Improved algorithmic bounds for discrepancy of sparse set systems

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Download date: 17. Nov. 2023

Improved Algorithmic Bounds for Discrepancy of Sparse Set Systems

Nikhil Bansal [*]   Shashwat Garg[†]

January 6, 2018

Abstract

We consider the problem of finding a low discrepancy coloring for sparse set systems where each element lies in at most t sets. We give an algorithm that finds a coloring with discrepancy $O((t \log n \log s)^{1/2})$ where s is the maximum cardinality of a set. This improves upon the previous constructive bound of $O(t^{1/2} \log n)$ based on algorithmic variants of the partial coloring method, and for small s (e.g. $s = \mathrm{poly}(t)$) comes close to the non-constructive $O((t \log n)^{1/2})$ bound due to Banaszczyk. Previously, no algorithmic results better than $O(t^{1/2} \log n)$ were known even for $s = O(t^2)$. Our method is quite robust and we give several refinements and extensions. For example, the coloring we obtain satisfies the stronger size-sensitive property that each set S in the set system incurs an $O((t \log n \log |S|)^{1/2})$ discrepancy. Another variant can be used to essentially match Banaszczyk's bound for a wide class of instances even where s is arbitrarily large. Finally, these results also extend directly to the more general Komlós setting.

1 Introduction

Let $(V, \mathcal{S})$ be a finite set system, with $V = \{1, \ldots, n\}$ and $\mathcal{S} = \{S_1, \ldots, S_m\}$ a collection of subsets of V. For a two-coloring $\chi : V \to \{-1, 1\}$, the discrepancy of χ for a set S is defined as $\chi(S) = |\sum_{j \in S} \chi(j)|$ and measures the imbalance from an even-split for S. The discrepancy of the system $(V, \mathcal{S})$ is defined as

$$\mathrm{disc}(\mathcal{S}) = \min_{\chi : V \to \{-1, 1\}} \max_{S \in \mathcal{S}} \chi(S).$$

That is, it is the minimum imbalance for all sets in $\mathcal{S}$, over all possible two-colorings χ.

Discrepancy is a widely studied topic and has applications to many areas in mathematics and computer science. For more background we refer the reader to the books [Cha00, Mat09, CST+14]. In particular, discrepancy is closely related to the problem of rounding fractional solutions of a linear system of equations to integral ones [LSV86, Rot12], and is widely studied in approximation algorithms and optimization.

Until recently, most of the results in discrepancy were based on non-algorithmic approaches and hence were not directly useful for algorithmic applications. However, in the last few years there has been remarkable progress in our understanding of the algorithmic aspects of discrepancy

[Ban10, CNN11, LM12, Rot14, HSS14, ES14, NT15]. In particular, we can now match or even improve upon all known applications of the widely used partial-coloring method [Spe85, Mat09] in discrepancy. This has led to several other new results in approximation algorithms [Rot13, BCKL14, BN15, NTZ13].

Sparse Set Systems: Despite the algorithmic progress, one prominent question that has remained open is to match the known non-constructive bounds on discrepancy for low degree or sparse set systems. These systems are parametrized by t, that denotes the maximum number of sets that contain any element. Beck and Fiala [BF81] proved, using an algorithmic iterated rounding approach, that any such set system has discrepancy at most $2t - 1$. They also conjectured that the discrepancy in this case is $O(t^{1/2})$. Despite much effort, this remains an elusive open question.

The best known result in this direction is due to Banaszczyk [Ban98], who proved a discrepancy bound of $O((t \log n)^{1/2})$. Unlike most results in discrepancy that are based on the partial-coloring method, Banaszczyk's proof is based on a very different and elegant convex geometric argument, and it is not known how to make it algorithmic. Prior to Banaszczyk's result, the best known non-algorithmic bound[1] was $O(t^{1/2} \log n)$ [Sri97], based on the partial-coloring method. This bound was made algorithmic in [Ban10]. The question of matching Banaszczyk's bound algorithmically and several related variants of the problem have received a lot of attention in recent years [Nik13, Buk13, ES14, EL15].

The regime where the above questions are most interesting is perhaps when $t = \log^\beta n$ for some moderate value of β. If particular, if $\beta < 1$, then the $O(t)$ Beck-Fiala bound already beats the $O((t \log n)^{1/2})$ bound of Banaszczyk. On the other hand, if $\beta = 1/\epsilon$ for some small ϵ, then the partial coloring bound already gives $O(t^{1/2} \log n) = O(t^{1/2+\epsilon})$, and becomes asymptotically close to the conjectured $O(t^{1/2})$ bound.

1.1 Our Results

In this paper we give new algorithmic results that improve upon the previously known algorithmic bounds for the Beck-Fiala problem in various ways. Our first result is the following:

Theorem 1. *Given any t-sparse set system on n elements, there is an efficient algorithm that with high probability, finds a coloring where each set $S \in \mathcal{S}$ incurs an $O((t \log n \log |S|)^{1/2})$ discrepancy.*

In particular, this gives an $O((t \log n \log s)^{1/2})$ discrepancy coloring, where s is the size of the largest set in the system.

This bound is never worse than the previous $O(t^{1/2} \log n)$ algorithmic bound, and essentially matches Banaszczyk's bound if all sets have size poly(t). Previously, bounds better than $O(t^{1/2} \log n)$ were known only for a very restricted range of s. In particular, an $O((s \log t)^{1/2})$ (algorithmic) bound follows by a direct application of the Lovász Local Lemma (LLL). This gives an improvement over previous algorithmic bounds for $s \ll \min(t \log^2 n, t^2)$, but is much worse even for moderately larger values of s such as $s = t^{10}$. It is unclear how to avoid the $s^{1/2}$ loss with LLL-based approaches[2], and the problem of improving upon the $O(t^{1/2} \log n)$ bound was open even for the case of $s = t^2$.

As a concrete example to summarize the status of the known bounds: for $t = O(\log^2 n)$ and $s = t^2 = O(\log^4 n)$, the conjectured Beck-Fiala bound is $O(\log n)$, the best known non-constructive bound is $O(\log^{3/2} n)$, all previous algorithms incur $\Omega(\log^2 n)$ discrepancy, and Theorem 1 above achieves $\tilde{O}(\log^{3/2} n)$ where $\tilde{O}(\cdot)$ hides poly$(\log \log n)$ factors.

---

[1] Throughout this paper we will assume that $t \geq \log^2 n$, to avoid taking minimum with the $O(t)$ bound every time.

[2] For a set of size s, a random coloring incurs $\Omega(s^{1/2})$ discrepancy with probability almost 1.

The algorithm is based on a random walk based approach, similar to that in [Ban10], where each step of the walk is guided by a solution to a semidefinite program (SDP). However, as we describe in Section 3.1, the previous partial coloring based approaches seem inherently incapable of improving upon the $O(t^{1/2} \log n)$ bound (in fact this is true even for $s = t$, where LLL based techniques perform much better). Moreover, the previous approaches also cannot give a size-sensitive discrepancy bound for each set S, as in Theorem 1. A key new idea of our algorithm is to add some extra constraints to the SDP which ensure that *every* large enough set makes enough "progress" towards getting colored, and to use a result of Nikolov [Nik13] to argue the feasibility of this SDP. Roughly, these constraints ensure that a set of size $|S|$ incurs an $O(\sqrt{t})$ discrepancy in expectation for only $O(\log |S|)$ phases. We give an overview of the algorithm in Section 3 after defining the relevant background in Section 2.

While Theorem 1 does not give anything new directly if s is arbitrarily large, our method is quite versatile and readily lends to many extensions. These can be used to match or almost match Banaszczyk's bound in many cases even when s is arbitrary. These are discussed in Theorems 2, 5 and their corollaries below.

Theorem 2. *For $i = 1, \ldots, \log n$, let m_i denote the number of sets in $\mathcal{S}$ with size in $(2^{i-1}, 2^i]$. Call these class-i sets. Then, there is an algorithm that finds a coloring with discrepancy $O((ti(\log m_i + \log \log n))^{1/2})$ for each class-i set.*

The bound above refines Theorem 1 (which follows as $m_i \leq n$ for each i) and can lead to substantially improved bounds when the number of large sets is not too high. Recall that in general we can assume that m_i is not too large, as by a standard linear algebraic argument [BF81], one can assume that $n \leq m$ which implies that the average row size is also at most t and that $m_i \leq nt/2^{i-1}$ for every i. The following corollary of Theorem 2 shows that if the system satisfies the stronger bound of $m_i \leq n^{(\log^{O(1)} t)/i}$ instead, then one can essentially match Banaszczyk's bound.

Corollary 3. *If the number of class-i sets is at most $n^{(\log^{O(1)} t)/i}$, then there is an algorithm that finds an $\tilde{O}((t \log n)^{1/2})$ discrepancy coloring, where $\tilde{O}(\cdot)$ hides poly$(\log t \log \log n)$ factors.*

Theorem 1 also extends to the more general setting of the Komlós conjecture (defined in Section 2) where the matrix entries can be arbitrary reals instead of only 0-1. For this setting, Banaszczyk's result [Ban98] implies an $O(\log^{1/2} n)$ non-constructive bound.

Theorem 4. *Given a matrix A with n columns such that the ℓ_2-norm of any column is at most 1, then there is an algorithm to find a coloring where each row j incurs a discrepancy of $O((\log n \log s_j)^{1/2})$, where s_j is the ℓ_1-norm of row j.*

The following extension of Theorem 4 allows further flexibility in choosing different weights w_j for rows to allow improved bounds in many cases.

Theorem 5. *Suppose there is a reweighting of the rows j with non-negative weights w_j such that the ℓ_2-norm of any column is at most β, then the algorithm gets a discrepancy of $O(\beta(\log n \log(w_j s_j))^{1/2}/w_j)$ for each row j.*

The following corollary of Theorem 5 shows that Banaszczyk's bound can be almost matched algorithmically if the sparsity of set system decreases suitably when restricted to very large sets, i.e. no element lies essentially in only very large sets.

Corollary 6. *Given a t-sparse set system with the property that each element lies in $O(t/i^{1+\epsilon})$ sets of size more than $t2^{i-1}$ for each $i \geq 1$, for some fixed $\epsilon > 0$, then there is an algorithm to find an $O((t \log n \log t)^{1/2})$ discrepancy coloring.*

2 Preliminaries

We describe some of the basic concepts that we will need.

2.1 Vector Discrepancy

If A is the $m \times n$ incidence matrix of the set-system $(V, \mathcal{S})$ with rows corresponding to the sets $S_1, \ldots, S_m$ and columns to the elements $[n]$[3], then we can write the discrepancy of A as

$$\operatorname{disc}(A) = \min_{x \in \{-1,1\}^n} \|Ax\|_\infty$$

The above definition of discrepancy generalizes to any matrix A with real entries. Let us index the rows of A by j and the columns by i. If we relax the entries x_i to be unit vectors v_i in $\mathbb{R}^n$ instead of ± 1, we obtain the relaxation

$$\operatorname{vecdisc}(A) = \min_{v_1, \ldots, v_n \in S^{n-1}} \max_{j=1}^m \| \sum_{i=1}^n A_{ji} v_i \|_2.$$

We refer to this as the vector discrepancy of A. The vector discrepancy can be computed efficiently (to arbitrarily high accuracy) by solving the following semidefinite program (SDP) and doing a binary search on λ.

$$\| \sum_i A_{ji} v_i \|_2^2 \;\; \leq \;\; \lambda^2 \qquad \forall j \in [m]$$

$$\|v_i\|_2^2 \;\; = \;\; 1 \qquad \forall i \in [n]$$

Komlós Conjecture: The following conjecture generalizes the Beck-Fiala conjecture.

Conjecture 7 (Komlós). *Any matrix A with columns of ℓ_2-norm at most 1 has discrepancy $O(1)$.*

The Beck-Fiala conjecture follows by scaling the columns of the incidence matrix A by $\sqrt{t}$. While the best known bound for the Komlós setting is $O(\log^{1/2} n)$ [Ban98], Nikolov [Nik13] showed that the Komlós conjecture holds for vector colorings. In particular, he showed the following result:

Theorem 8 ([Nik13]). *For any $m \times n$ matrix A with columns of ℓ_2-norm at most 1, $\operatorname{vecdisc}(A) \leq 1$.*

As pointed out to us by Raghu Meka, an easy proof of $\operatorname{vecdisc}(A) \leq O(1)$ also follows from Banaszczyk's result [Ban98].

2.2 Martingales

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables on some probability space, and let Y_t be a function of $X_1, \ldots, X_t$. The sequence $Y_0, Y_1, Y_2, \ldots, Y_n$ is called a martingale with respect to the sequence $X_1, \ldots, X_n$ if for all $t \in [n]$, $\mathbb{E}[|Y_t|]$ is finite and $\mathbb{E}[Y_t | X_1, X_2, ..., X_{t-1}] = Y_{t-1}$. We will use $\mathbb{E}_{t-1}[Y_t]$ to denote $\mathbb{E}[Y_t | X_1, X_2, ..., X_{t-1}]$. We will need the following martingale concentration inequality.

Theorem 9 (Freedman [Fre75]). *Let $Y_1, \ldots, Y_n$ be a martingale with respect to $X_1, \ldots, X_n$ such that $|Y_t - Y_{t-1}| \leq M$ for all t, and let $W_t = \sum_{j=1}^t \mathbb{E}_{j-1}[(Y_j - Y_{j-1})^2] = \sum_{j=1}^t Var[Y_j | X_1, \ldots, X_{j-1}]$. Then for all $\lambda \geq 0$ and $\sigma^2 \geq 0$, we have*

$$\Pr[|Y_n - Y_0| \geq \lambda \text{ and } W_n \leq \sigma^2] \leq 2 \exp\left(-\frac{\lambda^2}{2(\sigma^2 + M\lambda/3)} \right).$$

---

[3]We use $[n]$ to denote the set $\{1,2,\ldots,n\}$.

3 The Main Result

In this section we describe the main algorithm and prove Theorem 1.

3.1 High-level Overview

The algorithm has a similar structure to previous random walk based approaches. The algorithm starts with the coloring $x_0 = 0^n$ at time 0, and at each time step k, updates the coloring at time $k-1$ by adding a small increment. This increment is determined by solving an appropriate SDP. If a variable reaches -1 or 1 it is frozen, and its value is not updated any more. The variables that are not frozen are called alive.

Limitations of partial coloring based approaches: However, the previous algorithms based on these approaches are only able to give a discrepancy of $O(t^{1/2} \log n)$ for the following reason. Roughly speaking, the execution of the algorithm can be divided into $O(\log n)$ phases. In each phase, about half the variables get frozen, and the updates of the coloring at each time step are chosen so that each set incurs an expected discrepancy of $O(\sqrt{t})$ during a phase. So after $O(\log n)$ phases each set incurs an expected discrepancy of $O((t \log n)^{1/2})$, and to bound the maximum discrepancy, one then takes a union bound over the m sets and loses an additional $\log^{1/2} m \approx \log^{1/2} n$ factor.

If one could show a stronger statement that in each phase, the number of alive variables for *every* set reduces by a constant factor, then each set S_j would get colored in at most $O(\log |S_j|)$ phases, and the expected discrepancy for a set would be $O((t \log |S_j|)^{1/2}$ (instead of the $O((t \log n)^{1/2})$ previously). Theorem 1 would then follow after losing another $\log^{1/2} n$ factor for the union bound over the sets.

However, the property that every set reduces in size is non-trivial to guarantee. For example, even in the non-constructive partial coloring based methods [Mat09], in each partial coloring step (i.e. a phase) one can only guarantee that half the remaining variables (globally) get colored, and there could be several sets that incur $\Omega(\sqrt{t})$ discrepancy even though very few of their elements get colored. In particular, these methods incur an $O(t^{1/2} \log n)$ discrepancy even if all the sets are of size t.

New idea: The main idea of our algorithm is to add additional SDP constraints, given by (2), that we call the energy constraints that ensure that the "energy" of *each* (large enough) set increases significantly in each phase. This can be used to show that for each set $S \in \mathcal{S}$, essentially all of its elements will get colored in $O(\log |S|)$ phases. In contrast, the previous algorithms only ensured a global energy constraint over all the variables to measure progress of the algorithm. Moreover, the feasibility problem for this SDP can be viewed as a vector discrepancy problem for another matrix where the ℓ_2-norm remains $O(1)$. Theorem 8 then guarantees feasibility. The main technical part of the analysis is to show that these new energy constraints are sufficient to guarantee the progress property for each set.

Theorems 1 and 4 follows quite directly from the above approach. The extensions in theorems 2 and 5 also follow the general approach above, but use some additional flexibility in the approach to trade off and balance the different parameters.

3.2 Algorithm

We will index time by k. Let $x_k \in [-1, 1]^n$ denote the coloring at the end of time step k. During the algorithm, variables which get set to at least $(1 - 1/n)$ in absolute value are called frozen and their values are not changed any more. The remaining variables are called alive. We denote by $A(k)$ the set of alive variables at the end of time step k. Initially all variables are alive. Let $\gamma = n^{-2}$, and let $T = (10/\gamma^2) \log n$.

1. Initialize $x_0(i) = 0$ for all $i \in [n]$ and $A(0) = \{1, 2, ..., n\}$.

2. For each time step $k = 1, 2, \ldots, T$ repeat the following:

   (a) Find a feasible solution to the following semidefinite program:

$$\| \sum_{i \in S_j} v_i \|_2^2 \;\; \leq \;\; 2t \qquad \text{for each set } S_j \tag{1}$$

$$\| \sum_{i \in S_j} x_{k-1}(i) v_i \|_2^2 \;\; \leq \;\; 2t \qquad \text{for each set } S_j \tag{2}$$

$$\|v_i\|_2^2 \;\; = \;\; 1 \qquad \forall i \in A(k-1)$$

$$\|v_i\|_2^2 \;\; = \;\; 0 \qquad \forall i \notin A(k-1)$$

   (b) Construct $\gamma_k \in \mathbb{R}^n$ as follows: let $r \in \mathbb{R}^n$ be a random ± 1 vector, obtained by setting each coordinate $r(i)$ independently to -1 or 1 with probability $1/2$.
   For each $i \in [n]$, let $\gamma_k(i) = \gamma \langle r, v_i \rangle$. Update $x_k = x_{k-1} + \gamma_k$.

   (c) Initialize $A(k) = A(k-1)$.
   For each i, if $x_k(i) \geq 1 - 1/n$ or $x_k(i) \leq -1 + 1/n$, update $A(k) = A(k) \setminus \{i\}$.

3. Generate the final coloring as follows. Set $x_T(i) = 1$ if $x_T(i) \geq 1 - 1/n$ and set $x_T(i) = -1$ if $x_T(i) \leq -1 + 1/n$. If $i \in A(T)$, set $x_T(i)$ arbitrarily to ± 1.

3.3 Analysis

We begin with some simple observations.

Lemma 10. *The SDP is feasible at each time step k.*

Proof. Consider the incidence matrix A of the set system. For each row A_j of A corresponding to S_j, we add another row $A'_j = A_j \, \text{diag}(x_{k-1})$, where $\text{diag}(x_{k-1})$ is the $n \times n$ diagonal matrix with the (i, i) entry $x_{k-1}(i)$. In other words, A' is obtained by scaling the i-th column of A by $x_{k-1}(i)$. As $|x_{k-1}(i)| \leq 1$ for each i, the ℓ_2-norm of this augmented matrix $A \cup A'$ is at most $\sqrt{2t}$.

As the SDP in the algorithm above corresponds to finding a vector coloring (restricted to the alive elements) with discrepancy $\sqrt{2t}$ for the $2m \times n$ augmented matrix $A \cup A'$, Theorem 8 implies that it has a feasible solution. □

Lemma 11. *For any vector $v \in \mathbb{R}^n$ and a random ± 1 vector $r \in \mathbb{R}^n$, $\mathbb{E}[\langle r, v \rangle^2] = \|v\|_2^2$ and $|\langle r, v \rangle| \leq \sqrt{n} \|v\|_2$.*

Proof. Writing v in terms of its coordinates $v = (v(1), \ldots, v(n))$,

$$\mathbb{E}[\langle r, v \rangle^2] = \mathbb{E}[(\sum_i r(i) v(i))^2] = \sum_{i,j} \mathbb{E}[r(i) r(j)] v(i) v(j) = \|v\|_2^2$$

6

where the last equality uses that $\mathbb{E}[r(i)r(j)] = 0$ for $i \neq j$ and $\mathbb{E}[r(i)^2] = 1$.

The second part follows by Cauchy-Schwarz inequality, as $|\langle r, v \rangle| \leq \|r\|_2 \|v\|_2 = \sqrt{n}\|v\|_2$. $\qquad \square$

This implies the following.

Observation 12. *As $\|v_i\| \leq 1$, $|\gamma \langle r, v_i \rangle| \leq \gamma\sqrt{n}\|v_i\|_2 \leq 1/n$, the rounding in step 3 affects the discrepancy of any set by at most $n \cdot (1/n) = 1$. So we can ignore this rounding error. Moreover, $|\gamma \langle r, v_i \rangle| \leq 1/n$ also implies that no $x_k(i)$ goes out of the range $[-1, 1]$ during any step of the algorithm.*

Let us divide the execution of the algorithm into $p = 10 \log n$ phases each consisting of $L := 1/\gamma^2$ time steps. The following lemma says that the discrepancy added in a few phases cannot be too large.

Lemma 13. *Fix a set $S \in \mathcal{S}$ and some $\ell \in [p]$. Then, for any $0 \leq \beta \leq 1/(\gamma\sqrt{n}) = n^{3/2}$, the discrepancy of S, $\mathrm{disc}(S)$, after ℓ phases satisfies*

$$\Pr[\mathrm{disc}(S) \geq \beta\sqrt{2t\ell}] \leq 2\exp(-\beta^2/3)$$

Proof. Let $Y_k = \mathrm{disc}_k(S)$ denote the discrepancy of S after time k. Then, $Y_k = Y_{k-1} + \gamma \langle r^k, \sum_{i \in S} v_i^k \rangle$, where r^k denotes the random ± 1 vector at time k and v_i^k is the vector for i in the SDP solution at time k. As $\mathbb{E}[r^k, v] = 0$ for any vector v and the r^k are independent, Y_k forms a martingale with respect to $r^1, r^2 \ldots$.

So we can apply Theorem 9. By the SDP constraint (1) we can bound,

$$M \leq |\gamma \langle r^k, \sum_{i \in S} v_i^k \rangle| \leq \gamma \|r^k\|_2 \| \sum_{i \in S} v_i^k \|_2 \leq \gamma n^{1/2} \cdot \sqrt{2t}$$

Similarly, by Lemma 11 ,

$$\mathbb{E}_{k-1}[(Y_k - Y_{k-1})^2] = \mathbb{E}[\gamma^2 \langle r^k, \sum_{i \in S} v_i^k \rangle^2] = \gamma^2 \| \sum_{i \in S} v_i^k \|_2^2 \leq 2t\gamma^2.$$

This implies that $W_k \leq 2kt\gamma^2$ for all k. Applying Freedman's inequality from Theorem 9 with these bounds on M and W_k, we obtain

$$
\begin{aligned}
\Pr[\mathrm{disc}_{\ell L}(S) \geq \beta\sqrt{2t\ell}] \quad &\leq \quad 2\exp\left(-\frac{\beta^2 2t\ell}{2(W_{\ell L} + M\beta\sqrt{2t\ell}/3)}\right) \\
&\leq \quad 2\exp\left(-\frac{2t\ell\beta^2}{2(2\ell Lt\gamma^2 + \gamma\beta\sqrt{2nt}\sqrt{2t\ell}/3)}\right) \\
&\leq \quad 2\exp\left(-\frac{\beta^2}{(2 + \gamma\beta\sqrt{n})}\right) \leq 2\exp(-\beta^2/3).
\end{aligned}
$$

The last step uses that $\beta\gamma\sqrt{n} \leq 1$. $\qquad \square$

For a set S, let $A_S(k)$ denote the number of alive elements in set S at the end of time step k.

Definition 14 (Energy Deficit). *For a set $S \in \mathcal{S}$ and given a coloring x_k at the end of time k, the energy deficit of S is defined as $D_S(k) = \sum_{i \in S}(1 - x_k^2(i))$.*

Initially at time $k = 0$, the energy deficit of any set $S \in \mathcal{S}$ is $|S|$, and is always non-negative. We also note the following.

Lemma 15. *For any set $S \in \mathcal{S}$ and time k, it holds that the number of alive elements $A_S(k) \geq D_S(k) - 2$.*

Proof. Any frozen element can contribute at most $1 - (1 - 1/n)^2 \leq 2/n$ to the energy deficit. Moreover, since each alive alive element can contribute at most 1, $D_S(k) \leq A_S(k) + n \cdot (2/n)$ and the result follows. $\square$

We now come to the key part of the proof, which shows that for each large enough set, its energy deficit decreases substantially in each phase, with sufficiently high probability.

Theorem 16. *Fix a set $S \in \mathcal{S}$ and a time k. For any $\lambda \geq 0$, with probability at least $1 - 4\exp(-\lambda^2/100t)$, it holds that*

$$D_S(k + L) \leq D_S(k) - A_S(k + L) + \lambda. \tag{3}$$

As shown in the corollary below, this theorem implies that a set S shrinks to size $O(\lambda)$ in $O(\log |S|)$ phases with high probability. While these results are stated for general λ (which is useful for the extensions later), the reader may think of $\lambda = O((t \log n)^{1/2})$.

Corollary 17. *Fix a set $S \in \mathcal{S}$ and $\lambda \geq 1$. Let $\ell = 1 + 3\log(|S|/\lambda)$. Then,*

$$\Pr\left[A_S(\ell L) \geq 2\lambda + 2\right] \leq 16 \log n \exp\left(-\frac{\lambda^2}{100t}\right).$$

That is, the probability that S has more than $2\lambda + 2$ alive elements after ℓ phases is at most $16 \log n \exp(-\lambda^2/100t)$.

Proof. Let us apply Theorem 16 at $k = 0, L, \ldots, (\ell - 1)L$. Then, with probability at least $1 - 4\ell \exp(-\lambda^2/100t) \geq 1 - 16 \log n \exp(-\lambda^2/100t)$, we can assume that (3) holds at each of these steps. Let us condition on this event. If $A_S(\ell L) \geq 2\lambda + 2$, then for each of the time steps $k = 0, L, \ldots, (\ell - 2)L$,

$$
\begin{aligned}
D_S(k + L) \quad &\leq \quad D_S(k) - A_S(k + L) + \lambda \\
&\leq \quad D_S(k) - A_S(k + L)/2 - 1 \quad \text{(as } \lambda \leq A_S(\ell L)/2 - 1 \leq A_S(k + L)/2 - 1\text{)} \\
&\leq \quad D_S(k) - D_S(k + L)/2 \quad \text{(By Lemma 15)}
\end{aligned}
$$

This implies that $D_S((\ell - 1)L) \leq D_S(0)(3/2)^{-\ell+1} \leq \lambda$. However, applying (3) again at $k = (\ell - 1)L$ gives that

$$D_S(\ell L) \leq D_S((\ell - 1)L - A_S(\ell L) + \lambda \leq \lambda - (2\lambda + 2) + \lambda < 0$$

which contradicts that the energy deficit is always non-negative. $\square$

Before we prove theorem 16, let us first see how this implies Theorem 1.

Proof (Theorem 1). Applying corollary 17 with $\lambda = 20(t \log n)^{1/2}$, it follows that a set S_j reduces to size at most $O((t \log n)^{1/2})$ in at most $1 + 3(\log |S_j|)$ phases with probability at least $1 - 16 \log n \exp(-\lambda^2/100t) \geq 1 - 1/n^3$. By a union bound, this holds for all the (at most nt) sets with probability at least $1 - 1/n$.

By Lemma 13 with $\beta = 4 \log^{1/2} n$ and $\ell_j = 1 + 3 \log |S_j|$, the discrepancy S_j during the first ℓ_j phases is at most $O(\beta(t\ell_j)^{1/2}) = O((t \log n \log |S_j|)^{1/2})$ with probability $1 - O(n^{-5})$. When S_j has fewer than $\lambda = O((t \log n)^{1/2})$ alive elements, it is *safe* and can incur at most λ additional

8

discrepancy during the rest of the algorithm. Taking a union bound over all the sets gives the result. □

We now prove Theorem 16.

Proof (Theorem 16.) Let v_i^h denote the vector v_i in the SDP solution at time step h, and let r^h denote the random vector at time h. Let $\Delta_S(h) := D_S(h) - D_S(h-1)$ denote the change in energy deficit of S at time h. Then,

$$
\begin{aligned}
\Delta_S(h) &= \sum_{i \in S}(x_{h-1}(i)^2 - x_h(i)^2) = \sum_{i \in S}(x_{h-1}(i)^2 - (x_{h-1}(i) + \gamma\langle r^h, v_i^h\rangle)^2) \\
&= -2\gamma\langle r^h, \sum_{i \in S} x_{h-1}(i)v_i^h\rangle - \sum_{i \in S}\gamma^2\langle r^h, v_i^h\rangle^2
\end{aligned}
\tag{4}
$$

Let us denote the first and second terms on the right hand side of (4) by

$$
T_h := -2\gamma\langle r^h, \sum_{i \in S} x_{h-1}(i)v_i^h\rangle \quad \text{and} \quad R_h := -\sum_{i \in S}\gamma^2\langle r^h, v_i^h\rangle^2
$$

We need to show that with probability at least $1 - 4\exp(-\lambda^2/100t)$,

$$
D_S(k+L) - D_S(k) = \sum_{h=k+1}^{k+L}\Delta_S(h) = \sum_{h=k+1}^{k+L}(T_h + R_h) \leq -A_S(k+L) + \lambda.
$$

To do this, we will show that T_h is a martingale with small variance (this is where the SDP constraints (2) will be used), and that $\sum_{h=k+1}^{k+L} R_h$ is almost deterministic (its variance can be made arbitrarily small by making γ small) with $\mathbb{E}[\sum_{h=k+1}^{k+L} R_h] \leq -A_S(k+L)$.

Let $Y_h = \sum_{h'=1}^h T_{k+h'}$ and $X_h = r^{k+h}$. Then $Y_0 = 0$ and $Y_1, \dots, Y_L$ form a martingale with respect to $X_1, \dots, X_L$. This is the same martingale as in the proof of Lemma 13, except that it is multiplied by a factor of -2. So applying Freedman's inequality as in the proof of Lemma 13 with $\ell = 1$ gives

$$
\Pr[|Y_L| \geq 2\beta\sqrt{2t}] \leq 2\exp(-\beta^2/3).
$$

Choosing $\beta = \lambda/(4\sqrt{2t})$ gives

$$
\Pr[|Y_L| \geq \lambda/2] \leq 2\exp(-\lambda^2/96t).
\tag{5}
$$

We now bound $\sum_h R_h$. Let $Z_h' = \sum_{h'=1}^h R_{k+h'}$, and define the martingale

$$
Z_h = Z_h' - \sum_{h'=1}^h \mathbb{E}_{h'-1}[Z_{h'}' - Z_{h'-1}'] = Z_h' - \sum_{h'=1}^h \mathbb{E}_{h'-1}[R_{k+h'}]
\tag{6}
$$

with respect to $X_1, \dots, X_h$ (this is the standard Doob decomposition of Z_h').

We now apply Freedman's inequality to Z_h. By (6)

$$
|Z_h - Z_{h-1}| = |R_{k+h} - \mathbb{E}_{h-1}[R_{k+h}]| \leq 2|R_{k+h}| \leq 2\gamma^2 n|S| \leq 2\gamma^2 n^2.
$$

and thus $M \leq 2\gamma^2 n^2$. Moreover using the trivial bound $\mathbb{E}_{h-1}[(Z_h - Z_{h-1})^2] \leq M^2 = 4\gamma^4 n^4$, we obtain that total variance $W_L \leq LM^2 = 4\gamma^2 n^4 = 4$. By Freedman's inequality with $\lambda \leq n$, we get

$$
\Pr[Z_L > \lambda/2] \leq 2\exp(-\lambda^2/8(W_L + M\lambda/6)) \leq 2\exp(-\lambda^2/40) \leq 2\exp(-\lambda^2/40t).
\tag{7}
$$

9

By Lemma 11 for each $h \in [k+1, k+L]$,

$$\mathbb{E}[R_h] = -\gamma^2 \sum_{i \in S} \|v_i^h\|^2 = -\gamma^2 A_S(h-1) \leq -\gamma^2 A_S(k+L).$$

So we can bound $\sum_{h'=1}^{h} R_{k+h'}$ as

$$
\begin{aligned}
\sum_{h'=1}^{h} R_{k+h'} &= Z_L' = Z_L + \sum_{h'=1}^{L} \mathbb{E}_{h'-1}[R_{k+h'}] \\
&\leq Z_L - L\gamma^2 A_S(k+L) = Z_L - A_S(k+L).
\end{aligned}
\tag{8}
$$

This gives that

$$
\begin{aligned}
\Pr[\sum_{h=k+1}^{k+L} (T_h + R_h) \geq -A_S(k+L) + \lambda] &= \Pr[Y_L + Z_L' \geq -A_S(k+L) + \lambda] \\
&\leq \Pr[Y_L + Z_L \geq \lambda] \qquad \text{(by (8))} \\
&\leq \Pr[|Y_L| \geq \lambda/2] + \Pr[Z_L \geq \lambda/2] \\
&\leq 4\exp(-\lambda^2/100t) \qquad \text{(by (5) and (7))}
\end{aligned}
$$

$\square$

4   Extensions

We now describe the various extensions. The first extension is to the Komlós setting.

Theorem 4 (restated): *Given a matrix A with n columns such that the ℓ_2-norm of any column is at most 1, then there is an algorithm to find a coloring where each row j incurs a discrepancy of $O((\log n \log s_j)^{1/2})$, where s_j is the ℓ_1-norm of row j.*

The proof follows along the same lines as before, and only needs a modification in the definition of the energy deficit. We describe the relevant differences.

Proof. (Sketch) For the Komlós setting, we replace the SDP constraints (1) by $\|\sum_i a_{ji} v_i\|^2 \leq 2t$ and the constraints (2) by $\|\sum_i |a_{ji}| x_{k-1}(i) v_i\|^2 \leq 2t$ (note the absolute value in the second set of constraints) where a_{ji} denotes the $(j,i)^{th}$ element of A. To use the latter constraint, we define the energy deficit for row j at time k as $D_j(k) = s_j - \sum_i |a_{ji}| x_k(i)^2$. The energy deficit is always non-negative and is s_j at time $k=0$. We also define $A_j(k)$ now as the ℓ_1-norm of row j restricted to the alive variables. This gives that,

$$D_j(k) - D_j(k-1) = -2\gamma \sum_i |a_{ji}| \langle r^k, x_{k-1}(i) v_i^k \rangle - \gamma^2 \sum_i |a_{ji}| \langle r^k, v_i^k \rangle^2.$$

Exactly as in the proof of Theorem 16, the energy constraints of the SDP ensure that the first term behaves as a martingale with deviation in each step $O(\gamma t^{1/2})$. Similarly the second term causes the deficit to decrease in expectation by at least $\gamma^2 \sum_{i \in A_j(k)} |a_{ji}| \langle r^k, v_i^k \rangle^2$, which exactly as before, gives us the required decrease when the ℓ_1-norm of the alive variables in higher than our threshold $O(\lambda)$. $\square$

The next extension follows by using the flexibility in choosing the λ's differently for different sets.

Theorem 2 (restated): *For $i = 1, \ldots, \log n$, let m_i denote the number of sets in $\mathcal{S}$ with size in $(2^{i-1}, 2^i]$. Call these class-i sets. Then there is an algorithm that with probability at least $1 - 1/\log n$ finds a coloring with discrepancy $O((ti(\log m_i + \log \log n))^{1/2})$ for each class-i set.*

Proof. For class-i sets S, we set $\lambda_i = 20(t(\log m_i + \log \log n))^{1/2}$. Then by Corollary 17, the probability that a class-i set does not become smaller than $2\lambda_i$ in $3 \log 2^i = 3i$ phases is at most $16/(m_i^4 \log^3 n)$. By a union bound over all the sets in class i, and over all the $\log n$ choices of i, each set decreases to a size at most $2\lambda_i$ after $O(i)$ phases with probability at least $1 - 1/(2 \log n)$.

Using lemma 13 with $\beta = 3(\log m_i + \log \log n)^{1/2}$, probability that a class-i set S gets discrepancy more than $3(6 \cdot ti(\log m_i + \log \log n))^{1/2}$ in the first $3i$ phases is at most $2/(m_i^3 \log^3 n)$. Taking a union bound over all the m_i sets in class i and over all the $\log n$ choices of i, we get that each set gets discrepancy $O((ti(\log m_i + \log \log n))^{1/2})$ in the first $3i$ phases with probability at least $1 - 1/(2 \log n)$.

Together, this implies that with probability at least $1 - 1/\log n$, the discrepancy for every class-i set is $O((ti(\log m_i + \log \log n))^{1/2})$. $\qquad\square$

Corollary 3 follows directly by plugging the bounds on m_i above.

For theorem 5, we use another flexibility that we can reweight the rows of incidence matrix of the set system. In particular, if we scale the row j by w_j and still require the (weighted) vector discrepancy to be $\sqrt{t}$, then this implies a $\sqrt{t}/w_j$ discrepancy on the original row. As long as this reweighting does not affect the ℓ_2-norms by too much, the previous bounds and analysis still goes through, while allowing us to get tighter bounds on selected rows. In particular we have the following.

Theorem 5 (restated): *Suppose there is a reweighting of the rows j with non-negative weights w_j such that the ℓ_2-norm of any column is at most β, then the algorithm gets a discrepancy of $O(\beta(\log n \log(w_j s_j))^{1/2}/w_j)$ for each row j.*

Proof. (Sketch) The proof follows directly from Theorem 4. We work with the incidence matrix of reweighted system, and then translate the bounds back to the original system. $\qquad\square$

Corollary 6 follows from Theorem 5 as follows.

Proof. (Corollary 6) Call a set a class-i set for $i = 1, \ldots, \log n$, if its size lies in the range $[t2^{i-1}, t2^i]$. We choose $w_i = \sqrt{i + \log t}$ for class i sets, and $w_i = 1$ for sets of size less than t. Fix some element v, and let $m_{\geq i}$ denote the number of sets in classes i or more that contain v. Then, the squared ℓ_2-norm of columns in the weighted incidence matrix is at most

$$t + \sum_{i=1}^{\log n} w_i^2(m_{\geq i} - m_{\geq i+1}) = t + \sum_{i=1}^{\log n} (w_i^2 - w_{i-1}^2)(m_{\geq i}) \leq t + \sum_{i=1}^{\log n} m_{\geq i} = O(t)$$

where we use that $m_{\geq i} \leq O(t/i^{1+\epsilon})$ for any element v.

Applying theorem 5, we obtain that weighted discrepancy for a class-i set is $O((t(i + \log t) \log n)^{1/2})$ and thus the actual discrepancy is $O((t(i + \log t) \log n)^{1/2}/w_i) = O((t \log n)^{1/2})$. For sets of size at most t the discrepancy is $O((t \log n \log t)^{1/2})$. $\qquad\square$

We remark that there are lots of additional flexibility in the approach that we did not use in the extensions above. These could be helpful in some other special instances, and hopefully perhaps even for the general Beck-Fiala problem. For example, we can choose the weights w_i adaptively over time for sets that become dangerous (i.e. incur much more discrepancy than expected), as was done in [Ban10] to obtain the bounds for Spencer's result [Spe85], provided not too many sets become dangerous. Another observation is that the concentration of energy deficit is so strong for the sets of size $\Omega(t \log n)^{1/2}$, that we can essentially assume that the number of alive variables in a set decreases deterministically by a constant factor in each phase. Once the residual size of a set falls below $O((t \log n)^{1/2})$, we can simply discard it from the set system (as is done in iterated rounding, as it can only incur $O((t \log n)^{1/2})$ discrepancy henceforth). So what really matters at any time in the algorithm is the bound on the ℓ_2-norm of any column of this reduced system, rather than the ℓ_2-norm of columns in the original system. This reduction in the ℓ_2-norm can potentially be used to tradeoff the high discrepancy incurred by some sets.

Acknowledgements

References

[Ban98] Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Foundations of Computer Science (FOCS)*, pages 3–10, 2010.

[BCKL14] Nikhil Bansal, Moses Charikar, Ravishankar Krishnaswamy, and Shi Li. Better algorithms and hardness for broadcast scheduling via a discrepancy approach. In *SODA*, pages 55–71, 2014.

[BF81] József Beck and Tibor Fiala. Integer-making theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[BN15] Nikhil Bansal and Viswanath Nagarajan. Approximation-friendly discrepancy rounding. *CoRR*, abs/1512.02254, 2015.

[Buk13] Boris Bukh. An improvement of the beck-fiala theorem. *CoRR*, abs/1306.6081, 2013.

[Cha00] Bernard Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, 2000.

[CNN11] Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In *SODA*, pages 1607–1614, 2011.

[CST+14] William Chen, Anand Srivastav, Giancarlo Travaglini, et al. *A Panorama of Discrepancy Theory*, volume 2107. Springer, 2014.

[EL15] Esther Ezra and Shachar Lovett. On the beck-fiala conjecture for random set systems. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:190, 2015.

[ES14]   Ronen Eldan and Mohit Singh. Efficient algorithms for discrepancy minimization in convex sets. *CoRR*, abs/1409.2913, 2014.

[Fre75]   David A. Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

[HSS14]   Nicholas J. A. Harvey, Roy Schwartz, and Mohit Singh. Discrepancy without partial colorings. In *APPROX/RANDOM 2014*, pages 258–273, 2014.

[LM12]   Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. In *Foundations of Computer Science (FOCS)*, pages 61–67, 2012.

[LSV86]   László Lovász, Joel Spencer, and Katalin Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, 1986.

[Mat09]   Jiri Matousek. *Geometric discrepancy: An illustrated guide*. Springer Science, 2009.

[Nik13]   Aleksandar Nikolov. The komlós conjecture holds for vector colorings. *arXiv preprint arXiv:1301.4039*, 2013.

[NT15]   Aleksandar Nikolov and Kunal Talwar. Approximating hereditary discrepancy via small width ellipsoids. In *Symposium on Discrete Algorithms, SODA*, pages 324–336, 2015.

[NTZ13]   Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Symposium on Theory of Computing , STOC*, pages 351–360, 2013.

[Rot12]   Thomas Rothvoss. The entropy rounding method in approximation algorithms. In *Symposium on Discrete Algorithms (SODA)*, pages 356–372, 2012.

[Rot13]   Thomas Rothvoss. Approximating bin packing within o(log OPT * log log OPT) bins. In *FOCS*, pages 20–29, 2013.

[Rot14]   Thomas Rothvoss. Constructive discrepancy minimization for convex sets. In *Foundations of Computer Science (FOCS)*, pages 140–145, 2014.

[Spe85]   Joel Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289(2):679–706, 1985.

[Sri97]   Aravind Srinivasan. Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems. In *Symposium on Discrete Algorithms (SODA)*, pages 692–701, 1997.