Lecture 5

Matrix completion

Multi spike model with missing data.

<u>parameters</u>: rank $r \geq 2$, dimension $d \geq 2$, sample size $n \geq 2$, variance σ^2

<u>unknown</u>: $\Theta^0 \in \mathbb{R}^{d \times d}$, $\text{rank}(\Theta^0) \leq r$

<u>observe</u>: few missing entries of Θ^0

$$(\underline{X}_1, y_1), \cdots, (\underline{X}_n, y_n)$$

$$y_i = \langle \underline{X}_i, \Theta^0 \rangle + \underline{w}_i, \quad \underline{X}_i = e_{\underline{a}(i)} e_{\underline{b}(i)}^T$$

$$\underline{w}_1, \cdots, \underline{w}_n \overset{iid}{\sim} N(0, \sigma^2), \quad \underline{a}(1), \cdots, \underline{a}(n), \underline{b}(1), \cdots, \underline{b}(n) \sim \text{Unif}(\{1, \cdots, d\})$$

<u>error measure</u>: $\frac{1}{d^2} \| \hat{\Theta} - \Theta^0 \|_F^2$

<u>necessary</u>: $n \gg r \cdot d$, # observation $\gg$ # dof

$n \geq d \cdot \log d$, all rows & columns covered

<u>additional obstruction</u> $r = 1$, $\sigma = 0$

Suppose $\underline{\Theta}^0 \in_R \{\pm M \cdot e_1 e_1^T\}$

$$\mathbb{P}(\text{loc } 1,1 \text{ not observed}) = \left(1 - \frac{1}{d^2}\right)^n \geq 1 - \frac{n}{d^2} \longrightarrow 1$$

$$\mathbb{E}\left[\frac{1}{d^2} \| \hat{\underline{\Theta}} - \Theta^0 \|_F^2 \mid \text{loc } 1,1 \text{ not observed}\right] \geq \frac{M^2}{d^2}$$

$\longrightarrow$ will allow error to grow with $\|\Theta^0\|_{max}$

> <u>Theorem</u> $\exists$ poly-time estimator $\hat{\underline{\Theta}}$, if $n \geq d \cdot \log d$, then w.h.p.
>
> $$\frac{1}{d^2} \| \hat{\underline{\Theta}} - \Theta^0 \|_F^2 \lesssim \frac{(\sigma^2 + \|\Theta^0\|_{max}^2) \cdot r \cdot d \cdot \log(d)}{n}$$

<u>Strategy for poly-time estimator</u>

two steps

$$Y = \frac{1}{n} \sum_{i=1}^n y_i \cdot d^2 \cdot X_i$$

① Rough estimation (simple function of observation) $\mathbb{E}[Y] = \frac{1}{d^2} \sum_{a,b=1}^d \Theta_{ab}^0 \cdot d^2 \cdot e_a e_b^T = \Theta^0$

② Boost error bound (reduce to our multi-spike matrix estimator / analysis)

Compute $\underline{Y}$ as function of $(\underline{X}_1, \underline{y}_1), \cdots, (\underline{X}_n, \underline{y}_n)$ ensure $\mathbb{E}[\underline{Y}] = \Theta^0$

– Compute: $\underline{Y} = \frac{1}{n}(\underline{y}_1 \cdot \underline{X}_1 \cdot d^2 + \underline{y}_2 \underline{X}_2 \cdot d^2 + \cdots + \underline{y}_n \cdot \underline{X}_n \cdot d^2)$

$\underline{Y} = \Theta^0 + \underline{W}$, $\mathbb{E}[\underline{W}] = 0$

– output: $\hat{\underline{\Theta}} = \text{argmin}\{\| \Theta - \underline{Y} \|_F^2 \mid \text{rank}(\Theta) \leq r\}$

* A noisy version of Θ^0

<u>error analysis</u>

let $\underline{U} = \hat{\underline{\Theta}} - \Theta^0$, $\underline{W} = \underline{Y} - \Theta^0$

<u>claim</u>: $\| \underline{U} \|_F^2 \leq 8r \cdot \| \underline{W} \|^2$

$$\underline{W} = \underline{Y} - \Theta^0 = \frac{1}{n} \sum_{i=1}^n (\underline{y}_i d^2 \underline{X}_i - \Theta^0)$$

* $\underline{y}_i = \langle \underline{X}_i, \Theta^0 \rangle + \underline{w}_i$, $\underline{w}_i \sim N(0, \sigma^2)$

$$\underbrace{(\langle \underline{X}_i, \Theta^0 \rangle d^2 \underline{X}_i)}_{\underline{Z}_i^{(1)}} + \underbrace{\underline{w}_i \cdot d^2 \underline{X}_i}_{\underline{Z}_i^{(2)}}$$

$$= \frac{1}{n} \sum_{i=1}^n \underline{Z}_i^{(1)} + \frac{1}{n} \sum_{i=1}^n \underline{Z}_i^{(2)}$$

claim if $n \geq d \cdot \log d$, w.h.p.

1. $\left\| \frac{1}{n} \sum_{i=1}^{n} Z_i^{(1)} \right\|^2 \lesssim \|\theta^\circ\|_{max} \cdot \frac{d^3 \log d}{n}$
2. $\left\| \frac{1}{n} \sum_{i=1}^{n} Z_i^{(2)} \right\|^2 \lesssim \sigma^2 \frac{d^3 \log d}{n}$

assuming the claim

$$\frac{1}{d^2}\|U\|_F^2 \leq \frac{8r}{d^2}\|W\|^2$$

$$\overset{\text{claim}}{\lesssim} \frac{r}{d^2} \cdot \left(\|\theta\|_{max}^2 + \sigma^2\right) \frac{d^3 \cdot \log d}{n}$$

Matrix Bernstein bound

Z random $d \times d$ matrix, $d \geq 2$

$\mathbb{E}[Z] = 0$, $\{Z\} = \{Z^T\}$

(σ, b)-Bernstein condition

$\forall j \in \mathbb{N}_{\geq 2}$, $\| \mathbb{E}[\|Z\|^{j-2}] Z Z^T \| \leq j! \cdot b^{j-2} \cdot \sigma^2$

$$\boxed{\begin{array}{l} \text{Theorem If } Z \text{ satisfies } (\sigma, b)\text{-Bernstein condition, then with prob.} \geq 1 - d^{-100} \\ \left\| \frac{1}{n} \sum_i Z_i \right\| \lesssim \sigma \sqrt{\frac{\log d}{n}} + b \frac{\log d}{n} \end{array}}$$

Loewner order A, B symmetric matrices

$A \preceq B \iff B - A$ is p.s.d.

$\iff \forall v, \langle v, Av \rangle \leq \langle v, Bv \rangle$

fact if $-A \leq B \leq A$, then $\|B\| \leq \|A\|$

$Z^{(1)} = \langle X_i, \theta^\circ \rangle \cdot d^2 \cdot X - \theta^\circ$, $X = e_a \cdot e_b^T$

claim $Z^{(1)}$ is (σ_1, b_1)-Bernstein for $\sigma_1 \lesssim \|\theta^\circ\|_{max} \cdot d^{3/2}$ and $b_1 \lesssim \|\theta^\circ\|_{max} \cdot d^2$

Proof

$$\|Z^{(1)}\| \leq \underbrace{\|\langle X, \theta^\circ \rangle \cdot d^2 \cdot X\|}_{\underbrace{|\langle X, \theta^\circ \rangle| \cdot d^2}_{\|\theta^\circ\|_{max} \cdot d^2}} + \underbrace{\|\theta^\circ\|}_{\leq \|\theta^\circ\|_{max} \cdot d^2}$$

$$\leq 2\|\theta^\circ\|_{max} \cdot d^2$$

p.s.d.

$$\underbrace{0 \leq \mathbb{E}[Z^{(1)}(Z^{(1)})^T]}_{\text{p.s.d.}} = \mathbb{E}[\langle X, \theta^\circ \rangle^2 \cdot d^4 X X^T + \theta^\circ(\theta^\circ)^T - 2\theta^\circ(\theta^\circ)^T]$$

$* \ \mathbb{E}[(X - \mathbb{E}[X])]^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$= \mathbb{E}[\langle X, \theta^\circ \rangle^2 \cdot d^4 X X^T - \theta^\circ(\theta^\circ)^T]$$

$$= \mathbb{E}[(\theta^\circ_{a,b})^2 \cdot d^4 \cdot e_a e_a^T]$$

$e_a e_b^T (e_a e_b^T)^T = e_a \overset{1}{\cancel{e_b^T e_b}} e_a^T = e_a e_a^T$

$$\preceq \|\theta^\circ\|_{max}^2 \cdot d^4 \cdot \mathbb{E}[e_a \cdot e_a^T] = \|\theta^\circ\|_{max}^2 d^3 \cdot I_d$$

$\|\mathbb{E}[Z Z^T]\| \leq \|\theta^\circ\|_{max}^2 \cdot d^3$

$$\left\| \mathbb{E}\left[\|\mathbf{Z}^{(1)}\|^{j-2} \cdot \mathbf{Z}^{(1)}(\mathbf{Z}^{(1)})^T\right]\right\| \le \left(2\|\theta^0\|_{max}\cdot d^2\right)^{j-2} \cdot \underbrace{\underbrace{\left\|\mathbb{E}\left[\mathbf{Z}^{(1)}(\mathbf{Z}^{(1)})^T\right]\right\|}_{\|\theta^0\|_{max}^2 \cdot d^3}}_{\sigma_1^2}$$

Theorem

W.h.p., $\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i\right\| \le \sigma\sqrt{\frac{\log d}{n}} + b\,\frac{\log d}{n}$

$\lesssim \|\theta^0\|_{max}\cdot\left(\frac{d^{3/2}\sqrt{\log d}}{\sqrt{n}} + \frac{d^2\log d}{n}\right)$

$\mathbf{Z}^{(2)} = \mathbf{W}\cdot d^2 \cdot \mathbf{X}$, $\mathbf{W} \sim N(0,\sigma^2)$

claim $\mathbf{Z}^{(2)}$ is (σ_2, b_2)-Bernstein, for $\sigma_2 =$, $b_2 =$

Proof $\|\mathbf{Z}^{(2)}\| = |\mathbf{W}|\cdot d^2$

$$\mathbb{E}\Big[\underbrace{\|\mathbf{Z}^{(2)}\|^{j-2}\cdot \mathbf{Z}^{(2)}(\mathbf{Z}^{(2)})^T}_{|\mathbf{W}|^j\cdot(d^2)^j\cdot e_{\mathbf{a}} e_{\mathbf{a}}^T}\Big] = \underbrace{\mathbb{E}\left[|\mathbf{W}|^j\right]}_{(j-1)(j-3)\cdots\cdots 1 \le j!\,\sigma^j} \cdot d^{2j}\cdot \frac{1}{d} I_d \preceq j!\,(\underbrace{\sigma d^2}_{b_2})^{j-2}\cdot \underbrace{\sigma^2\cdot d^3}_{\sigma_2^2}$$