2012 IEEE 53rd Annual Symposium on Foundations of Computer Science

Constructive Discrepancy Minimization by Walking on The Edges

Shachar Lovett
Institute for Advanced Study
slovett@math.ias.edu

Raghu Meka
Institute for Advanced Study
raghu@math.ias.edu

***Abstract*—Minimizing the discrepancy of a set system is a fundamental problem in combinatorics. One of the cornerstones in this area is the celebrated six standard deviations result of Spencer (AMS 1985): In any system of n sets in a universe of size n, there always exists a coloring which achieves discrepancy $6\sqrt{n}$. The original proof of Spencer was existential in nature, and did not give an efficient algorithm to find such a coloring. Recently, a breakthrough work of Bansal (FOCS 2010) gave an efficient algorithm which finds such a coloring. His algorithm was based on an SDP relaxation of the discrepancy problem and a clever rounding procedure. In this work we give a new randomized algorithm to find a coloring as in Spencer's result based on a restricted random walk we call *Edge-Walk*. Our algorithm and its analysis use only basic linear algebra and is "truly" constructive in that it does not appeal to the existential arguments, giving a new proof of Spencer's theorem and the *partial coloring lemma*.**

I. Introduction

Minimizing the discrepancy of a set system is a fundamental problem in combinatorics with many applications in computer science (see [1], [2]). Here, we are given a collection of sets $\mathcal{S}$ from a universe $V = \{1, \ldots, n\}$ and the goal is to find a *coloring* $\chi : V \to \{1, -1\}$ that minimizes the maximum discrepancy $\chi(\mathcal{S}) = \max_{S \in \mathcal{S}} |\sum_{i \in S} \chi(i)|$. We denote the minimum discrepancy of $\mathcal{S}$ by $\mathsf{disc}(\mathcal{S})$.

There is by now a rich body of literature on discrepancy minimization with special focus on the 'discrete' formulation described above. One of the cornerstones in this area is the celebrated six standard deviations result of Spencer [3].

Theorem 1. *For any set system $(V, \mathcal{S})$ with $|V| = n$, $|\mathcal{S}| = m$, there exists a coloring $\chi : V \to \{1, -1\}$ such that $\chi(\mathcal{S}) < K\sqrt{n \cdot \log_2(m/n)}$, where K is a universal constant (K can be 6 if $m = n$).*

The above bound is in fact the best possible upto constant factors (cf. [1]). One remarkable aspect of the above theorem is that for $m = O(n)$, the discrepancy is just $O(\sqrt{n})$, whereas a random coloring has discrepancy $O(\sqrt{n \log n})$. Spencer's original proof relied on an ingenious pigeon-hole principle argument based on Beck's partial coloring approach [4]. However, due to the use of the pigeon-hole principle, the proof was non-constructive: Spencer's proof does not give an efficient (short of enumerating all possible colorings) way to find a good coloring χ as in the theorem. This was a long-standing open problem in discrepancy minimization and it was even conjectured that such an algorithm cannot exist [5]. In a recent breakthrough work, Bansal [6] disproved this conjecture and gave the first randomized polynomial time algorithm to find a coloring with discrepancy $O(\sqrt{n} \cdot \log(m/n))$, thus matching Spencer's bound up to constant factors for the important case of $m = O(n)$. Note that as shown recently by Charikar et al. [7], Bansal's result is algorithmically tight as it is NP-hard to distinguish between set systems with discrepancy zero and those with discrepancy $\Omega(\sqrt{n})$.

In this work we give a new elementary constructive proof of Spencer's result. Our algorithm and its analysis use only basic linear algebra and perhaps more importantly is "truly" constructive. Bansal's algorithm while giving a constructive solution, still implicitly uses Spencer's original non-constructive proof to argue the correctness of the algorithm. Our algorithm on the other hand also gives a new (constructive) proof of Spencer's original result.

Theorem 2. *For any set system $(V, \mathcal{S})$ with $|V| = n$, $|\mathcal{S}| = m$, there exists a randomized algorithm running in time $\tilde{O}((n+m)^3)$* [1] *that with probability at least $1/2$, computes a coloring $\chi : V \to \{1, -1\}$ such that $\chi(\mathcal{S}) < K\sqrt{n \cdot \log_2(m/n)}$, where K is a universal constant.*

The constant K above can be taken as 13 for the case of $m = n$. Observe that our bound matches Spencer's result for all ranges of m, n, whereas Bansal's result loses an additional factor of $\Omega(\sqrt{\log(m/n)})$.

The above discrepancy bound is in fact tight up to

[1] Throughout, $\tilde{O}()$ hides polylogarithmic factors.

0272-5428/12 $26.00 © 2012 IEEE
DOI 10.1109/FOCS.2012.23

constant factors: there exist systems with

We also get a similar constructive proof of Srinivasan's result [8] for minimizing discrepancy in the "Beck-Fiala Setting" where each variable is constrained to occur in a bounded number of sets. Bansal was able to use his SDP based approach to give a constructive proof of Srinivasan's result. Our techniques for Theorem 2 also extend to this setting matching the best known constructive bounds.

Theorem 3. *Let $(V, \mathcal{S})$ be a set-system with $|V| = n$, $|\mathcal{S}| = m$ and each element of V contained in at most t sets from $\mathcal{S}$. Then, there exists a randomized algorithm running in time $\tilde{O}((n+m)^5)$ that with probability at least $1/2$ computes a coloring $\chi : V \to \{1, -1\}$ such that $\chi(\mathcal{S}) < K\sqrt{t} \cdot \log n$, where K is a universal constant.*

We remark that non-constructively, a better bound of $O(\sqrt{t \cdot \log n})$ was obtained by Banaszczsyk [9] using techniques from convex geometry. Beck and Fiala [10] proved that $\mathsf{disc}(\mathcal{S}) < 2t$ and conjectured that $\mathsf{disc}(\mathcal{S}) = O(\sqrt{t})$ (which if true would be tight, cf. [1]) and this remains a major open problem in discrepancy minimization.

II. Outline of Algorithm

To describe the algorithm we first set up some notation. Fix a set system $(V, \mathcal{S})$ with $V = \{1, \ldots, n\}$ and $|\mathcal{S}| = m$. As is usually done, we shall assume that $m \geq n$ – the general case can be easily reduced to this situation. Similar to Spencer's original proof our algorithm also works by first finding a "partial coloring": $\chi : V \to [-1, 1]$ such that

- For all $S \in \mathcal{S}$, $|\chi(S)| = O(\sqrt{n \log(m/n)})$.
- $|\{i : |\chi(i)| = 1\}| \geq cn$, for a fixed constant $c > 0$.

Given such a partial coloring, we can then recurse (as in Spencer's original proof) by running the algorithm on the set of variables assigned values in $(-1, 1)$ without changing the colors of variables assigned values in $\{1, -1\}$. Eventually, we will converge to a full coloring and the total discrepancy (a geometrically decreasing series with ratio roughly $\sqrt{1-c}$) can be bounded by $O(\sqrt{n \log(m/n)})$. Henceforth, we will focus on obtaining such a partial coloring.

Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be the indicator vectors of the sets in $\mathcal{S}$. Then, the discrepancy of χ on $\mathcal{S}$ is $\chi(\mathcal{S}) = \max_{i \in [m]} |\langle \chi, v_i \rangle|$. Our partial coloring algorithm (as does Spencer's approach) works in the more general context of arbitrary vectors, and we will work in this general context.

Theorem 4 (Main Partial Coloring Lemma)**.** *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors, and $x_0 \in [-1, 1]^n$ be a "starting" point. Let $c_1, \ldots, c_m \geq 0$ be thresholds such that $\sum_{j=1}^m \exp(-c_j^2/16) \leq n/16$. Let $\delta > 0$ be a small approximation parameter. Then there exists an efficient randomized algorithm which with probability at least 0.1 finds a point $x \in [-1, 1]^n$ such that*

(i) $|\langle x - x_0, v_j \rangle| \leq c_j \|v_j\|_2$.
(ii) $|x_i| \geq 1 - \delta$ *for at least* $n/2$ *indices* $i \in [n]$.

Moreover, the algorithm runs in time $O((m+n)^3 \cdot \delta^{-2} \cdot \log(nm/\delta))$.

Note that the probability of success 0.1 can be boosted by simply running the algorithm multiple times. Given the above result, we can get the desired partial coloring needed for minimizing set discrepancy by applying the theorem to the indicator vectors of the sets $S \in \mathcal{S}$ with $\delta = 1/n$, and $x_0 = 0^n$. Combining the above with the recursive analysis gives Theorem 2 with a running time of $\tilde{O}((n+m)^5)$. It was pointed to us by Spencer that we can in fact take $\delta = 1/\log n$ and then use randomized rounding to get the running time stated in Theorem 2.

We stress that Spencer's original approach shows the existence of a true partial coloring (the colors take values in $\{-1, 0, 1\}$), whereas our approach gives a fractional coloring—the colors take values in $[-1, 1]$ though many of the colors are close to $\{-1, 1\}$.

The constructive proof of Srinivasan's result, Theorem 3, follows a similar outline starting from our partial coloring lemma. We defer the details to Section VI.

We now describe the proof of the partial coloring lemma.

A. Partial Coloring by Walking on The Edge

We will find the desired vector x by performing a constrained random walk that we refer to as *Edge-Walk* for reasons that will become clear later.

We first describe the algorithm conceptually, ignoring the approximation parameter δ. We will assume throughout that $\|v_1\|_2 = \ldots = \|v_m\|_2 = 1$ as this normalization does not change the problem. Consider the following polytope $\mathcal{P}$ which describes the legal values for $x \in \mathbb{R}^n$,

$$\mathcal{P} := \{x \in \mathbb{R}^n : |x_i| \leq 1 \ \forall i \in [n], \\ |\langle x - x_0, v_j \rangle| \leq c_j \ \forall j \in [m]\}.$$

We will refer to the constraints $|x_i| \leq 1$ as *variable constraints* and to the constraints $|\langle x - x_0, v_j \rangle| \leq c_j$ as *discrepancy constraints*. The partial coloring lemma can be rephrased in terms of the polytope $\mathcal{P}$ as follows: there exists a point $x \in \mathcal{P}$ that satisfies at least $n/2$ variable constraints without any slack. Intuitively, this corresponds to finding a point x in $\mathcal{P}$ that is as far away from origin as possible; the hope being that if $\|x\|_2$ is large, then in fact many of the coordinates of x will be close to 1 in absolute value. We find such a point (and show it's existence) by simulating a constrained Brownian motion in $\mathcal{P}$. (If uncomfortable with Brownian motion, the reader can view the walk as taking very small discrete Gaussian steps, which is what we will do in the actual analysis.)

Consider a random walk in $\mathcal{P}$ corresponding to the Brownian motion starting at $x = x_0$. Whenever the random walk reaches a face of the polytope, it continues inside this face. We continue the walk until we reach a vertex $x \in \mathcal{P}$. The idea being that we want to get away from origin, but do not want to cross the polytope – so whenever a constraint (variable or discrepancy) becomes tight we do not want to change the constraint and continue in the subspace orthogonal to the defining constraint. We call this random walk the "Edge-Walk" in $\mathcal{P}$.

By definition, the random walk is constrained to $\mathcal{P}$, and $|\langle x - x_0, v_j\rangle| \leq c_j$ for all $j \in [m]$. We show that as long as $\sum \exp(-c_j^2) \ll n$, the random walk hits many variable constraints with good probability. That is, the end vertex x has $x_i \in \{-1, 1\}$ for many indices. This step relies on a martingale tail bound for Gaussian variables and an implicit use of the ℓ_2-norm as a potential function for gauging the number of coordinates close to 1 in absolute value.

The actual algorithm differs slightly from the above description. First, we will not run the walk until we reach a vertex of $\mathcal{P}$, but after a certain 'time' has passed, which will still guarantee the above conditions. Second, we will approximate the continuous random walk by many small discrete steps.

III. Comparison with Entropy Method

Here we contrast our result with Beck's partial coloring lemma [4] based on the Entropy method which has many applications in discrepancy theory. While similar in spirit, our partial coloring lemma is incomparable and in particular, even the existence of the vector x as in Theorem 4 does not follow from Beck's partial coloring lemma.

We first state Beck's partial coloring lemma as formulated in [11].

Theorem 5 (Entropy Method)**.** *Let $(V, \mathcal{S})$ be a set-system with $V = \{1, \ldots, n\}$. Let $\Delta : \mathcal{S} \to \mathbb{R}_+$ be such that $\sum_{S\in\mathcal{S}} g(\Delta_S/\sqrt{|S|}) \leq n/5$, where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by,*

$$g(\lambda) = \begin{cases} Ke^{-\lambda^2/9}, & \lambda > 0.1 \\ K\ln(1/\lambda), & \lambda \leq 0.1 \end{cases},$$

where K is an absolute constant. Then, there exists $\chi \in \{-1, 0, 1\}^n$ with $|\{i : \chi_i \neq 0\}| \geq n/2$ such that $|\sum_{i\in S} \chi_i| \leq \Delta_S$ for every $S \in \mathcal{S}$.

By applying our Theorem 4 to the indicator vectors of the sets in $\mathcal{S}$ and $\delta = 1/\text{poly}(n)$ sufficiently small we get the following corollary.

Corollary 6. *Let $(V, \mathcal{S})$ be a set-system with $V = \{1, \ldots, n\}$. Let $\Delta : \mathcal{S} \to \mathbb{R}_+$ be such that*

$$\sum_{S\in\mathcal{S}} \exp(-\Delta_S^2/16|S|) \leq n/16.$$

Then, there exists $\chi \in [-1, 1]^n$ with $|\{i : |\chi_i| = 1\}| \geq n/2$, such that $|\sum_{i\in S} \chi_i| \leq \Delta_S + 1/\text{poly}(n)$, for every $S \in \mathcal{S}$. Moreover, there exists a randomized $\text{poly}(|\mathcal{S}|, n)$-time algorithm to find χ.

The above result strengthens the Entropy method in two important aspects. Firstly, our method is constructive. In contrast, the entropy method is non-constructive and the constructive discrepancy minimization algorithms of Bansal do not yield the full partial coloring lemma as in Theorem 5. Secondly, the above result can tolerate many more *stringent constraints* than the Entropy method. For instance, the entropy method can only allow $O(n/\log n)$ of the sets in $\mathcal{S}$ to have discrepancy $1/n$, whereas our result can allow $\Omega(n)$ of the sets to have such small discrepancy. We believe that this added flexibility in achieving much smaller discrepancy for a constant fraction of sets could be useful elsewhere.

One weakness of Theorem 4 is that we do not strictly speaking get a proper partial coloring: the non $\{1, -1\}$ variables in our coloring χ can take any value in $(-1, 1)$. This however does not appear to be a significant drawback, as Corollary 6 can also be made to work from an arbitrary *starting point* x_0 as in the statement of Theorem 4.

IV. Preliminaries

We start with some notation and few elementary properties of the Gaussian distributions.

A. Notation

Let $[n] = \{1, \ldots, n\}$. Let $e_1, \ldots, e_n$ denote the standard basis for $\mathbb{R}^n$. We denote random variables by capital letters and distributions by calligraphic letters. We write $X \sim \mathcal{D}$ for a random variable X distributed according to a distribution $\mathcal{D}$.

B. Gaussian distribution

Let $\mathcal{N}(\mu, \sigma^2)$ denote the Gaussian distribution with mean μ and variance σ^2. A Gaussian distribution is called *standard* if $\mu = 0$ and $\sigma^2 = 1$. If $G_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $G_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ then for $t_1, t_2 \in \mathbb{R}$ we have

$$t_1 G_1 + t_2 G_2 \sim \mathcal{N}(t_1\mu_1 + t_2\mu_2, t_1^2\sigma_1^2 + t_2^2\sigma_2^2).$$

Let $V \subseteq \mathbb{R}^n$ be a linear subspace. We denote by $G \sim \mathcal{N}(V)$ the standard multi-dimensional Gaussian distribution supported on V: $G = G_1 v_1 + \ldots + G_d v_d$, where $\{v_1, \ldots, v_d\}$ is an orthonormal basis for V and $G_1, \ldots, G_d \sim \mathcal{N}(0, 1)$ are independent standard Gaussian variables. It is easy to check that this definition is invariant of the choice of the basis $\{v_1, \ldots, v_d\}$. We will need the following simple claims.

Claim 7. *Let $V \subseteq \mathbb{R}^n$ be a linear subspace and let $G \sim \mathcal{N}(V)$. Then, for all $u \in \mathbb{R}^n$, $\langle G, u\rangle \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 \le \|u\|_2^2$.*

Proof: Let $G = G_1 v_1 + \ldots + G_d v_d$ where $\{v_1, \ldots, v_d\}$ is an orthonormal basis for V and $G_1, \ldots, G_d \sim \mathcal{N}(0, 1)$ are independent. Then $\langle G, u\rangle = \sum_{i=1}^d \langle u, v_i\rangle \cdot G_i$ is Gaussian with mean zero and variance $\sum_{i=1}^d \langle u, v_i\rangle^2 \le \|u\|_2^2$. ∎

Claim 8. *Let $V \subseteq \mathbb{R}^n$ be a linear subspace and let $G \sim \mathcal{N}(V)$. Let $\langle G, e_i\rangle \sim \mathcal{N}(0, \sigma_i^2)$. Then $\sum_{i=1}^n \sigma_i^2 = \dim(V)$.*

Proof: Let $G = G_1 v_1 + \ldots + G_d v_d$ where $v_1, \ldots, v_d$ are an orthonormal basis for V and $G_1, \ldots, G_d \sim \mathcal{N}(0, 1)$ are independent. Then, $\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \mathbb{E}[|\langle G, e_i\rangle|^2] = \mathbb{E}[\|G\|_2^2] = \sum_{i=1}^d \|v_i\|_2^2 \cdot \mathbb{E}[G_i^2] = d = \dim(V)$. ∎

The following is a standard tail bound for Gaussian variables.

Claim 9. *Let $G \sim N(0, 1)$. Then, for any $\lambda > 0$, $\Pr[|G| \ge \lambda] \le 2\exp(-\lambda^2/2)$.*

We will also need the following tail bound on martingales with Gaussian steps. It is a mild generalization of Lemma 2.2 in [6] and we omit the proof.

Lemma 10 ([6])**.** *Let $X_1, \ldots, X_T$ be random variables. Let $Y_1, \ldots, Y_T$ be random variables where each Y_i is a function of X_i. Suppose that for all $1 \le i \le T$, $x_1, \ldots, x_{i-1} \in \mathbb{R}$, $Y_i | (X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1})$ is Gaussian with mean zero and variance at most one (possibly different for each setting of $x_1, \ldots, x_{i-1}$). Then for any $\lambda > 0$,*

$$\Pr[|Y_1 + \ldots + Y_T| \ge \lambda\sqrt{T}] \le 2\exp(-\lambda^2/2).$$

V. Main Partial Coloring Lemma

We are now ready to present our main *partial coloring* algorithm and prove Theorem 4. We shall use the notation from the theorem statement and Section II-A.

Let $\gamma > 0$ be a small step size so that $\delta = O(\gamma\sqrt{\log(nm/\gamma)})$. We note that the correctness of the algorithm is not affected by the choice of γ, as long as it is small enough; only the running time is affected.

Let $T = K_1/\gamma^2$, where $K_1 = 16/3$. We assume that $\delta < 0.1$. The algorithm will produce intermediate steps $X_0 = x_0, X_1, \ldots, X_T \in \mathbb{R}^n$ according to the following update process[2]

a) Edge-Walk:: For $t = 1, \ldots, T$ do

- Let $\mathcal{C}_t^{\text{var}} := \mathcal{C}_t^{\text{var}}(X_{t-1}) = \{i \in [n] : |(X_{t-1})_i| \ge 1 - \delta\}$ be the set of variable constraints 'nearly hit' so far.
- Let $\mathcal{C}_t^{\text{disc}} := \mathcal{C}_t^{\text{disc}}(X_{t-1}) = \{j \in [m] : |\langle X_{t-1} - x_0, v_j\rangle| \ge c_j - \delta\}$ be the set of discrepancy constraints 'nearly hit' so far.
- Let $\mathcal{V}_t := \mathcal{V}(X_{t-1}) = \{u \in \mathbb{R}^n : u_i = 0\ \forall i \in \mathcal{C}_t^{\text{var}},\ \langle u, v_j\rangle = 0\ \forall j \in \mathcal{C}_t^{\text{disc}}\}$ be the linear subspace orthogonal to the 'nearly hit' variable and discrepancy constraints.
- Set $X_t := X_{t-1} + \gamma U_t$, where $U_t \sim \mathcal{N}(\mathcal{V}_t)$.

The following lemma captures the essential properties of the random walk.

Lemma 11. *Consider the random walk described above. Assume that $\sum_{j=1}^m \exp(-c_j^2/16) \le n/16$. Then, with probability at least 0.1,*

1) $X_0, \ldots, X_T \in \mathcal{P}$.
2) $|(X_T)_i| \ge 1 - \delta$ *for at least* $n/2$ *indices* $i \in [n]$.

Theorem 4 follows immediately from Lemma 11 by setting $x = X_T$. Note that computing $\mathcal{C}_t^{\text{var}}, \mathcal{C}_t^{\text{disc}}$, given X_{t-1} takes time $O(nm)$. Further, once we know the set of constraints defining $\mathcal{V}_t$, we can sample from $\mathcal{N}(\mathcal{V}_t)$ in time $O((n+m)^3)$ by first constructing an orthogonal basis U for $\mathcal{V}_t$ and setting $U_t = \sum_{u \in U} G_u u$, where $G_u \sim \mathcal{N}$ are chosen independently.

We prove Lemma 11 in the remainder of this section. We start with a simple observation that $\mathcal{C}_t^{\text{var}}, \mathcal{C}_t^{\text{disc}}$ can only increase during the random walk.

Claim 12. *For all $t < T$ we have $\mathcal{C}_t^{var} \subseteq \mathcal{C}_{t+1}^{var}$ and $\mathcal{C}_t^{disc} \subseteq \mathcal{C}_{t+1}^{disc}$. In particular, for $1 \le t < T$, $dim(\mathcal{V}_t) \ge dim(\mathcal{V}_{t+1})$.*

Proof: Let $i \in \mathcal{C}_t^{\text{var}}$. That is, $|(X_{t-1})_i| \ge 1 - \delta$. Then by definition of the random walk, $U_t \in \mathcal{V}_t$ and

[2]We call the random walk "Edge-Walk" because geometrically, once the walk (almost) hits an edge (face) of the polytope $\mathcal{P}$, it stays on the edge.

$(U_t)_i = 0$. Thus, $(X_t)_i = (X_{t-1})_i$ and $i \in \mathcal{C}^{\text{var}}_{t+1}$. The argument for discrepancy constraints is analogous. ∎

We next show that the walk stays inside $\mathcal{P}$ with high probability.

Claim 13. *For $\gamma \le \delta/\sqrt{C\log(mn/\gamma)}$ and C a sufficiently large constant, with probability at least $1 - 1/(mn)^{C-2}$, $X_0, \ldots, X_T \in \mathcal{P}$.*

Proof: If one of the points $X_0, .., X_T$ is not in $\mathcal{P}$, then it is necessary that at least once the random walk makes a jump of size at least δ in the direction of one of the vectors in $W := \{e_1, \ldots, e_n, v_1, \ldots, v_m\}$. Consider a single step t, $1 \le t \le T$ and a vector $w \in W$ and condition on any subspace $\mathcal{V}_t$. Since $U_t \sim \mathcal{N}(\mathcal{V}_t)$ we have by Claim 7 that $\langle U_t, w\rangle$ is Gaussian with mean 0 and variance at most 1. Hence by Claim 9,

$$\Pr[|\langle U_t, w\rangle| \ge \delta/\gamma] \le 2\exp(-(\delta/\gamma)^2/2).$$

Hence, by a union bound,

$$\Pr[\exists t,\ X_t \notin \mathcal{P}] \le T \cdot (m+n) \cdot 2\exp(-(\delta/\gamma)^2/2) < 1/(mn)^{C-2},$$

for C large enough. ∎

We are now ready to prove Lemma 11. The intuition behind the proof is as follows. We first use the hypothesis on the thresholds $c_j, j \in [m]$, to argue that $\mathbb{E}[\,|\mathcal{C}^{\text{disc}}_T|\,] \ll n$. This follows from the definition of the walk and a simple application of the martingale tail bound of Lemma 10. Note that to prove the lemma it essentially suffices to argue that $\mathbb{E}[|\mathcal{C}^{\text{var}}_T|] = \Omega(n)$ (we can then use Markov's inequality). Roughly speaking, we do so by a "win-win" analysis. Consider an intermediate update step $t \le T$. Then, either $|\mathcal{C}^{\text{var}}_t|$ is large, in which case we are done, or $|\mathcal{C}^{\text{var}}_t|$ is small in which case $dim(\mathcal{V}_{t-1})$ is large so that $\mathbb{E}[\|X_t\|^2]$ increases significantly (with noticeable probability) due to Claim 8. On the other hand, $\|X_t\|^2 \le n$ as all steps stay within the polytope $\mathcal{P}$ (with high probability). Hence, $|\mathcal{C}^{\text{var}}_t|$ cannot be small for many steps and in particular $|\mathcal{C}^{\text{var}}_T|$ will be large with noticeable probability.

We first argue that $\mathbb{E}[\,|\mathcal{C}^{\text{disc}}_T|\,]$ is small. That is, on average only a few discrepancy constraints are ever nearly hit.

Claim 14. $\mathbb{E}[|\mathcal{C}^{\textit{disc}}_T|] < n/4$.

Proof: Let $J := \{j : c_j \le 10\delta\}$. To bound the size of J, we have

$$n/16 \ge \sum_{j\in J} \exp(-c_j^2/16) \ge |J| \cdot \exp(-100\delta^2/16) \ge |J| \cdot \exp(-1/16) > 9|J|/10,$$

and hence $|J| \le 1.2n/16$. Now, for $j \notin J$, if $j \in \mathcal{C}^{\text{disc}}_T$, then $|\langle X_T - x_0, v_j\rangle| \ge c_j - \delta \ge 0.9c_j$. We will bound the probability that this occurs. Recall that $X_T = x_0 + \gamma(U_1 + \ldots + U_T)$ and define $Y_i = \langle U_i, v_j\rangle$. Then, for $j \notin J$, we have

$$\Pr[j \in \mathcal{C}^{\text{disc}}_T] \le \Pr[\,|Y_1 + \ldots + Y_T| \ge 0.9c_j/\gamma].$$

We next apply Lemma 10. Note that the conditions of the lemma apply, since $U_1, \ldots, U_T$ is a sequence of random variables, Y_i is a function of U_i and $Y_i|(U_1, \ldots, U_{i-1})$ is Gaussian with mean zero and variance at most one (by Claim 7). Hence,

$$\Pr[j \in \mathcal{C}^{\text{disc}}_T] \le 2\exp(-(0.9c_j)^2/2\gamma^2 T) = 2\exp(-(0.9c_j)^2/2K_1 T) < 2\exp(-c_j^2/16).$$

So

$$\mathbb{E}[|\mathcal{C}^{\text{disc}}_T|] \le |J| + \sum_{j\notin J} \Pr[j \in \mathcal{C}^{\text{disc}}_T] \le \frac{1.2n}{16} + \frac{2n}{16} < \frac{n}{4}.$$

∎

Claim 15. $\mathbb{E}[\|X_T\|_2^2] \le n$.

Proof: We will show that $\mathbb{E}[(X_T)_i^2] \le 1$ for all $i \in [n]$. Conditioning on the first t for which $i \in \mathcal{C}^{\text{var}}_t$ (or that no such t exists), we get

$$\mathbb{E}[(X_T)_i^2] = \Pr[i \notin \mathcal{C}^{\text{var}}_T]\,\mathbb{E}[(X_T)_i^2 | i \notin \mathcal{C}^{\text{var}}_T] + \sum_{t=1}^{T} \Pr[i \in \mathcal{C}^{\text{var}}_t \setminus \mathcal{C}^{\text{var}}_{t-1}]\,\mathbb{E}[(X_T)_i^2 | i \in \mathcal{C}^{\text{var}}_t \setminus \mathcal{C}^{\text{var}}_{t-1}].$$

Clearly $\mathbb{E}[(X_T)_i^2 | i \notin \mathcal{C}^{\text{var}}_T] \le 1$. For $t \le T$, we have

$$\mathbb{E}[(X_T)_i^2 | i \in \mathcal{C}^{\text{var}}_t \setminus \mathcal{C}^{\text{var}}_{t-1}] = \mathbb{E}[(X_t)_i^2 | i \in \mathcal{C}^{\text{var}}_t \setminus \mathcal{C}^{\text{var}}_{t-1}] \le 1 - \delta + \gamma\mathbb{E}[|(U_t)_i|_2^2] \le 1,$$

where we used the fact that $(U_t)_i$ is a Gaussian variable with mean zero and variance at most one (by Claim 7). ∎

Finally, we show that $\mathbb{E}[|\mathcal{C}^{\text{var}}_T|]$ is large. That is, on average we will nearly hit a constant fraction of the variable constraints.

Claim 16. $\mathbb{E}[\,|\mathcal{C}^{\textit{var}}_T|\,] \ge 0.56n$.

Proof: We start by computing the average norm of X_t.

$$\mathbb{E}[\|X_t\|_2^2] = \mathbb{E}[\|X_{t-1} + \gamma U_t\|_2^2] = \mathbb{E}[\|X_{t-1}\|_2^2] + \gamma^2\mathbb{E}[\|U_t\|_2^2] = \mathbb{E}[\|X_{t-1}\|_2^2] + \gamma^2\mathbb{E}[\dim(\mathcal{V}_t)],$$

where we used that fact that given X_{t-1}, $\mathbb{E}[U_t|X_{t-1}] = 0$ and $\mathbb{E}[\|U_t\|_2^2|X_{t-1}] = \dim(\mathcal{V}_t)$, by Claim 8. Hence, by Claim 15,

$$n \geq \mathbb{E}[\|X_T\|_2^2] \geq \gamma^2 \sum_{t=1}^{T} \mathbb{E}[\dim(\mathcal{V}_t)] \geq \gamma^2 |T| \cdot \mathbb{E}[\dim(\mathcal{V}_T)] = K_1 \cdot \mathbb{E}[\dim(\mathcal{V}_T)] = K_1 \mathbb{E}[(n - |\mathcal{C}_T^{\text{var}}| - |\mathcal{C}_T^{\text{disc}}|)].$$

Therefore, $\mathbb{E}[|\mathcal{C}_T^{\text{var}}|] \geq n(1 - 1/K_1) - \mathbb{E}[|\mathcal{C}_T^{\text{disc}}|] \geq n(1 - 1/K_1 - 1/4) > (0.56)n$, where the second inequality follows from Claim 14. ∎

Lemma 11 now follows immediately from Claim 13 and Claim 16.

Proof of Lemma 11: From Claim 16 and the fact that $|\mathcal{C}_T^{\text{var}}| \leq n$, it follows that $\mathbb{P}[|\mathcal{C}_T^{\text{var}}| \geq n/2] \geq 0.12$. Combining with Claim 13, with probability at least $0.12 - 1/poly(m,n) > 0.1$, $|\mathcal{C}_T^{\text{var}}| \geq n/2$ and $X_T \in \mathcal{P}$ which shows the lemma. ∎

VI. Discrepancy Minimization from Partial Coloring

We now derive Theorem 2 and Theorem 3 from our partial coloring lemma.

Proof of Theorem 2: Let $(V, \mathcal{S})$ be a system with $|V| = n$ and $|\mathcal{S}| = m$. Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be the indicator vectors of the sets in $\mathcal{S}$. We set $\delta = 1/(8 \log m)$. Let $\alpha(m,n) = 8\sqrt{\log(m/n)}$. Then, $m \cdot \exp(-\alpha(m,n)^2/16) < n/16$. Therefore, by Theorem 4 applied to $v_1, \ldots, v_m$ and starting point $x_0 = 0^n$, with probability at least 0.1 we find a vector $x_1 \in [-1,1]^n$ such that $|\langle v_j, x_1 \rangle| < \sqrt{n} \cdot \alpha(m,n)$ for all $j \in m$ and $|\{i : |(x_1)_i| \geq 1 - \delta\}| \geq n/2$. We can boost this probability further by repeating the algorithm $O(\log n)$ times; from now on we will ignore the probability that the algorithm does not find such a vector.

Let $I_1 = \{i : |(x_1)_i| < 1 - \delta\}$ be the coordinates not 'fixed' in the first step and set $n_1 = |I_1|$. We now iteratively apply Theorem 4 to the restricted system described by the vectors $v_1^1 = (v_1)_{I_1}, \ldots, v_m^1 = (v_m)_{I_1} \in \mathbb{R}^{n_1}$ and starting point $(x_1)_{I_1}$ to get another vector $x_2 \in [-1,1]^{n_1}$ such that $|\langle v_j^1, x_2 \rangle| < \sqrt{n_1} \cdot \alpha(m, n_1)$ for all $j \in [m]$ and $|\{i : |(x_2)_i| \geq 1 - \delta\}| \geq n_1/2$. By iterating this procedure for at most $t = 2 \log n$ times and concatenating the resulting vectors appropriately we get $x \in \mathbb{R}^n$ such that $|x_i| \geq 1 - \delta$ for all $i \in [n]$ and for every $j \in [m]$,

$$\begin{aligned} |\langle v_j, x \rangle| &< \sqrt{n} \cdot \alpha(m,n) + \cdots + \sqrt{n_t} \cdot \alpha(m, n_t) \\ &< \sqrt{n} \sum_{r=0}^{\infty} \frac{8\sqrt{\log(m \cdot 2^r/n)}}{2^{r/2}} \\ &< C\sqrt{n \cdot \log(m/n)}, \end{aligned}$$

for C a universal constant.

We now round x to get a proper coloring $\chi \in \{1,-1\}^n$. Let $\chi \in \{1,-1\}^n$ be obtained from x as follows: for $i \in [n]$, $\chi_i = sign(x_i)$ with probability $(1+|x_i|)/2$ and $-sign(x_i)$ with probability $(1-|x_i|)/2$, so that $\mathbb{E}[\chi_i] = x_i$. Let $Y = \chi - x$. Fix some $j \in [m]$. Then, the discrepancy of χ with v_j is

$$|\langle \chi, v_j \rangle| \leq |\langle x, v_j \rangle| + |\langle Y, v_j \rangle| \leq C\sqrt{n \log(m/n)} + |\langle Y, v_j \rangle|.$$

We will show that with high probability, $|\langle Y, v_j \rangle| \leq \sqrt{n}$ for all $1 \leq j \leq m$. Fix some $j \in [m]$ and consider $\langle Y, v_j \rangle$. We have that $|Y_i| \leq 2$, $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) \leq \delta$. We also have $\|v_j\|_2 \leq \sqrt{n}$ and $\|v_j\|_\infty \leq 1$. Thus, by a standard Chernoff bound (see e.g., Theorem 2.3 in [12]),

$$\mathbb{P}\left[|\langle Y, v_j \rangle| > 2\sqrt{2 \log m} \cdot \sqrt{n\delta}\right] \leq 2\exp(-2 \log m) < 1/2m.$$

Therefore, by the union bound and our choice of δ, with probability at least $1/2$ we have that $|\langle Y, v_j \rangle| \leq \sqrt{n}$ for all $1 \leq j \leq m$. Therefore, $|\langle \chi, v_j \rangle| \leq C\sqrt{n \log(m/n)} + \sqrt{n}$ for all $1 \leq j \leq m$.

The running time is dominated by the $O(\log^2 n)$ uses of Theorem 4. Thus, the total running time is $O((n+m)^3 \log^5(mn)) = \tilde{O}((n+m)^3)$. ∎

The constant in the theorem can be sharpened to be 13 by fine tuning the parameters. We do not dwell on this here. We next prove Theorem 3.

Proof of Theorem 3: The proof is similar to the above argument and we only sketch the full proof. Set $\delta = 1/n$. Let $(V, \mathcal{S})$ be the set system. Let $v_1, \ldots, v_m$ be the indicator vectors of the sets in $\mathcal{S}$ and let $c_j = C\sqrt{t}/\|v_j\|_2$ for C to be chosen later. Observe that $\sum_j \|v_j\|_2^2 \leq nt$ as each element appears in at most t sets. In particular, the number of vectors v_j with $\|v_j\|_2^2$ in $[2^r t, 2^{r+1} t]$ is at most $n/2^r$. Therefore,

$$\sum_j \exp(-c_j^2/16) < \sum_{r=0}^{\infty} \frac{n \cdot \exp(-C^2/16 \cdot 2^{r+1})}{2^r} < \frac{n}{16},$$

for C a sufficiently large constant. Thus, by applying Theorem 4 to the vectors v_j and thresholds c_j for $j \in [m]$, with probability at least 0.1 we get a vector $x_1 \in [-1,1]^n$ such that $|\langle v_j, x_1 \rangle| < C\sqrt{t}$ for all $j \in [m]$ and $|\{i : |(x_1)_i| \geq 1 - \delta\}| > n/2$.

By iteratively applying the same argument as in the proof of Theorem 4 for $2 \log n$ steps, we get a vector $x \in [-1,1]^n$ with $|x_i| \geq 1 - \delta$ for all i and $|\langle v_j, x \rangle| < 2C\sqrt{t} \log n$ for all $j \in [m]$. The theorem now follows by rounding the x to the nearest integer coloring χ: $\chi_i = sign(x_i)$ for all $i \in [m]$. ∎

b) Acknowledgments: We would like to thank Oded Regev for many discussions and collaboration at the early stages of this work. We thank Joel Spencer for his encouragement and enthusiasm about this work: part of our presentation is inspired by a lecture he gave on this result at the Institute for Advanced Study, Princeton. We also thank him for the observation on improving the run time of Theorem 4 and allowing us to include it here. We thank Nikhil Bansal for valuable comments and discussions.

REFERENCES

[1] J. Matoušek, *Geometric Discrepancy: An Illustrated Guide*, ser. Algorithms and Combinatorics. Springer, 1999.

[2] B. Chazelle, *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2002.

[3] J. Spencer, "Six standard deviations suffice," *Transactions of the American Mathematical Society*, vol. 289, no. 2, pp. 679–706, 1985.

[4] J. Beck, "Roths estimate of the discrepancy of integer sequences is nearly sharp," *Combinatorica*, vol. 1, no. 4, pp. 319–325, 1981.

[5] N. Alon and J. Spencer, *The Probabilistic Method*, ser. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, 2011. [Online]. Available: http://books.google.com/books?id=XLwRrn4rpk4C

[6] N. Bansal, "Constructive algorithms for discrepancy minimization," in *FOCS*, 2010, pp. 3–10.

[7] M. Charikar, A. Newman, and A. Nikolov, "Tight hardness results for minimizing discrepancy," in *SODA*, 2011, pp. 1607–1614.

[8] A. Srinivasan, "Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems," in *ACM-SIAM Symposium on Discrete Algorithms*, 1997, pp. 692–701.

[9] W. Banaszczyk, "Balancing vectors and gaussian measures of n-dimensional convex bodies," *Random Struct. Algorithms*, vol. 12, no. 4, pp. 351–360, 1998.

[10] J. Beck and T. Fiala, "Integer-making theorems," *Discrete Applied Mathematics*, vol. 3, no. 1, pp. 1–8, 1981.

[11] J. Matoušek, "An l_p version of the beck-fiala conjecture," *Eur. J. Comb.*, vol. 19, no. 2, pp. 175–182, 1998.

[12] F. Chung and L. Lu, *Complex Graphs and Networks*. American Mathematical Society, 2006.