Constructive Discrepancy Minimization by Walking on The Edges

by *Shachar Lovett and Raghu Meka*

presented by Martin Balko

The authors introduce a new randomized algorithm which finds a coloring which achieves discrepancy $C\sqrt{n}$. The 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 the *partial coloring lemma*.

Definitions:

- We are given a collection of m sets $\mathcal{S}$ from a universe $V = \{1, \ldots, n\}$. Let no element from V be in more than t sets of $\mathcal{S}$.
- The goal is to find a *coloring* $\chi\colon V \to \{-1, 1\}$ that minimizes the maximum *discrepancy* $\chi(\mathcal{S}) = \max_{S \in \mathcal{S}} \left|\sum_{i \in S} \chi(i)\right|$. The minimum discrepancy of $\mathcal{S}$ is denoted as $\mathrm{disc}(\mathcal{S}) = \min_\chi \chi(\mathcal{S})$.

Known:

- A random coloring has discrepancy $O(\sqrt{n \log m})$.
- For t bounded $\mathrm{disc}(\mathcal{S}) < 2t$ holds [Beck and Fiala, 1981] and $\mathrm{disc}(\mathcal{S}) = O(\sqrt{t})$ is conjectured.
- For t bounded $\mathrm{disc}(\mathcal{S}) = O(\sqrt{t \cdot \log n})$ holds [Banaszczsyk, 1998], non-constructively.

  Theorem 1 (Standard deviation result, Spencer 1985)**.** *For any set system $(V, \mathcal{S})$ with $|V| = n$, $|\mathcal{S}| = m$, there exists a coloring $\chi\colon V \to \{-1, 1\}$ such that $\chi(\mathcal{S}) < K\sqrt{n \cdot \log_2(m/n)}$, where K is a universal constant (K can be six if $m = n$).*

- Spencer's original proof was non-constructive. A longstanding problem: is there an efficient way to find a good coloring as in Theorem 1?
- Bansal gave the first randomized polynomial time algorithm to find coloring with discrepancy $O(\sqrt{n} \cdot \log(m/n))$ [Bansal, 2010].

The results:

- A new algorithm which gives a new constructive proof of Spencer's original result.

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

- A similar constructive result for minimizing discrepancy in the "Beck-Fiala setting" where each variable is constrained to occur in a bounded number of sets.

  Theorem 3. *Let (V, 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 in running time $\tilde{O}((n+m)^5)$ that with probability at least $1/2$ computes a coloring $\chi\colon V \to \{-1, 1\}$ such that $\chi(\mathcal{S}) < K\sqrt{t} \cdot \log n$, where K is a universal constant.*

Outline of the Edge-Walk Algorithm:

- A *partial coloring* $\chi\colon V \to [-1, 1]$ such that for all $S \in \mathcal{S}$, $|\chi(S)| = O(\sqrt{n \log(m/n)})$ and $|\{i : |\chi(i)| = 1\}| \geq cn$ for a fixed constant $c > 0$.

  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 tresholds 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 $|\langle x - x_0, v_j\rangle| \leq c_j \|v_j\|_2$ and $|x_i| \geq 1 - \delta$ for at least $n/2$ indices $i \in [n]$. Moreover, the algorithm runs in time $O((m+n)^3 \cdot \delta^{-2} \cdot \log(nm/\delta))$.*

- Theorem 4 implies Theorem 2 and Theorem 3.
- A polytope $\mathcal{P} = \{x \in \mathbb{R}^n : |x_i| \leq 1\ \forall i \in [n], |\langle x - x_0, v_j\rangle| \leq c_j\ \forall j \in [m]\}$ defined by *variable constraints* $|x_i| \leq 1$ and *discrepancy constraints* $|\langle x - x_0, v_j\rangle| \leq c_j$.

Preliminaries for the proof of Theorem 4:

- Let $\mathcal{N}(\mu, \sigma^2)$ denote the *Gaussian distribution* with mean μ and variance σ^2. For $\mu = 0$ and $\sigma^2 = 1$ we call it *standard.*
- For a linear subspace $V \subseteq \mathbb{R}^n$ we denote by $G \sim \mathcal{N}(V)$ the *standard multi-dimensional Gaussian distribution supported on* V: $G = G_1 v_1 + \cdots + G_d v_d$ where $\{v_1, \ldots, v_d\}$ is an orthonormal basis for V and $G_1, \ldots, G_d \sim \mathcal{N}(0,1)$.

  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 $\sigma^2 \leq \|u\|_2^2$.*

  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 $\sum_{i=1}^n \sigma_i^2 = \dim(V)$.*

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

  Lemma 10 (Bansal, 2010)**.** *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 $1 \leq t \leq T$, $x_1, \ldots, x_{i-1} \in \mathbb{R}$, $Y_i \mid (X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$ 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 $\lambda > 0$, $\Pr[|Y_1 + \cdots + Y_T| \geq \lambda\sqrt{T}] \leq 2\exp(-\lambda^2/2)$.*

Proof of Theorem 4:

- In each step t, $1 \leq t \leq T$, set

$$\begin{aligned}
\mathcal{C}_t^{var} &= \{i \in [n] : (X_{t-1})_i \geq 1 - \delta\},\\
\mathcal{C}_t^{disc} &= \{j \in [m] : |\langle X_{t-1} - x_0, v_j\rangle| \geq c_j - \delta\},\\
\mathcal{V}_t &= \{u \in \mathbb{R}^n : u_i = 0 \ \forall i \in \mathcal{C}_t^{var}, \langle u, v_j\rangle = 0 \ \forall j \in \mathcal{C}_t^{disc}\}.
\end{aligned}$$

- A crucial lemma:

  Lemma 11. *Assume that $\sum_{j=1}^m \exp(-c_j^2/16) \leq n/16$. Then in our random walk with probability at least 0.1 we have $X_0, \ldots, X_T \in \mathcal{P}$ and $|(X_T)_i| \geq 1 - \delta$ for at least $n/2$ indices $i \in [n]$.*

- Auxiliary results:

  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 $1 \leq t \leq T$, $\dim(\mathcal{V}_t) \geq \dim(\mathcal{V}_{t+1})$.*

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

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

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

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