CSE 525: Randomized Algorithms **Spring 2025**

Lecture 12: Discrepancy Theory and the Lovett Meka Algorithm

Lecturer: Shayan Oveis Gharan *05/15/2025*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

These notes are based on notes by Sasha Nikolov.

12.1 Introduction

Let A be an $m \times n$ real-valued matrix, the discrepancy of A is

$$\text{disc}(A) = \min_{x \in \{\pm 1\}^n} \|Ax\|_\infty.$$

When the entries of A are in $[-1, 1]$, we have an upper bound $\text{disc}(A) = O(\sqrt{n \log m})$, with high probability, by taking a uniformly random $x \in \{\pm 1\}^n$. We don't prove this here and we leave it as an exercise.

The following existential proved by Sperner in 1985.

Theorem 12.1 (Six Standard Deviations Suffice Sperner'85)**.** *Let A be any $m \times n$ matrix with entries in $[-1, 1]$. Then*

$$\text{disc}(A) = O(\sqrt{n \log(m/n)}).$$

For decades it remained an open problem whether this result can be proven algorithmically, i.e., whether there is an efficient algorithm to find a vector x satisfying the conclusion of the above theorem. This was resolved by Bansal in 2010 where he came with a randomized algorithm using SDP. In this lecture we describe a simpler proof/algorithm in a follow up work by Lovett and Meka.

12.2 The Algorithm

Theorem 12.2. *Let m, n be positive integers with $m \geq n$ and let A be any $m \times n$ matrix with entries from $[-1, 1]$. There is a randomized algorithm running in time polynomial in m, n which, when given A as input, outputs an $x \in \{\pm 1\}^n$ such that*

$$\|Ax\|_\infty = O\left(\sqrt{n \log \frac{m}{n}}\right)$$

with high probability.

In specifying and analyzing the algorithm, we take a geometric viewpoint. Let C be a universal constant to be specified later, and consider the convex polytope:

$$K = \left\{x \in \mathbb{R}^n : \|Ax\|_\infty \leq C\sqrt{n \log(8m/n)}\right\}.$$

If $a_1, a_2, \ldots, a_m$ are the rows of A, we can re-write this as:

$$K = \left\{ x \in \mathbb{R}^n : \forall i \in [m], |\langle a_i, x \rangle| \leq C\sqrt{n \log(8m/n)} \right\}.$$

Intuitively, K is a convex polytope defined as the intersection of "slabs" of the form:

$$|\langle a_i, x \rangle| \leq C\sqrt{n \log(8m/n)}.$$

The algorithm performs an approximate continuous random walk, starting from the origin, inside $K \cap [-1,1]^n$. When the walk intersects a facet of this polytope, it restricts further steps to that facet, until reaching a vertex of $K \cap [-1,1]^n$. Ideally, this vertex is in $\{\pm 1\}^n$, but we will show it has a constant fraction of coordinates in $\{\pm 1\}$.

We reduce Theorem 2 to the following result:

Theorem 12.3. *Let $m \geq n$ be positive integers and let $\delta = 1/\sqrt{n}$. There is a randomized polynomial-time algorithm and a constant C such that, given an $m \times n$ matrix A and a vector $x^{(0)} \in [-1,1]^n$, it finds an $x \in [-1,1]^n$ such that with probability at least $1/6 - \varepsilon$ (for any $1/6 > \varepsilon > 0$):*

1. *For each $i = 1, \ldots, m$, we have*

$$|\langle a_i, x - x^{(0)} \rangle| \leq C \|a_i\|_2 \sqrt{\log(8m/n)}.$$

2. *$|x_i| > 1 - \delta$ for at least $n/10$ coordinates i.*

Having this we prove Theorem 12.2.

Proof Sketch of Theorem 12.2 Start with $x^{(0)} = 0$, run the algorithm from Theorem 3 and obtain an $x \in [-1,1]^n$. Let x' be the vector where coordinates failing condition (2) are unset, and recursively apply the algorithm on $x^{(0)} = x'$ and the submatrix A' (removing satisfied columns). Since we fix a constant fraction of coordinates each time, after $S = 10 \log n$ steps, all indices satisfy condition (2). The total discrepancy is:

$$\begin{aligned}
\|Ax^*\|_\infty &\leq C\sqrt{n}\sqrt{\log(8m/n)} + C\sqrt{n/10}\sqrt{\log(8m/(n/10))} + \ldots \\
&< \sqrt{n} \sum_{s=0}^{\infty} \frac{C\sqrt{\log(8m \cdot 10^s/n)}}{10^{s/2}} < C'\sqrt{n \log(m/n)}
\end{aligned}$$

for some constant C'. Finally, rounding x^* to $\{\pm 1\}^n$ adds at most $O(\sqrt{n})$ discrepancy since $n/\delta = \sqrt{n}$.

Let $\mathcal{N}(\mu, \sigma^2)$ denote the mean μ Gaussian distribution with variance σ^2. The algorithm is formally described in Algorithm 1.

Algorithm 1 Main Algorithm

Input: $A \in \mathbb{R}^{m \times n}$, $x^{(0)} \in [-1, 1]^n$
Output: $x \in [-1, 1]^n$
for $t = 1$ to T **do**
  $D_t = \{i : |\langle a_i, x^{(t-1)} - x^{(0)} \rangle| \geq C\|a_i\|_2 \sqrt{\log(8m/n)} - \delta\}$
  $V_t = \{j : |x_j^{(t-1)}| > 1 - \delta\}$
  $W_t = \{y \in \mathbb{R}^n : \langle a_i, y \rangle = 0, \forall i \in D_t;\ y_j = 0, \forall j \in V_t\}$
  Let $\{w_1, \ldots, w_k\}$ be orthonormal basis of W_t
  Sample $g_1, \ldots, g_k \sim \mathcal{N}(0, 1)$
  $\Delta x^{(t)} = \sum_{i=1}^{k} g_i w_i$
  $x^{(t)} = x^{(t-1)} + \gamma \Delta x^{(t)}$
end for
return $x^{(T)}$

- The parameters δ, γ are small real values with $1/\sqrt{n} \geq \delta \gg \gamma > 0$. The integer T is chosen on the order of $1/\gamma^2$.
- The set D_t contains the facets of K for which $x^{(t-1)}$ is almost tight.
- The set V_t contains facets of the cube $[-1, 1]^n$ that are nearly tight.
- The subspace W_t is the orthogonal complement of those constraints, meaning movement along W_t maintains validity with respect to D_t and V_t.

In each iteration, we perturb $x^{(t-1)}$ by Gaussian noise projected onto W_t. This ensures:

- We do not increase discrepancy with respect to D_t.
- We do not alter coordinates already close to ± 1 (i.e., in V_t).

Discussion of Parameters:

- δ defines how close a coordinate must be to ± 1 to be considered "fixed." This helps us manage rounding error.
- γ is the step size in our random walk. We take a small step in a Gaussian direction within the valid subspace.

Gaussian Stability: If $g_1, \ldots, g_k$ are i.i.d. samples from $\mathcal{N}(0, 1)$, and $g = (g_1, \ldots, g_k)$, then for any $a \in \mathbb{R}^k$:

$$\langle a, g \rangle \sim \mathcal{N}(0, \|a\|_2^2).$$

This property is usually called rotation invariance property of Gaussians. We don't discuss the proof.

The following inequality is an analogue of the Azuma's inequality for Gaussian random variables.

Lemma 12.4. *Let $0 < \sigma \leq \tau$. Suppose $y_1 \sim \mathcal{N}(0, \sigma^2)$ and for all $i > 1$, $y_i - y_{i-1} \mid y_1, \ldots, y_{i-1} \sim \mathcal{N}(0, \sigma_i^2)$ with $0 < \sigma_i \leq \tau$. Then:*

$$\Pr\left[|y_\ell| > \lambda \cdot \tau \cdot \sqrt{\ell}\right] \leq 2e^{-\lambda^2/2}, \quad \textit{for all } \lambda > 0.$$

The main lemma now follows from the martingale concentration bound and the Stability of Gaus- sians. It says that if the random walk is not "too long" then, on average, the number of tight discrepancy constraints will be small.

Lemma 12.5. *If $T = O(1/\gamma^2)$, then there exists a constant C such that*

$$\mathbb{E}\left[|D_{T+1}|\right] = \mathbb{E}\left[\left|\left\{i \in [m] : |\langle a_i, x^{(T)} - x^{(0)}\rangle| \geq C\|a_i\|_2\sqrt{\log(8m/n)}\right\}\right|\right] \leq \frac{n}{4}.$$

Proof. Let C be a constant to be fixed later. By linearity of expectation,

$$\mathbb{E}[|D_{T+1}|] = \sum_{i=1}^{m} \Pr\left[|\langle a_i, x^{(T)} - x^{(0)}\rangle| \geq C\|a_i\|_2\sqrt{\log(8m/n)}\right].$$

We expand the inner product as:

$$\langle a_i, x^{(T)} - x^{(0)}\rangle = \sum_{t=1}^{T} \gamma\langle a_i, \Delta x^{(t)}\rangle,$$

where $\Delta x^{(t)} = \sum_{j=1}^{k} g_j w_j$ for i.i.d. $g_j \sim \mathcal{N}(0,1)$. By Gaussian stability,

$$\gamma\langle a_i, \Delta x^{(t)}\rangle = \gamma\sum_{j=1}^{k} g_j\langle a_i, w_j\rangle \sim \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \gamma^2\sum_{j=1}^{k}\langle a_i, w_j\rangle^2 \leq \gamma^2\|a_i\|_2^2$ since $\{w_j\}$ is orthonormal.

Let $y_t = \langle a_i, x^{(t)} - x^{(0)}\rangle$. The sequence $y_1, y_2, \ldots, y_T$ satisfies the conditions of the martingale concentration lemma with $\tau = \gamma\|a_i\|_2$. Thus,

$$\Pr\left[|\langle a_i, x^{(T)} - x^{(0)}\rangle| > \lambda \cdot \gamma\|a_i\|_2\sqrt{T}\right] \leq 2e^{-\lambda^2/2}.$$

Since $T = O(1/\gamma^2)$, choose $\lambda = C\sqrt{\log(8m/n)}$ with $C \geq 1/(\gamma\sqrt{T})$. Then

$$\Pr\left[|\langle a_i, x^{(T)} - x^{(0)}\rangle| > C\|a_i\|_2\sqrt{\log(8m/n)}\right] \leq \frac{n}{4m}.$$

Summing over all $i \in [m]$, we conclude:

$$\mathbb{E}[|D_{T+1}|] \leq \frac{n}{4}.$$

$\square$

Theorem 12.6 (Restated Theorem 12.3 – continued)**.** *Let $x = x^{(T)}$ be the output of Algorithm 1. For γ sufficiently small compared to δ, with high probability $x^{(t)} \in K \cap [-1,1]^n$ for all t. This implies $x^{(T)} \in [-1,1]^n$, satisfying the first property of Theorem 12.3.*

Let $x = x(T)$ be the output of Algorithm 1. It is not hard to show that for γ much smaller than δ, $x^{(t)} \in K \cap [-1,1]^n$ for all t with high probability. We leave this detail as an exercise. This immediately gives that $x^{(T)} \in [-1,1]^n$ and that the first property in Theorem 12.3 is satisfied.

We now prove the second property: that at least $n/10$ coordinates of x satisfy $|x_i| > 1 - \delta$.

Let us estimate $\mathbb{E}[|V_T|]$ using $\|x^{(T)} - x^{(0)}\|_2^2 \le n$. Recall that

$$x^{(T)} - x^{(0)} = \sum_{t=1}^{T} \gamma \Delta x^{(t)}.$$

For each t, let $\{w_1, \ldots, w_k\}$ be an orthonormal basis of W_t. Then:

$$\mathbb{E}[\|\Delta x^{(t)}\|_2^2] = \sum_{j=1}^{k} \mathbb{E}[g_j^2] = k = \dim(W_t) \ge n - \mathbb{E}[|D_t|] - \mathbb{E}[|V_t|].$$

The samples g_j are independent and have variance 1, so:

$$\mathbb{E}[\|x^{(T)} - x^{(0)}\|_2^2] = \gamma^2 \sum_{t=1}^{T} \mathbb{E}[\|\Delta x^{(t)}\|_2^2].$$

Since $|D_t| \le |D_{T+1}|$ and $|V_t| \le |V_T|$ since whenever the random walk is tight to a facet it remains tight, we get:

$$n \ge \gamma^2 \sum_{t=1}^{T} (n - \mathbb{E}[|D_T|] - \mathbb{E}[|V_T|]) = \gamma^2 T \left(n - \mathbb{E}[|D_T|] - \mathbb{E}[|V_T|]\right).$$

Choose $T = 2/\gamma^2$ to obtain:

$$\mathbb{E}[|V_T|] \ge \frac{1}{2}(n - 2\mathbb{E}[|D_T|]) \ge \frac{n}{4}.$$

Applying Markov's inequality:

$$\Pr\left[|V_T| < \frac{n}{10}\right] \le \frac{n - \mathbb{E}[|V_T|]}{9n/10} \le \frac{3n/4}{9n/10} = \frac{5}{6}.$$

Thus, with probability at least $1/6$, we have $|V_T| \ge n/10$. It is easy to verify that the algorithm runs in polynomial time.

References

[1] Shachar Lovett and Raghu Meka. *Constructive discrepancy minimization by walking on the edges.* FOCS 2012.

[2] Zbynek sidák. *Rectangular confidence regions for the means of multivariate normal distributions.* JASA, 1967.

[3] Joel Spencer. *Six standard deviations suffice.* Transactions of the AMS, 1985.