Lecture 11: Singular Value Thresholding

Isabella Zhu

13 March 2025

§1 Matrix Denoising Problem

Observe $Y = \Theta^* + E$, all $m \times n$ matrices. We will assume that

$$u^T E v \sim \mathrm{subG}(\sigma^2 |u|_2^2 |v|_2^2)$$

which means E is a subGaussian matrix. Assume that $\mathrm{rank}(\Theta^*) = r < \min(m, n)$. Goal is to estimate Θ^*. Error is Frobenius norm.

§2 Singular Value Thresholding

The SVD of Θ^* is

$$\Theta^* = \sum_j \lambda_j u_j v_j^T$$

The SVD of Y is

$$Y = \sum_j \hat{\lambda}_j \hat{u}_j \hat{v}_j^T$$

Definition 2.1. The **SVT estimator** is

$$\hat{\Theta}^{SVT} = \sum_j 1(\hat{\lambda}_j \geq 2\tau) \hat{\lambda}_j \hat{u}_j \hat{v}_j^T$$

We want to pick τ so that $\lambda_j(E) \leq \tau$ for all j with probability $1 - \delta$. Note that $\max \lambda_j \lesssim \sigma\sqrt{m+n}$ (see lecture 5).

Lemma 2.2

We have

$$||E||_p \leq 2\sigma\sqrt{5m + 5n} + 2\sigma\sqrt{2\log(1/\delta)} \triangleq \tau$$

with probability at least $1 - \delta$.

Proof.

$$||E||_{op} = \max_{x \in B_2} |Ex|_2$$

Define N_m to be a $\frac{1}{4}$-net for B_2^m and similarly for N_n. We have $|N_m| \leq 9^m$ and

$|N_n| \leq 9^n$. Then, we have

$$||E||_{op} \leq \max_{z \in N_n} |Ez|_2 + \frac{1}{4}||E||_{op}$$

and then we have

$$\max_{z \in N_n} |Ez|_2 = \max y^T Ez \leq \max_{w \in N_m} w^T Ez + \frac{1}{4}||E||_{op}$$

Putting this together, we have

$$||E||_{op} \leq 2 \max_{x \in N_n, y \in N_m} y^T Ex$$

Using our subgaussian bounds, we get

$$\mathbb{P}(||E||_{op} > t) \leq 9^n \cdot 9^m \cdot \mathbb{P}(w^T Ez > \frac{t}{2}) \leq 9^{n+m} e^{-t^2/8\sigma^2} \triangleq \delta$$

and then we can just find t in front of δ and we're done.

Theorem 2.3

Define $\tau = 2\sigma\sqrt{5m + 5n} + 2\sigma\sqrt{2\log(1/\delta)}$. Then,

$$\frac{1}{mn}||\hat{\Theta}^{SVT} - \Theta^*||_F^2 \lesssim \frac{\sigma^2 \text{rank}(\Theta^*)}{mn}\left(m + n + \log\frac{1}{\delta}\right)$$

Proof. Define the support $S = \{j : |\hat{\lambda}_j| > 2\tau\}$. Define the event $A = \{||E||_{op} \leq \tau\}$. We already know $\mathbb{P}(A) \geq 1 - \delta$.

On A, we have from Weyl's inequality, $|\hat{\lambda}_j - \lambda_j| \leq ||E||_{op} \leq \tau$. For $j \in S$, we have that

$$|\hat{\lambda}_j| > 2\tau \implies |\lambda_j| \geq |\hat{\lambda}_j| - |\hat{\lambda}_j - \lambda_j| \geq \tau.$$

Similarly, we have

$$|\hat{\lambda}_j| \leq 2\tau \implies |\lambda_j| \leq |\hat{\lambda}_j| + |\hat{\lambda}_j - \lambda_j| \leq 3\tau$$

Define the **oracle**

$$\bar{\Theta} = \sum_{j \in s} \lambda_j u_j v_j^T$$

Then, we have

$$||\hat{\Theta} - \bar{\Theta}||_F^2 \leq ||\hat{\Theta} - \bar{\Theta}||_{op}^2 \cdot \text{rank}(\hat{\Theta} - \bar{\Theta}) \leq 2|S| \cdot ||\hat{\Theta} - \bar{\Theta}||_{op}^2$$

So we get that

$$||\hat{\Theta} - \bar{\Theta}||_{op} \leq ||\hat{\Theta} - Y||_{op} + ||Y - \Theta^*||_{op} + ||\bar{\Theta} - \Theta^*||_{op}$$

We have $||\hat{\Theta} - Y||_{op} \leq 2\tau$ by definition. We also have $||Y - \Theta^*||_{op} = ||E||_{op} \leq \tau$ by definition. Finally, $||\bar{\Theta} - \Theta^*||_{op} \leq 3\tau$. Therefore,

$$||\hat{\Theta} - \bar{\Theta}||_{op}^2 \leq 6\tau \implies ||\hat{\Theta} - \bar{\Theta}||_F^2 \leq 72 \cdot |S| \cdot \tau^2$$

Then we bound

$$\begin{aligned} ||\hat{\Theta} - \Theta^*||_F^2 &\leq 2||\hat{\Theta} - \bar{\Theta}||_F^2 + 2||\bar{\Theta} - \Theta^*||_F^2 \\ &\leq 144 \sum_{j \in S} \tau^2 + 2 \sum_{j \in S^c} \lambda_j^2 \\ &\leq 144 \sum_j \min(\tau^2, \lambda_j^2) \\ &\leq 144 \text{rank}(\Theta^*) \tau^2 \end{aligned}$$

Remark 2.4. Note that for inequalities, we assumed A was symmetric. If A is not symmetric, define matrix

$$\tilde{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

and do the same inequalities, since singular values are the same. So, Weyl works for non-symmetric matrices as well.

§3 Spectral Properties

Theorem 3.1

(David-Kahan) Let $A, \hat{A}$ be symmetric $n \times n$ matrices such that we have SVDs $A = \sum_{j=1}^n \lambda_j u_j u_j^T$ and $\hat{A} = \sum_{j=1}^n \hat{\lambda}_j \hat{u}_j \hat{u}_j^T$, ordered largest to smallest. Then,

$$|\sin \angle(u_1, \hat{u_1})| \leq \frac{2||\hat{A} - A||_{op}}{\max(\lambda_1 - \lambda_2, \hat{\lambda_1} - \hat{\lambda_2})}$$

Remark 3.2. We can get direct bound on distances $\min_{\epsilon=\pm 1} |u_1 - \epsilon \hat{u_1}|_2 \leq \sqrt{2} |\sin \angle(u_1, \hat{u_1})|$.

Proof. We'll alias $u_1 \to u$ and $\hat{u_1} \to \hat{u}$. Take any x such that $|x|_2 = 1$. We can write

$$x = \sum_{j=1}^n \langle x, u_j \rangle u_j, \quad \sum_{j=1}^n \langle x, u_j \rangle^2 = 1.$$

Then we have

$$x^T A x = \sum_{j=1}^n \lambda_j \langle x, u_j \rangle^2 = \lambda_1 \langle x, u \rangle^2 + \sum_{j \geq 2} \lambda_j \langle x, u_j \rangle^2 \leq \lambda_1 \langle x, u \rangle^2 + \lambda_2 \sum_{j \geq 2} \lambda_j \langle x, u_j \rangle^2$$

which can be rewritten as

$$x^T A x \leq (\lambda_1 - \lambda_2) \langle x, u \rangle^2 + \lambda_2 = (\lambda_1 - \lambda_2)(1 - \sin^2 \angle(x, u)) + \lambda_2$$

Substituting $u^T A u = \lambda_1$, we get

$$(\lambda_1 - \lambda_2) \sin^2 \angle(x, u) \leq u^T A u - x^T A x$$

Take $x = \hat{u}$. So, we get

$$\begin{aligned}
(\lambda_1 - \lambda_2) \sin^2 \angle(\hat{u}, u) &\leq u^T A u - \hat{u}^T A \hat{u} \\
&= u^T \hat{A} u - \hat{u}^T A \hat{u} + u^T (A - \hat{A}) u \\
&\leq \hat{u}^T \hat{A} \hat{u} - \hat{u}^T A \hat{u} + u^T (A - \hat{A}) u \\
&= \langle A - \hat{A}, uu^T - \hat{u}\hat{u}^T \rangle \\
&\leq ||\hat{A} - A||_{op} \cdot ||uu^T - \hat{u}\hat{u}^T||_1 \ \text{ (Holder)} \\
&\leq ||\hat{A} - A||_{op} \cdot \sqrt{2} ||uu^T - \hat{u}\hat{u}^T||_F \ \text{ (Cauchy-Scharwz)}
\end{aligned}$$

Now we note that

$$||uu^T - \hat{u}\hat{u}^T||_F^2 = ||uu^T||_F^2 + ||\hat{u}\hat{u}^T||_F^2 - 2\langle uu^T, \hat{u}\hat{u}^T \rangle = 2 \sin^2 \angle(u, \hat{u})$$

so plugging everything back in we get the desired result.