Stanford University	**Spring 2016**
Math 233: Non-constructive methods in combinatorics	**Instructor: Jan Vondrák**
Lecture date: May 6, 2016	**Scribe: Evita Nestoridi**

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 $l : [n] \to \{-1, +1\}$, 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 $l : [n] \to \{\pm 1\}$ such that for every $i \in [m]$,*

$$|l(A_i)| \leq K\sqrt{n \log_2 \frac{2m}{n}}.$$

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 $\mathbf{v}_j = \frac{1}{|A_j|}\chi_{A_j}$ and think about the following polytope:

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

where $c_j = K\sqrt{\log_2 \frac{2m}{n}}$.

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 $\mathbf{x}_0$ of the polytope. Once we violate one of the conditions $|x_i| \leq 1 - \frac{1}{\delta}$ or $|\langle \mathbf{v}_j, \mathbf{x}\rangle| \leq c_j - \delta$ 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 $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_m \in \mathbb{R}^n$ 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 $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_T$ such that with constant probability, $\mathbf{X}_T \in [-1, 1]^n$ and*

1. *for at least $n/2$ coordinates $|(\mathbf{X}_T)_i| \geq 1 - \delta$*
2. *$\forall j \in [m]$; $|\langle \mathbf{X}_T - \mathbf{x}_0, \boldsymbol{v}_j\rangle| \leq c_j$*

Remark 12.3 *Notice that only $n/2$ of the coordinates will be labeled ± 1 so we will need to iterate.*

The random walk. Let $\gamma > 0$ and $\delta = \Theta(\gamma\sqrt{\log \frac{mn}{\gamma}})$. Define

$$\mathcal{C}_t^{var} = \{i \in [n] : |(\mathbf{X}_{t-1})_i| \geq 1 - \delta\}$$

be the nearly-tight variable constraints and

$$\mathcal{C}_t^{disc} = \{j \in [m] : |\langle \mathbf{v}_j, \mathbf{X}_{t-1} - x_0 \rangle| \geq c_j - \delta\}$$

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*

$$\mathcal{V}_t = \{\mathbf{y} \in \mathbb{R}^n : y_i = 0 \;\; \forall i \in \mathcal{C}_t^{var} \;\; \text{and} \; \langle \mathbf{v}_j, \mathbf{y} \rangle = 0 \; \forall j \in \mathcal{C}_t^{disc}\}.$$

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 $\boldsymbol{u}_1, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d$ then define*

$$N(\mathcal{V}_t) = \text{the distribution of } \mathbf{G} = G_1\mathbf{u}_1 + \ldots + G_d\mathbf{u}_d, \;\; \text{where } G_i \sim N(0,1) \text{ are independent.}$$

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

Remark 12.6 *$N(\mu, \sigma^2)$ denotes the normal distribution of mean μ and variance σ^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 $a_1G_1 + a_2G_2 \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$.*

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

Proof: If $\mathbf{G} = \sum_{j=1}^d G_j \mathbf{u}_j$ then for every i, $\langle \mathbf{G}, \mathbf{e}_i \rangle = \sum_j G_j \langle \mathbf{u}_j, \mathbf{e}_i \rangle$ which has mean zero. According to Remark 12.6, $\langle \mathbf{G}, \mathbf{e}_i \rangle$ is a Gaussian of variance

$$Var[\langle \mathbf{G}, \mathbf{e}_i \rangle] = \sum_j \langle \mathbf{u}_j, \mathbf{e}_i \rangle^2 \leq ||\mathbf{e}_i||^2 = 1$$

since the $\mathbf{u}_j$ are orthonormal. □

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

Proof: Let $d = dim(\mathcal{V}_t)$:

$$\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \mathbb{E}[\langle \mathbf{G}, \mathbf{e}_i \rangle^2] = \mathbb{E}[||\mathbf{G}||^2] = \sum_{i=1}^d Var(G_i) = d = dim(\mathcal{V}_t).$$

□

A Martingale Bound.

Lemma 12.9 *Let $X_1, X_2, \dots X_T$ be (correlated) real random variables and let $Y_1, Y_2, \dots Y_T$ be defined as $Y_i = f_i(X_i)$, where each f_i is a deterministic function. If for all $\xi_1, \xi_2, \dots \xi_i$ it is true that $Y_i | (X_1 = \xi_1, \dots, X_{i-1} = \xi_{i-1}) \sim N(0, \sigma_i^2(\xi_1, \dots, \xi_{i-1}))$ where $\sigma_i^2(\xi_1, \dots, \xi_{i-1}) \leq 1$ 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 $\sigma_i(\xi_1, \dots, \xi_{i-1})$ simply by σ_i. Let us consider a positive parameter $a > 0$. We have

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

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. □

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

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

Proof: Obvious. □

Claim 12.11 *For $\gamma \leq \frac{\delta}{\sqrt{c\log(mn/\gamma)}}$, we have that with high probability $\mathbf{X}_1, \mathbf{X}_2, \dots \mathbf{X}_T \in \mathcal{P}$.*

Proof: Conditioned on $\mathbf{X}_{t-1}$, we have $\mathbf{U}_t \sim N(\mathcal{V}_t)$. Recall that if a constraint gets tight within a distance of δ, 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 $\langle \mathbf{X}_t, \mathbf{v}_j \rangle$ increases by more than δ, which means that $\langle \mathbf{U}_t, \mathbf{v}_j \rangle$ is more than δ/γ. Since $\langle \mathbf{U}_t, \mathbf{v}_j \rangle$ is Gaussian with variance 1, we have

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

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

$$\mathbb{P}[\exists t; \mathbf{X}_t \notin \mathcal{P}] \leq \frac{1}{poly(m,n)}.$$

□

Claim 12.12 $\mathbb{E}[|\mathcal{C}_t^{disc}|] \leq \frac{n}{16}$.

Proof: Remember that $j \in \mathcal{C}_T^{disc}$ if $\langle \mathbf{X}_T - \mathbf{x}_0, \mathbf{v}_j \rangle > c_j - \delta \geq 0.9c_j$ (considering $c_j \geq 1$ and our choice of a small δ). Using Lemma 12.9,

$$\mathbb{P}[j \in \mathcal{C}_T^{disc}] \leq \mathbb{P}[\gamma \sum_{t=1}^{T} \langle \mathbf{U}_t, \mathbf{v}_j \rangle \geq 0.9c_j] \leq e^{-(0.9c_j)^2/2\gamma^2 T} \leq e^{-cj^2/16}.$$

The last inequality holds because we take $T = \frac{16}{3\gamma^2}$. So $\mathbb{E}[|\mathcal{C}_t^{disc}|] \leq \sum_{j=1}^{T} e^{-c_j/16} \leq \frac{n}{16}$. □

We complete the analysis in the next lecture.