Lecture 8

Plan for today

1. Basics about Gaussian r.vs.
2. Dimension reduction using Johnson-Lindenstrauss Flattening Lemma.
3. A recap of martingale concentration
4. Discrepancy Minimization using Random Walks + Martingales + Gaussians.

—— x ——

Facts about Gaussian r.vs

- $X \sim N(\mu, \sigma^2) \Rightarrow cX \sim N(c\mu, c^2\sigma^2)$

- $X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \sum_i X_i \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$.

$\Rightarrow$ if $G_i \sim N(0,1)$ and $\bar{G} = (G_1, G_2 \dots G_n) \in \mathbb{R}^n$ $\quad x \in \mathbb{R}^n$

then $\langle x, G \rangle \sim N(0, \sum_i x_i^2) = N(0, \|x\|^2)$

So if $\|x\| = 1 \Rightarrow \langle x, G \rangle \sim N(0,1)$.

Similarly:

Suppose V is a subspace of $\mathbb{R}^n$ of dim $d \leq n$.

Pick a basis $\bar{v}_1, \bar{v}_2 \dots \bar{v}_d$ for V

and define $g = g_i \bar{v}_i$ where $g_i \sim N(0,1)$ indep

then we say g is a gaussian from subspace V. $\quad g \sim \boxed{N(V)}$.

Fact: for vector $x \in \mathbb{R}^n$ and $G \sim N(V)$

$\langle x, G \rangle \sim N(0, \sigma^2)$ for some $\sigma^2 \leq 1$.

Recall our bound from Azuma-Hoeffding

Let $Z_1, Z_2 \ldots\ldots$ be independent r.v.s.

and X_i = function of $Z_1 Z_2 \ldots Z_i$

and say that $X_i \mid Z_1 \ldots Z_{i-1}$ behaves like a Rademacher.

i.e. $Pr(X_i = 1 \mid Z_1 \ldots Z_{i-1}) = Pr(X_i = -1 \mid Z_1 \ldots Z_{i-1})$
$= 1/2$.

the Azuma-Hoeffding says:-

$$\boxed{Pr\left(\left| \sum_{i=1}^{T} X_i \right| \geq \lambda \right) \leq 2 \exp\left(-\lambda^2/2T\right)}$$

This extends to gaussian RVs:

sps. $X_i \mid Z_1 \ldots Z_{i-1}$ is gaussian. with mean 0 and var 1.

$$\Rightarrow Pr\left(\left| \sum_{i=1}^{T} X_i \right| \geq \lambda \right) \leq 2 \exp\left(-\lambda^2/2T\right)$$

And if $X_i \mid Z_1 \ldots Z_{i-1} \sim N(0, \sigma_i^2)$ for $\sigma_i^2 \leq \varepsilon$ then

$$Pr\left(\left| \sum_{i=1}^{T} X_i \right| \geq \lambda \right) \leq 2 \exp\left(-(\lambda/\varepsilon)^2/2T\right).$$

(Gaussian concentration for Martingales)

—— x ——

Crucially: also holds when T is a "stopping" time (not a fixed quantity).

Def: T is stopping time wrt martingale $X_1, X_2, \ldots$

if $\{T = t\}$ depends only on $X_1 \ldots X_t$.

Discrepancy:

Given a setsystem $\mathcal{S} = (S_1, S_2 \ldots S_m)$ where each $S_i \subseteq [n]$ $[n] = \{1, 2 \ldots n\}$

A 2-coloring of $[n]$ is a map $\chi : [n] \to \{-1, 1\}$

the discrepancy of this coloring is

$$\max_{i \in [m]} \left| \sum_{j \in S_i} \chi(j) \right|$$

Want small discrepancy (good balance).

Fact: Sps set $\chi(j) \in \{-1, 1\}$ uniformly at random.

then $\mathbb{E}\left[\sum_{j \in S_i} \chi(j)\right] = 0$ and $\Pr\left(\left|\sum_{j \in S_i} \chi(j)\right| \geq \lambda\right) \leq \exp\left(-\frac{2\lambda^2}{n}\right)$

so set $\lambda = O(\sqrt{n \log m})$ to get that discrepancy $\leq \lambda$ $\forall$ sets $\in \mathcal{S}$ wp. $1/\mathrm{poly}(m)$.

Thm [Joel Spencer]

$\exists$ a coloring with $\mathrm{disc}_{\mathcal{S}}(\chi) \leq O\left(\sqrt{n \log(m/n)}\right)$

$\Rightarrow$ for $m = n$ get $O(\sqrt{n})$ discrepancy.

even $\leq 6\sqrt{n}$

"six standard deviations suffice".

Original proof was non-constructive (entropy + pigeonhole principle on exponentially large family)

- Bansal (2010) gave first algorithmic result (semidef. prog + rounding)
- today's proof due to Lovett + Meka. (uses LPs, Gaussians, Martingales).

Step 1: $\boxed{\text{"Relaxation" to Partial Colorings}}$

"convex" fractional relaxation

Instead of $\chi: [n] \to \{-1,1\}$ allow $\chi: [n] \to [-1,1]$.

Bad news: set $\chi(j) = 0 \ \forall j \Rightarrow \sum_{j \in S_i} \chi(j) = 0$

"zero fractional discrep".

Fix: allow only small fraction of vars to be "far from" $\{-1,1\}$.

Most variables should be "close to" ± 1.

Lemma 1: Let $x_0 = 0^n$ be starting point in $[-1,1]^n$.

Then can find $x \in [-1,1]^n$ s.t.

① $|x(S_i)| = \left|\sum_{j \in S_i} x_j\right| \le \sqrt{|S_i|} \cdot \Delta + \frac{1}{poly(n)}$

$\Delta = c \sqrt{\log(m/n)}$ for constant c

② $\#(j \text{ s.t. } x_j \notin \{-1,1\}) \le n/2$.

Actually prove the following:—

"Lemma 2"

given any vectors $a_1, a_2 \ldots a_m \in \mathbb{R}^n$ and any starting point $\bar{x}_0 \in [-1,1]^n$

And $\delta > 0$ small $1/poly(n)$

can find $x \in [-1,1]^n$ s.t.

① $|\langle a_i, x - x_0 \rangle| \le \|a_i\|_2 \cdot \Delta$

$\Delta = c\sqrt{\log(m/n)}$. for constant c

② $\#\{j \text{ st } x_j \in [-(1-\delta), 1-\delta]\} \le n/2$.

Set a_i = Indicator vector of set $S_i \Rightarrow \|a_i\|_2 = \sqrt{|S_i|}$, set $\bar{x}_0 = \bar{0}$.

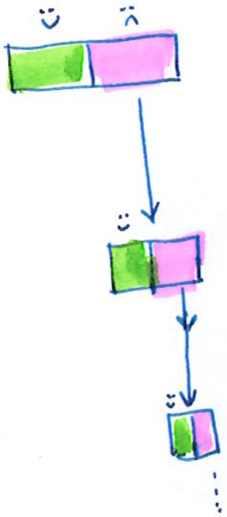
Now take solution given by Lemma 2, "round" vars close to ± 1, to get Lemma 1.

Great: before we prove Lemma 2 (and hence Lemma 1), Let's see how to get total coloring from partial coloring.

Start at $X_0 = 0$.

Lemma 1 $\Rightarrow$ set J of $\geq n/2$ vars at $\{\pm 1\}$

rest of vars ($\leq n/2$ vars) fractional, (values $x_j \in (-1,1)$)



Freeze these vars in J. on $[n] \setminus J$, start at where previous run stopped (use it as the X_0) and run Lemma 1 again on these $n/2$ vars.

$\Rightarrow$ J' of $\geq \frac{1}{2}([n]\setminus J)$ vars at $\{\pm 1\}$.

rest again fractional ..... repeat.

Net discrepancy may add up, but its at most

$$\chi(S_i) \leq c\sqrt{n \log m/n} + c\sqrt{\tfrac{n}{2} \log (m/(n/2))} + \cdots\cdots$$

$$= c \cdot \sum_{i \geq 0} \sqrt{\frac{n}{2^i} \log\left(\frac{2^i m}{n}\right)} = O\left(\sqrt{n \log m/n}\right)$$

first term dominates.

$\Rightarrow$ get Spencer's Theorem from Lemma 1 (and hence from Lemma 2).

How to prove Lemm 2?

Beautiful algo — uses random walks

Gaussians

Martingales.

Recall: Lemma 2

Given vectors $a_1 \dots a_m$, any start pt $x_0 \in [-1,1]^n$, any δ small

find $x \in [-1,1]^n$ st

$\Delta = c\sqrt{\lg(m/n)}$

$\boxed{① \ |\langle a_i, x - x_0\rangle| \leq \|a_i\|_2 \cdot \Delta}$

② $\#(j \text{ st } x_j \in [-(1-\delta), 1-\delta]) \leq \cancel{n/2} \cdot \frac{3n}{4}$ (same idea works)

Wlog: $\boxed{a_i \text{ unit vectors}}$ so that want $\boxed{|\langle a_i, x - x_0\rangle| \leq \Delta}$.

Algo: start at x_0

take tiny Gaussian steps. (so $x_t \leftarrow x_{t-1} + \text{gaussian}$)

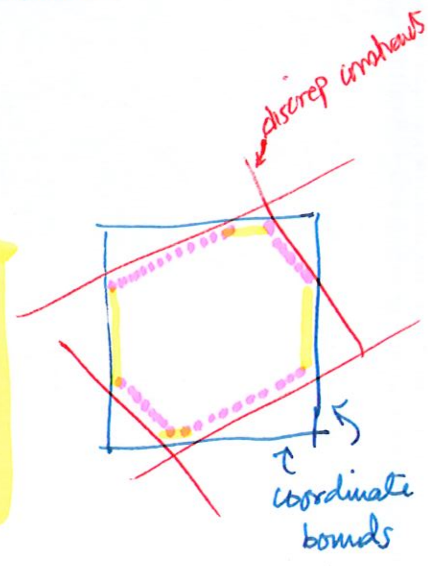
[If we just do this, not great.

But we want $x_j \in [-1,1]$ in each coord j

and want $\langle a_i, x - x_0\rangle \in [-\Delta, \Delta]$

So idea: if we get close to violating some constraint

freeze the solution to lie in that subspace.



Call:

variable j frozen if $|x_j| > 1-\delta$.

constraint a_i dangerous if $|\langle a_i, x - x^0\rangle| > \Delta - \delta$

Intuition: the variable bounds are much closer to origin than discrepancy bounds

so will hit a variable bound earlier. So freeze many more constraints.

At time t, say have solution x^t.

- Some variables ~~are~~ j have value $|x_j^t| \geq 1-\delta$. Call them Frozen.

  Some constraints i have $|\langle a_i, x^t - x^0\rangle| \geq \Delta - \delta$. Call them Dangerous

  So $F^t = \{j : |x_j^t| \geq 1-\delta\}$

  $D^t = \{i : |\langle a_i, x^t - x^0\rangle| \geq \Delta - \delta\}$.

- Let V_t = subspace orthogonal to all frozen vars and all dangerous constraints.

  $= \left(\text{span}(\{e_j : j \in F\} \cup \{a_i : i \in D\})\right)^\perp$

- As long as V_t has dimension $\geq n/2$ — (so: few frozen / dangerous vars / consts)

Algo

$$\boxed{\begin{array}{l} \text{Pick } g_t \sim N(V_t) \quad \text{gaussian from that subspace} \\ x^{t+1} \leftarrow x^t + \varepsilon \cdot g_t \end{array}}$$

$\varepsilon \leq \dfrac{\delta}{\sqrt{\log \frac{mn}{\delta}}}$

Stop when $\dim(V_t) < n/2$

$\Rightarrow$ many frozen / dangerous things (at least $n/2$)

Show that: most of them are frozen. So lots of vars (say $\geq n/4$ in expectation) are close to ± 1.

Proves what we want.

What Could go wrong with Algo?

① Could be that x^t jumps outside $[-1,1]^n \cap$ $\{x \mid |\langle a_i, x\rangle| \leq \Delta\}$

② Or that when # of frozent dangerous constraints exceed $n/2$

(or actually, when $\dim(V_t) < n/2$)

then very few frozen vars (mostly dangerous constraints)

---

① Know that x^{t-1} was good (for non frozen vars, $|x_j^{t-1}| \leq 1-\delta$

for non dangerous constraints, $|\langle a_i, x^{t-1}\rangle| \leq \Delta - \delta$)

So for x^t to go outside the feasible region, must have

that $|\varepsilon . g^t| \geq \delta$

$$\Pr(|\varepsilon g^t| \geq \delta) \leq 2\exp\left(-\left(\tfrac{\delta}{\varepsilon}\right)^2/2\right)$$

$$\text{So if } \delta \geq \varepsilon . \sqrt{\log\left(\tfrac{mn}{\varepsilon}\right)} \Rightarrow \quad \leq \left(\varepsilon/mn\right)^{O(1)}.$$

can now take union bound
over all steps
as long as steps $\leq \operatorname{poly}\left(\frac{mn}{\varepsilon}\right)$

② How many steps?

Fact: $\mathbb{E}[\|x^{T+1} - x_0\|^2] = \sum_{t \leq T} \varepsilon^2 . \dim(V_t)$

Pf: $\mathbb{E}[\|x^{T+1} - x^0\|^2] = \mathbb{E}[\|x^T - x^0\|^2] + 2\mathbb{E}[\cancel{\langle \varepsilon g^T, x^T - x^0\rangle}] + \mathbb{E}[\|\varepsilon g^T\|^2]$

$= \mathbb{E}[\|x^T - x^0\|^2] + \varepsilon . \dim(V_T)$. ∎

0 by independence and g^T is symmetric

$\| \; \varepsilon \dim(V_T)$ under reflections

So as long as $\dim(V_t) \geq n/2$,

$$E[\|x^{T+1} - x^0\|^2] \geq E[\|x^T - x^0\|^2] + n/2 \cdot \varepsilon^2$$

$$\geq n/2 \cdot \varepsilon^2 \cdot T$$

But since x^T must remain in $[-1,1]^n$, $\quad \|x^T - x^0\|^2 \leq 4n$

$$\Rightarrow T \leq 8/\varepsilon^2 .$$

$\Rightarrow$ Algo must stop after $O(1/\varepsilon^2)$ steps.

---

want this to be $\geq n/4$ wp $1/2$ (say)

OK: What happens at this stopping time: how many frozen / dangerous?

want this to be small.

Let's see what $\Pr(\text{constraint } i \text{ dangerous})$ is?

where $v = e_j$ for a "variable" constraint or a_i for a "set" constraint

$$\Pr\left(\langle v, \varepsilon(g_1 + \cdots + g_t)\rangle \geq \Delta - \delta\right)$$

(Martingale concentration for Gaussians)

$$\leq \Pr\left(\langle v, \varepsilon(g_1 + \cdots + g_t)\rangle \geq \Delta/2\right)$$

but each $\langle v, \varepsilon g_i\rangle \sim N(0, \varepsilon^2 \sigma_i^2)$ for some $\sigma_i^2 \leq 1$.

$$\leq 2\exp\left(-\frac{(\Delta/2)^2/\varepsilon^2}{T}\right) \quad \text{but } T = O(1/\varepsilon^2)$$

and $\Delta = c\sqrt{\log(m/n)}$.

$$\leq \frac{n}{10m} \quad \text{for large enough } c.$$

$\Rightarrow E[\text{dangerous constraints}] \leq n/10$.

$\Rightarrow$ $\boxed{\text{with prob} \geq 1/2,\ \#\text{dangerous} \leq n/5 \quad \Rightarrow \#\text{frozen} \geq n/2 - n/5 \geq n/4.}$ ☺

To wrap up:

Today we saw Gaussian r.v.s and used them for

(a) Dimension reduction (for distance preservation)

(b) Discrepancy minimization.

Along the way we need

- concentration bounds for sums of $\underbrace{\text{squares of Gaussians}}_{\text{Chi-squared}}$ (indep)

- concentration bounds for Gaussians (but now martingales)

- random walks using Gaussians.

These ideas are simple but very! powerful

- the dimension reduction technique (argument)

  is useful in another surprising context

  - compressive sensing (aka the "single-pixel camera").

(See notes or wikipedia)