Stats300b Étude 4 Solution

Due: Thursday, February 11 at 5:00pm on Gradescope.

**Question 7.10** (Low-rank matrix sensing)**:** In this question, we consider the problem of recovering a low-rank matrix from linear observations, showing that (with high probability) this is possible under a Gaussian random measurement model. We assume we observe triples $(X_i, Z_i, Y_i) \in \mathbb{R}^d \times R^d \times \mathbb{R}$ where

$$Y_i = \langle X_i Z_i^T, \Theta^\star \rangle = \mathrm{tr}(Z_i X_i^T \Theta^\star) = X_i^T \Theta^\star Z_i \tag{7.1}$$

for $X_i$ and $Z_i \stackrel{\text{iid}}{\sim} \mathsf{N}(0, I_d)$ and independent, where $\Theta^\star \in \mathbb{R}^d$ is an unknown rank $r$ matrix. (Here we use the standard notation on matrices that $\langle A, B \rangle = \mathrm{tr}(A^T B)$.) There is no noise in this observation model. We would like to recover $\Theta^\star$ from $n$ such measurements.

(a) Show that for any $d \times d$ matrix $A$,

$$\mathbb{E}[|X^T A Z|] \ge \frac{2}{\pi} \|A\|_{\mathrm{Fr}} \quad \text{and} \quad \mathbb{E}[|X^T A Z|^2] = \|A\|_{\mathrm{Fr}}^2 .$$

*Hint:* To prove the first inequality, first condition on $Z$. Then note that for any norm $\|\cdot\|$ and random vector $W$, $\mathbb{E}[\|W\|] \ge \|\mathbb{E}[|W|]\|$, where $|W|$ is the elementwise absolute value of $W$. Recognize that $\|w\| := \sqrt{\sum_{i=1}^d \sigma_i^2 w_i^2}$ is a norm on $w \in \mathbb{R}^d$.

(b) Argue that there exist numerical constants $c_0, c_1 > 0$ such that for any fixed matrix $A \in \mathbb{R}^{d\times d}$, we have

$$\mathbb{P}\left( \frac{1}{n} \sum_{i=1}^n |\langle X_i Z_i^T, A \rangle| \le c_0 \|A\|_{\mathrm{Fr}} \right) \le \exp(-c_1 n) .$$

*Hint:* For a constant $c > 0$, define the random variables $B_i = 1$ if $|\langle X_i Z_i^T, A\rangle| \ge c \|A\|_{\mathrm{Fr}}$ and $B_i = 0$ otherwise. Use the Paley-Zygmund inequality (Ex. 1.11) to show that $\mathbb{P}(B_i = 1) \ge p$, where $p > 0$ is a numerical constant, and then bound $\mathbb{P}(\overline{B}_n \le \mathbb{E}[B]/2)$.

(c) Using the covering number bounds in Ex. 7.8, show there exist numerical constants $0 < c_0, c_1$ and $C < \infty$ such that with probability at least $1 - e^{-c_1 n}$,

$$\frac{1}{n} \sum_{i=1}^n |X_i^T A Z_i| \ge c_0 \|A\|_{\mathrm{Fr}} \tag{7.2}$$

for all rank $r$ matrices $A \in \mathbb{R}^{d\times d}$ as long as $n \ge C dr \log(dr)$. You may assume $dr$ is large if that is convenient. You may also use that

$$\frac{1}{n} \sum_{i=1}^n \left\| Z_i X_i^T \right\|_{\mathrm{Fr}} = \frac{1}{n} \sum_{i=1}^n \|Z_i\|_2 \|X_i\|_2 \le \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \|Z_i\|_2^2 + \frac{1}{2} \|X_i\|_2^2 \right) \stackrel{(\star)}{\le} 2d$$

where inequality $(\star)$ holds with probability at least $1 - e^{-c_0 dn}$. *Hint:* note that inequality (7.2) is homogeneous in $A$.

(d) Assume that $\Theta^\star$ is rank $r$ in the sensing model (7.1). Argue that there exist numerical constants $0 < c_0, c_1$ and $C < \infty$ such that with probability at least $1 - e^{-cn}$,

$$\frac{1}{n} \sum_{i=1}^n |X_i^T \Theta Z_i - Y_i| \ge c_0 \|\Theta - \Theta^\star\|_{\mathrm{Fr}}$$

simultaneously for all rank $r$ matrices $\Theta$ as long as $n \ge C dr \log(dr)$.

(e) For loss $\ell(t) = |t|$, explain what part (d) tells us about the empirical minimizer

$$\widehat{\Theta}_n := \operatorname*{argmin}_{\Theta \in \mathbb{R}^{d \times d}} \left\{ P_n \ell(\langle XZ^T, \Theta \rangle - Y) \mid \text{rank}(\Theta) \le r \right\}.$$

In one sentence, compare the sample size $n$ versus the number of parameters in $\Theta^\star \in \mathbb{R}^{d \times d}$.

**Answer:**

(a) We have $\mathbb{E}[|X^T AZ| \mid Z = z] = \sqrt{2/\pi} \|Az\|_2$. For $Z \sim \mathsf{N}(0, I)$, we have $AZ \sim \mathsf{N}(0, AA^T)$. Letting $AA^T$ have eigenvalue decomposition $AA^T = U\Sigma U^T$ for a diagonal $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_d^2)$, we have $\mathbb{E}[\|AZ\|_2] = \mathbb{E}[\|\Sigma^{1/2} W\|_2]$ for $W \sim \mathsf{N}(0, I)$, and $\mathbb{E}[\|\Sigma^{1/2} W\|_2] = \mathbb{E}[\sqrt{\sum_{j=1}^d \sigma_j^2 W_j^2}]$. The function $t \mapsto \sqrt{\sum_{j=1}^d \sigma_j^2 t_j^2}$ is a Mahalanobis norm and so is convex, and it is invariant to signs, so Jensen's inequality gives

$$\mathbb{E}\left[\left\|\Sigma^{1/2} W\right\|_2\right] \ge \sqrt{\sum_{j=1}^d \sigma_j^2 \mathbb{E}[|W_j|]^2} = \sqrt{\frac{2}{\pi} \operatorname{tr}(\Sigma)},$$

which gives the first result.

The second is trivial: we have $\mathbb{E}[(X^T AZ)^2] = \mathbb{E}[\operatorname{tr}(X^T AZZ^T A^T X)] = \mathbb{E}[\operatorname{tr}(XX^T AZZ^T A^T)] = \operatorname{tr}(AA^T)$ as desired.

(b) By Exercise (1.11), we have

$$\mathbb{P}\left(|X^T AZ| \ge \frac{1}{2}\mathbb{E}[|X^T AZ|]\right) \ge \frac{1}{4} \frac{\mathbb{E}[|X^T AZ|]^2}{\mathbb{E}[(X^T AZ)^2]} \ge \frac{1}{\pi^2}$$

by part (a), and substituting the lower bound $\mathbb{E}[|X^T AZ|] \ge \frac{2}{\pi} \|A\|_{\text{Fr}}$ gives that with probability at least $\pi^{-2}$, we have $|X^T AZ| \ge \frac{1}{\pi} \|A\|_{\text{Fr}}$. Define the binary random variables $B_i = 1$ if $|X_i^T AZ_i| \ge \frac{1}{\pi} \|A\|_{\text{Fr}}$ and $B_i = 0$ otherwise, so that $\mathbb{E}[B_i] \ge \frac{1}{\pi^2}$. Then for $\overline{B}_n = \frac{1}{n} \sum_{i=1}^n B_i$,

$$\mathbb{P}(\overline{B}_n \le \mathbb{E}[B_1]/2) \le \exp(-c_1 n)$$

for a numerical constant $c_1 > 0$ by Hoeffding's inequality. As $\frac{1}{n} \sum_{i=1}^n |X_i^T AZ_i| \le \frac{1}{2\pi} \|A\|_{\text{Fr}}$ implies we must have $\overline{B}_n \le \frac{1}{2}\mathbb{E}[B_1]$, this gives the claimed bound.

(c) First, we use the covering number bounds. Let $\{A_1, \ldots, A_N\}$ be an $\epsilon$-cover of the rank $r$ matrices on the Frobenius sphere $\{A \mid \|A\|_{\text{Fr}} = 1\}$, which by Q. 7.8 has cardinality $N \le (Cr/\epsilon)^{2rd}$. Define the good event

$$\mathcal{E} := \bigcap_{j=1}^N \left\{ \frac{1}{n} \sum_{i=1}^n |\langle X_i Z_i^T, A_j \rangle| \ge c_0 \|A_j\|_{\text{Fr}} \right\} \cap \left\{ P_n \left\|XZ^T\right\|_{\text{Fr}} \le 2d \right\},$$

which by a union bound occurs with probability at least $1 - (N+1)e^{-cn}$. Then for any rank $r$ matrix $A \in \{A \mid \|A\|_{\text{Fr}} = 1\}$, there exists $A_j$ such that $\|A - A_j\|_{\text{Fr}} \le \epsilon$, and

$$P_n |\langle XZ^T, A \rangle| \ge P_n |\langle XZ^T, A_j \rangle| - P_n |\langle XZ^T, A - A_j \rangle| \stackrel{(i)}{\ge} P_n |\langle XZ^T, A_j \rangle| - \epsilon P_n \left\|XZ^T\right\|_{\text{Fr}} \stackrel{(ii)}{\ge} c_0 \|A_j\|_{\text{Fr}} - 2d\epsilon,$$

where inequality $(i)$ is Cauchy-Schwarz and inequality $(ii)$ uses the event $\mathcal{E}$. Notably $\min_j \|A_j\|_{\text{Fr}} \geq 1 - \epsilon$, and so taking $\epsilon = \frac{c_0}{4d}$ gives that

$$P_n|\langle XZ^T, A\rangle| \geq c_0 \|A\|_{\text{Fr}} \tag{7.3}$$

for all rank $r$ matrices $A$ with $\|A\|_{\text{Fr}} = 1$ on the event $\mathcal{E}$. This event occurs with probability at least

$$1 - (N+1)\exp(-c_1 n) \geq 1 - \exp\left(-cn + Cdr\log(dr)\right),$$

where the numerical constants $c, C$ may change their values from inequality to inequality.

Finally, note that inequality (7.3) is homogeneous in $\|A\|_{\text{Fr}}$ and so must hold for all rank $r$ matrices $A \in \mathbb{R}^{d\times d}$.

(d) The difference $\Theta - \Theta^\star$ is rank $2r$, so we simply apply the previous part of the question to

$$P_n|\langle XZ^T, \Theta\rangle - \langle XZ^T, \Theta^\star\rangle| = P_n|\langle XZ^T, \underbrace{\Theta - \Theta^\star}_{=:A}\rangle|.$$

(e) Evidently, with probability at least $1 - e^{-cn}$, as soon as $n \geq Cdr\log(dr)$ we have $\widehat{\Theta}_n = \Theta^\star$. This is both surprising—the raw number of parameters in $\Theta^\star$ is $d^2$, as it is a $d \times d$ matrix, so in $n = dr \log dr \ll d^2$ rank-1 observations of $\Theta^\star$ we get perfect recovery—and unsurprising, as the matrix is rank $r$ and so effectively only has $O(dr)$ parameters to estimate, and we are getting a bit more than that.

□