Lecture 12. Algorithmic Discrepancy

In the last lecture, we saw Spencer’s result on the discrepancy of set systems of n subsets of \([n]\): there is always a labeling \(l : [n] \rightarrow \{-1, +1\}\), such that \(|l(A_i)| = O(\sqrt{n})\) for each set \(A_i\). The question we discuss here is how to find such a labeling by an efficient algorithm. The first solution to this problem was due to Bansal (2008); his algorithm relied on semidefinite programming. Here we will follow a somewhat simpler approach developed later by Lovett-Meka (2010). We will prove “algorithmically” the following statement.

**Theorem 12.1** There is an absolute constant \(K > 0\) such that for any \(A_1, A_2, \ldots, A_m \subseteq [n]\) with \(m \geq n\), there is \(l : [n] \rightarrow \{\pm 1\}\) such that for every \(i \in [m]\),

\[|l(A_i)| \leq K \sqrt{n \log \frac{2m}{n}}.\]

We adopt a geometric approach, where each labeling is viewed as a point in \(\mathbb{R}^n\), and our constraints on it define a certain polytope. Let \(v_j = \frac{1}{|A_j|} \chi_{A_j}\) and think about the following polytope:

\[P = \{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid \forall i \in [n]; |x_i| \leq 1 \text{ and } \forall j \in [m] : |\langle v_j, x \rangle| < c_j\}\]

where \(c_j = K \sqrt{\log \frac{2m}{n}}\).

The algorithm is essentially a random walk that starts from the origin and moves inside the polytope as long as we don’t hit any constraint. If a constraint is hit then we continue inside the subspace defined by the constraint. More precisely, we will consider a Brownian motion starting at a point \(x_0\) of the polytope. Once we violate one of the conditions \(|x_i| \leq 1 - \frac{1}{2}\) or \(|\langle v_j, x \rangle| \leq c_j - \delta\) we freeze that constraint with equality (to the value attained) and we continue in a subspace defined by the respective constraint with equality. We hope to find a point which satisfies all the constraints and has many integer coordinates — that means we have constructed a partial coloring (similar to the one constructed in Spencer’s existential proof).

The following is a lemma that capture that main (desired) properties of our random walk.

**Lemma 12.2** Let \(\delta = \frac{1}{8 \log m}\). For every unit vectors \(v_1, v_2, \ldots, v_m \in \mathbb{R}^n\) and \(c_1, c_2, \ldots, c_m \geq 1\) such that \(\sum_{j=1}^{m} e^{-c_j^2/16} \leq \frac{n}{16}\) there is a random walk \(X_0, X_1, \ldots, X_T\) such that with constant probability, \(X_T \in [-1, 1]^n\) and

1. for at least \(n/2\) coordinates \(|(X_T)_i| \geq 1 - \delta\)
2. \(\forall j \in [m]; |\langle X_T - x_0, v_j \rangle| \leq c_j\)

**Remark 12.3** Notice that only \(n/2\) of the coordinates will be labeled \(\pm 1\) so we will need to iterate.
The random walk. Let $\gamma > 0$ and $\delta = \Theta(\sqrt{\log \frac{mn}{\gamma}})$. Define
\[
C_t^{\text{var}} = \{i \in [n] : |(X_{t-1})_i| \geq 1 - \delta\}
\]
be the nearly-tight variable constraints and
\[
C_t^{\text{disc}} = \{j \in [m] : |\langle v_j, X_{t-1} - x_0 \rangle| \geq c_j - \delta\}
\]
be the nearly-tight discrepancy constraints. The random walk behaves in such a way that once a constraint is nearly-tight, we preserve the value on that constraint — effectively getting stuck to the corresponding hyperplane. The next definition describes the subspace obtained by these restrictions.

**Definition 12.4** Define
\[
\mathcal{V}_t = \{y \in \mathbb{R}^n : y_t = 0 \ \forall i \in C_t^{\text{var}} \text{ and } \langle v_j, y \rangle = 0 \ \forall j \in C_t^{\text{disc}}\}.
\]
Each random step is Gaussian in this subspace. We define its distribution next.

**Definition 12.5** If a subspace $\mathcal{V}_t$ of $\mathbb{R}^n$ has an orthonormal basis $u_1, u_1, \ldots, u_d$ then define
\[
N(\mathcal{V}_t) = \text{the distribution of } G = G_1 u_1 + \ldots + G_d u_d, \text{ where } G_i \sim N(0, 1) \text{ are independent}.
\]

Recall the parameter $\gamma > 0$. The random step is $X_t = X_{t-1} + \gamma U_t$, where $U_t \sim N(\mathcal{V}_t)$, chosen independently of any prior history of the walk.

**Remark 12.6** $N(\mu, \sigma^2)$ denotes the normal distribution of mean $\mu$ and variance $\sigma^2$; the probability density of $X \sim N(\mu, \sigma^2)$ is $p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Also if $G_1 \sim N(\mu_1, \sigma_1^2)$ and $G_2 \sim N(\mu_2, \sigma_2^2)$ are two independent random variables then $a_1 G_1 + a_2 G_2 \sim N(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$.

**Lemma 12.7** If $e_i$ is the $i$th canonical basis vector and $G \sim N(\mathcal{V}_t)$ then $\langle G, e_i \rangle \sim N(0, \sigma^2)$, $\sigma^2 \leq 1$.

**Proof:** If $G = \sum_{j=1}^d G_j u_j$ then for every $i$, $\langle G, e_i \rangle = \sum_j G_j \langle u_j, e_i \rangle$ which has mean zero. According to Remark 12.6 $\langle G, e_i \rangle$ is a Gaussian of variance
\[
\text{Var}[\langle G, e_i \rangle] = \sum_j \langle u_j, e_i \rangle^2 \leq \|e_i\|^2 = 1
\]
since the $u_j$ are orthonormal. \qed

**Lemma 12.8** If $\sigma_i^2 = \text{Var}[\langle G, e_i \rangle]$ then $E[\|G\|^2] = \sum_{i=1}^n \sigma_i^2 = \text{dim}(\mathcal{V}_t)$.

**Proof:** Let $d = \text{dim}(\mathcal{V}_t)$:
\[
\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n E[\langle G, e_i \rangle^2] = E[\|G\|^2] = \sum_{i=1}^d \text{Var}(G_i) = d = \text{dim}(\mathcal{V}_t).
\]
\qed
A Martingale Bound.

Lemma 12.9 Let $X_1, X_2, \ldots X_T$ be (correlated) real random variables and let $Y_1, Y_2, \ldots Y_T$ be defined as $Y_i = f_i(X_i)$, where each $f_i$ is a deterministic function. If for all $\xi_1, \xi_2, \ldots, \xi_{i-1}$ it is true that $Y_i | (X_1 = \xi_1, \ldots, X_{i-1} = \xi_{i-1}) \sim N(0, \sigma_i^2(\xi_1, \ldots, \xi_{i-1}))$ where $\sigma_i^2(\xi_1, \ldots, \xi_{i-1}) \leq 1$ then for every $\lambda > 0$,

$$\mathbb{P} \left[ \sum_{i=1}^{T} Y_i \geq \lambda \sqrt{T} \right] \leq e^{-\lambda^2 / 2}.$$

Proof: Let us denote $\sigma_i(\xi_1, \ldots, \xi_{i-1})$ simply by $\sigma_i$. Let us consider a positive parameter $a > 0$.

We have

$$\mathbb{E}[e^{aY_i} | X_1 = \xi_1, \ldots, X_{i-1} = \xi_{i-1}] = \frac{1}{\sqrt{2\pi} \sigma_i} \int_{-\infty}^{\infty} e^{ay} e^{-\frac{y^2}{2\sigma_i^2}} dy = e^{a^2\sigma_i^2 / 2} \leq e^{a^2 / 2}$$

which gives by induction

$$\mathbb{E}[e^{a \sum_{i=1}^{T} Y_i}] \leq e^{Ta^2 / 2}.$$

Then Markov’s inequality gives:

$$\mathbb{P} \left[ \sum_{i=1}^{T} Y_i > \lambda \sqrt{T} \right] = \mathbb{P} \left[ e^{a \sum_{i=1}^{T} Y_i} > e^{a \lambda \sqrt{T}} \right] \leq \frac{\mathbb{E}[e^{a \sum_{i=1}^{T} Y_i}]}{e^{a \lambda \sqrt{T}}} \leq \frac{e^{Ta^2 / 2}}{e^{a \lambda \sqrt{T}}}.$$

Choosing $a = \frac{\lambda}{\sqrt{T}}$ in the above inequality gives

$$\mathbb{P} \left[ \sum_{i=1}^{T} Y_i > \lambda \sqrt{T} \right] \leq e^{-\lambda^2 / 2}$$

which finishes the proof. \[\square\]

The following claims will prepare the ground to prove Theorem 12.1.

Claim 12.10 For all $t$ it is true that $C_t^{\text{disc}} \subset C_{t+1}^{\text{disc}}$ and $C_t^{\text{var}} \subset C_{t+1}^{\text{var}}$. This means that $\text{dim } \mathcal{V}_t \leq \text{dim } \mathcal{V}_{t+1}$.

Proof: Obvious. \[\square\]

Claim 12.11 For $\gamma \leq \frac{\delta}{\sqrt{\log(mn / \gamma)}}$, we have that with high probability $X_1, X_2, \ldots X_T \in \mathcal{P}$.

Proof: Conditioned on $X_{t-1}$, we have $U_t \sim N(V_t)$. Recall that if a constraint gets tight within a distance of $\delta$, it is frozen and cannot be violated again. Hence the only way that a constraint of $\mathcal{P}$ could get violated is that in one step $\langle X_t, v_j \rangle$ increases by more than $\delta$, which means that $\langle U_t, v_j \rangle$ is more than $\delta / \gamma$. Since $\langle U_t, v_j \rangle$ is Gaussian with variance 1, we have

$$\mathbb{P} \left[ \langle U_t, v_j \rangle > \frac{\delta}{\gamma} \right] \leq e^{-\delta^2 / 2\gamma^2} \leq \left( \frac{\gamma}{mn} \right)^{c / 2}.$$

For $\text{poly}(m, n, \frac{1}{\gamma})$ many steps and $c > 0$ large enough,

$$\mathbb{P} \left[ \exists t ; X_t \notin \mathcal{P} \right] \leq \frac{1}{\text{poly}(m, n)}.$$

\[\square\]
Claim 12.12 \( \mathbb{E}[|C_T^{\text{disc}}|] \leq \frac{n}{16} \).

Proof: Remember that \( j \in C_T^{\text{disc}} \) if \( \langle X_T - x_0, v_j \rangle > c_j - \delta \geq 0.9c_j \) (considering \( c_j \geq 1 \) and our choice of a small \( \delta \)). Using Lemma 12.9,

\[
\mathbb{P}[j \in C_T^{\text{disc}}] \leq \mathbb{P}[\gamma \sum_{t=1}^{T} \langle U_t, v_j \rangle \geq 0.9c_j] \leq e^{-(0.9c_j)^2/2\gamma^2 T} \leq e^{-c_j^2/16}.
\]

The last inequality holds because we take \( T = \frac{16}{3\gamma^2} \). So \( \mathbb{E}[|C_T^{\text{disc}}|] \leq \sum_{j=1}^{T} e^{-c_j/16} \leq \frac{n}{16}. \]

We complete the analysis in the next lecture.