\delta > 0$ to be fixed shortly. Each call returns a vector γ. If the call to **PartialColor** succeeds, we update $x \leftarrow x + \gamma$ and continue. Otherwise, we discard γ and try again. Since **PartialColor** succeeds with constant probability, we expect to discard γ only a constant number of times. We stop once there are no live coordinates in x, i.e. all coordinates satisfy $1 - \delta \leq |x_i| \leq 1$.

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

$$\begin{aligned} \|Ax\|_2^2 &\leq \lg n (2^{21} m (\mathrm{herdisc}_2(A))^2) \Rightarrow \\ \|Ax\|_2 &= O(\sqrt{m \lg n}\, \mathrm{herdisc}_2(A)). \end{aligned}$$

We now round each coordinate to the nearest of 1 and -1 to obtain a coloring $\hat{x} \in \{-1,+1\}^n$. Let $\lambda_1 \geq \cdots \geq \lambda_n \geq 1$ be the eigenvalues of $A^T A$. Since $\|\hat{x}-x\|_\infty \leq \delta$, we have

$$\|A(\hat{x}-x)\|_2 \leq \sqrt{\lambda_1}\|\hat{x}-x\|_2 \leq \sqrt{\lambda_1}\delta\sqrt{n}.$$

From Corollary 2, we get that $\sqrt{\lambda_1} = O(\mathrm{herdisc}_2(A)\sqrt{mn})$. Thus $\sqrt{\lambda_1}\delta\sqrt{n} = O(n\delta\sqrt{m}\,\mathrm{herdisc}_2(A))$. We now fix $\delta = 1/n$ and get $\sqrt{\lambda_1}\delta\sqrt{n} = O(\sqrt{m}\,\mathrm{herdisc}_2(A))$. We then conclude:

$$\begin{aligned}
\mathrm{disc}_2(A,\hat{x}) &= (1/\sqrt{m})\|A\hat{x}\|_2 \\
&\leq (1/\sqrt{m})(\|Ax\|_2 + \|A(\hat{x}-x)\|_2) \\
&= O(\sqrt{\lg n}\cdot \mathrm{herdisc}_2(A)).
\end{aligned}$$

Since the running time of **PartialColor** is $O(mn^2 + n^3\delta^{-2}\lg(n/\delta))$ and we call it an expected $O(\lg n)$ times with $\delta = 1/n$, the final algorithm has an expected running time of $\tilde{O}(mn^2 + n^5)$. This concludes the proof of Theorem 7.

4 Acknowledgment

The author wishes to thank Nikhil Bansal and Aleksandar Nikolov for useful discussions and pointers to relevant literature.

References

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

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

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

[4] N. Bansal, D. Dadush, and S. 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] N. Bansal, D. Dadush, S. Garg, and S. Lovett. The gram-schmidt walk: A cure for the banaszczyk blues. *CoRR*, abs/1708.01079, 2017.

[6] N. Bansal and S. 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] M. Charikar, A. Newman, and A. Nikolov. Tight hardness results for minimizing discrepancy. In *Proc. 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 1607–1614, 2011.

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

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

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

[12] L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151 – 160, 1986.

[13] S. Lovett and R. Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.

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

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

[16] J. Matoušek and A. 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.

[17] J. Matoušek, A. Nikolov, and K. Talwar. Factorization norms and hereditary discrepancy. *CoRR*, abs/1408.1376, 2014.

[18] C. D. Meyer, editor. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[19] J. Spencer. Six standard deviations suffice. *Trans. Amer. Math. Soc.*, 289:679–706, 1985.