SUPPLEMENT TO "ROP: MATRIX RECOVERY VIA RANK-ONE PROJECTIONS" [1]

BY T. TONY CAI and ANRU ZHANG

Department of Statistics
The Wharton School
University of Pennsylvania

We prove in this supplement the technical lemmas used in the proofs of the main results.

Proof of Lemma 2.1. Suppose $\mathcal{X} : \mathbb{R}^{p_1\times p_2} \to \mathbb{R}^n$ is given by (1.6), we consider rank-1 matrices:

$$A_1 = e_1^{(p_1)} e_1^{(p_2)T}, \quad A_2 = \frac{\beta^{(1)}\gamma^{(1)}}{\|\beta^{(1)}\|_2\|\gamma^{(1)T}\|_2}$$

Here $e_1^{(p_1)}, e_2^{(p_2)}$ are the p_1- and p_2- dimensional vectors with first entry 1 and others 0, respectively. Then we have $\|A_1\|_F = \|A_2\|_F = 1$.

$$E\|\mathcal{X}(A_1)\|_2^2 = E\sum_{i=1}^n (\beta_1^{(i)})^2(\gamma_1^{(i)})^2 = n$$

$$\mathrm{Var}\|\mathcal{X}(A_1)\|_2^2 = \mathrm{Var}\sum_{i=1}^n (\beta_1^{(i)})^2(\gamma_1^{(i)})^2 = n\mathrm{Var}\left(\beta_1^2\gamma_1^2\right) = 8n.$$

By Chebyshev's inequality, we have for all $t > 1$,

$$P\left(\|\mathcal{X}(A_1)\|_2^2 \geq tn\right) \leq \frac{8}{n(t-1)^2} \tag{0.13}$$

On the other hand,

$$\begin{aligned}\|\mathcal{X}(A_2)\|_2^2 =& \|\beta^{(1)}\|_2^2\|\gamma^{(1)}\|_2^2 + \sum_{i=2}^n \Big(\frac{\beta^{(i)T}\beta^{(1)}}{\|\beta^{(1)}\|_2}\Big)^2\Big(\frac{\gamma^{(i)T}\gamma^{(1)}}{\|\gamma^{(1)}\|_2}\Big)^2 \\ \geq& \|\beta^{(1)}\|_2^2\|\gamma^{(1)}\|_2^2 \sim \chi^2(p_1)\cdot\chi^2(p_2)\end{aligned}$$

[1]The research was supported in part by NSF FRG Grant DMS-0854973, NSF Grant DMS-1208982 and NIH Grant R01 CA127334-05.

By Lemma 1 in [28], we know

$$P\left(\chi^2(p_1) \geq p_1/2\right) \leq \exp(-p_1/4), \quad P\left(\chi^2(p_2) \geq p_2/2\right) \leq \exp(-p_2/4)$$

Hence,

$$P\left(\|\mathcal{X}(A_2)\|_2^2 \geq p_1 p_2/4\right) \leq \exp(-p_1/4) + \exp(-p_2/4). \tag{0.14}$$

Combining (0.13) and (0.14), we can see

$$C_2/C_1 \geq \frac{\|\mathcal{X}(A_2)\|_2/(\sqrt{n}\|A_2\|_F)}{\|\mathcal{X}(A_1)\|_2/(\sqrt{n}\|A_1\|_F)} = \sqrt{\frac{\|\mathcal{X}(A_2)\|_2^2}{\|\mathcal{X}(A_1)\|_2^2}} \geq \sqrt{\frac{p_1 p_2/4}{tn}} \tag{0.15}$$

holds with probability at least $1 - e^{-p_1/4} - e^{-p_2/4} - \frac{8}{n(t-1)^2}$. □

Proof of Lemma 3.1. First, the common used definition for sub-Gaussian distribution of random variable X include the following two.

$$\exists c, C > 0, \text{ such that } P(|X| \geq t) \leq C\exp(-ct^2) \tag{0.16}$$

$$\exists c > 0, \text{ such that } Ee^{tX} \leq \exp(c^2t^2/2) \tag{0.17}$$

Suppose $\alpha_{\mathcal{P}}$ is finite, then we have

$$Ee^{tX} = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} EX^{2k} \leq \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \alpha^{2k}(2k-1)!! = \sum_{k=0}^{\infty} \frac{(t\alpha)^{2k}}{2^k k!} = \exp((\alpha t)^2/2)$$

namely X is sub-Gaussian. Now suppose X is sub-Gaussian, then

$$\begin{aligned} EX^{2k} &= (2k)\int_0^{\infty} P(|X| > t)t^{2k-1}dt \leq 2kC\int_0^{\infty} t^{2k-1}e^{-ct^2}dt \\ &= \frac{kC}{c^k}\int_0^{\infty} (ct^2)^{k-1}e^{-ct^2}d(ct^2) = \frac{k!C}{c^k} \leq \left(\frac{\max(C,1)}{c}\right)^k (2k-1)!! \end{aligned}$$

which implies that $\alpha_{\mathcal{P}} \leq \sqrt{\max(C,1)/c}$ is finite. □

Proof of Lemma 7.1. Without confusion, we simply use α to represent $\alpha_{\mathcal{P}}$ in the proof. Note that we can multiply A by a scale without loss of generality. So we assume throughout the proof that $\|A\|_F = 1$. We'll prove this lemma by steps. First, we show an inequality on the even moments of $|\beta^T A\gamma|$; next, we give a bound on the moment generation function of $|\beta^T A\gamma|$. Finally, we give the desired tail bound.

1. **Step 1: Even moments of $|\beta^T A\gamma|$.**
Assume that $x = (x_1, \cdots, x_{p_1}), y = (y_1, \cdots, y_{p_2})$ are two random i.i.d. standard normal distributed vectors. Based on the definition of $\alpha_{\mathcal{P}}$ in (3.2), we know

$$E\beta_i^{2k} = E\gamma_j^{2k} \le \alpha^{2k} E x_i^{2k} = \alpha^{2k} E y_j^{2k}; \qquad (0.18)$$
$$E\beta_i^{2k-1} = E\gamma_j^{2k-1} = E x_i^{2k-1} = E y_j^{2k-1} = 0$$

Consider the expansion of $E(\beta^T A\gamma)^{2k}$, where the non-zero terms can be written as

$$\prod_{l=1}^{2k} A_{i_l,j_l} \cdot \prod_{i=1}^{p_1} E\beta_i^{2s_i} \cdot \prod_{j=1}^{p_2} E\gamma_j^{2t_j}.$$

Here $s_1 + \cdots + s_{p_1} = t_1 + \cdots t_{p_2} = k$. By (0.18), this term can be bounded as

$$\left|\prod_{l=1}^{2k} A_{i_l,j_l} \cdot \prod_{i=1}^{p_1} E\beta_i^{2s_i} \cdot \prod_{j=1}^{p_2} E\gamma_j^{2t_j}\right| \le \prod_{l=1}^{2k} |A_{i_l,j_l}| \cdot \alpha^{4k} \cdot \prod_{i=1}^{p_1} E x_i^{2s_i} \cdot \prod_{j=1}^{p_2} E y_j^{2t_j}$$

The right hand side is exact the term in the expansion of $\alpha^{4k} E(x^T A_{abs} y)^{2k}$, where A_{abs} is the the element-wise absolute value of A. Therefore, we have

$$E[\beta^T A\gamma]^{2k} \le \alpha^{4k} E[x^T A_{abs} y]^{2k}. \qquad (0.19)$$

Now we suppose A_{abs} has singular value decomposition

$$A_{abs} = \sum_{i=1}^{p} a_i u_i v_i^T = U\text{diag}(a)V^T$$

where U, V are orthogonal and $a = (a_1, \cdots, a_p)$ is the singular value vector of A_{abs}. A well-known fact is that $\sum_i a_i^2 = \|A_{abs}\|_F^2 = \|A\|_F^2$. Since x, y are standard normal distributed, we can see that $x^T A_{abs} y$ and $x^T \text{diag}(a) y$ has the same distribution. So

$$E[x^T A_{abs} y]^{2k} = E[\sum_{i=1}^{p} a_i x_i y_i]^{2k}$$

Next, we note

$$z = \sum_{i=1}^{p} x_i \frac{a_i y_i}{\sqrt{\sum_{j=1}^{p} a_j^2 y_j^2}},$$

then z is standard normal distributed and independent of $\sqrt{\sum_{j=1}^{p} a_j^2 y_j^2}$ since

$$\sum_{i=1}^{p}\left(\frac{a_i y_i}{\sqrt{\sum_{j=1}^{p} a_j^2 y_j^2}}\right)^2 = 1.$$

and z given $y_1, \cdots, y_p$ is always standard normal distributed.
For integer $k \geq 1$,

$$\begin{aligned}
E[x^T A_{abs} y]^{2k} =& E\left| z \cdot \sqrt{\sum_{j=1}^{p} a_j^2 y_j^2} \right|^{2k} = E|z|^{2k} \cdot E(\sum_{j=1}^{p} a_j^2 y_j^2)^k \\
=& (2k-1)!! \cdot \sum_{\substack{k_1, k_2, \cdots, k_p \geq 0, \\ k_1 + \cdots + k_p = k}} \frac{k!}{\prod_{j=1}^{p} k_j!} \prod_{i=1}^{p} E(a_i^2 y_i^2)^{k_i} \\
=& \sum_{\substack{k_1, k_2, \cdots, k_p \geq 0, \\ k_1 + \cdots + k_p = k}} \frac{k!}{\prod_{j=1}^{p} (k_j)!} \prod_{i=1}^{p} a_i^{2k_i} \cdot \prod_{i=1}^{p} (2k_i - 1)!! \\
\leq & \sum_{\substack{k_1, k_2, \cdots, k_p \geq 0, \\ k_1 + \cdots + k_p = k}} \frac{k!}{\prod_{j=1}^{p} k_j!} \prod_{i=1}^{p} a_i^{2k_i} \cdot (2(k_1 + k_2 + \cdots k_p) - 1)!! \\
=& ((2k-1)!!)^2 \cdot \sum_{\substack{k_1, k_2, \cdots, k_p \geq 0, \\ k_1 + \cdots + k_p = k}} \frac{k!}{\prod_{j=1}^{p} k_j!} \prod_{i=1}^{p} (a_i^2)^{k_i} \\
=& ((2k-1)!!)^2 (\sum_{i=1}^{p} a_i^2)^k = ((2k-1)!!)^2 \|A\|_F^{2k}
\end{aligned}$$

Together with (0.19), we have

$$E[\beta^T A \gamma]^{2k} \leq \alpha^{4k} ((2k-1)!!)^2 \|A\|_F^{2k} \tag{0.20}$$

2. **Log-moment generation function of $|\beta^T A \gamma|$.**
By the bound of the even moments of $|\beta^T A \gamma|$, we can also give the estimate of odd moments, for integer $k \geq 1$,

$$\begin{aligned}
0 \leq E\left|\beta^T A \gamma\right|^{2k+1} &\leq \sqrt{E[\beta^T A \gamma]^{2k} \cdot E[\beta^T A \gamma]^{2k+2}} \\
&\leq \alpha^{4k+2} (2k-1)!! \cdot (2k+1)!! \leq \alpha^{4k+2} (2k+1)!
\end{aligned}$$

Also,

$$E\left[\beta^T A \gamma\right]^{2k} \leq \alpha^{4k} ((2k-1)!!)^2 \leq \alpha^{4k} (2k)!$$

So for all $k \geq 2$, $E|\beta^T A\gamma|^k \leq \alpha^{2k} k!$. Denote $\mu = E|\beta^T A\gamma|$, then for $0 \leq t < \frac{1}{\alpha^2}$,

$$Ee^{t|\beta^T A\gamma|} = 1+\sum_{k=1}^{\infty} \frac{t^k}{k!} E|\beta^T A\gamma|^k \leq 1+t\mu+\sum_{k=2}^{\infty} t^k \alpha^{2k} = 1+t\mu+\frac{t^2\alpha^4}{1-t\alpha^2}$$

For $-\frac{1}{\alpha^2} < t < 0$, we have

$$\begin{aligned} Ee^{t|\beta^T A\gamma|} =& 1 + t\mu + \sum_{k=1}^{\infty} t^{2k}\alpha^{4k} - \sum_{k=1}^{\infty} \frac{|t|^{2k+1}}{(2k+1)!} E|\beta^T A\gamma|^{2k+1} \\ \leq& 1 + t\mu + \frac{t^2\alpha^4}{1-|t|^2\alpha^4} \leq 1 + t\mu + \frac{t^2\alpha^4}{1-|t|\alpha^2} \end{aligned}$$

Hence, we have for all $-1/\alpha^2 < t < 1/\alpha^2$,

$$Ee^{t|\beta^T A\gamma|} \leq 1 + t\mu + \frac{t^2\alpha^4}{1-|t|\alpha^2}.$$

Note that $\log(1+x) \leq 1+x$ for all $-1 < x < \infty$, we have

$$\log E\exp(t(|\beta^T A\gamma| - \mu)) = \log E\exp(t|\beta^T A\gamma|) - t\mu \leq \frac{t^2\alpha^4}{1-|t|\alpha^2}. \tag{0.21}$$

for all $-1/\alpha^2 < t < 1/\alpha^2$.

3. **The tail bound of $\|\mathcal{X}(A)\|_1/n$.**
Finally, we estimate the tail bound of $\|\mathcal{X}(A)\|1/n$. Note that

$$\|\mathcal{X}(A)\|_1/n = \left(\sum_{j=1}^{n} \left|\beta^{(j)T} A\gamma^{(j)}\right|\right)/n,$$

based on (0.21), the logarithm of moment generating function of $\|\mathcal{X}(A)\|_1/n$ satisfies

$$\log E\exp\left(t(\|\mathcal{X}(A)\|_1/n - \mu)\right) = n\log E\exp\left(\frac{t}{n}(|\beta^T A\gamma| - \mu)\right) \leq \frac{t^2\alpha^4/n}{1-|t|\alpha^2/n}$$

By the proof of Lemma 1 in [28], we know $\|\mathcal{X}(A)\|_1/n - \mu$ has the tail bound,

$$\begin{aligned} P(\|\mathcal{X}(A)\|_1/n - \mu \geq \alpha^2(x/n + 2\sqrt{x/n})) \leq \exp(-x) \\ P(\|\mathcal{X}(A)\|_1/n - \mu \leq \alpha^2(x/n + 2\sqrt{x/n})) \leq \exp(-x) \end{aligned}$$

Finally we set $\delta = \sqrt{x/n}$, by Lemma 7.2, $1/(3\alpha^4) \leq \mu \leq 1$, we finish the proof of Lemma. □

Proof of Lemma 7.2. Since $\mathcal{P}$ is symmetric and of variance 1, we have $E\beta_i = E\gamma_j = 0$, $E\beta_i^2 = E\gamma_j^2 = 1$, $E\beta_i^4 = E\gamma_j^4 \le 3\alpha_{\mathcal{P}}^4$ for all i, j. Then by some expansions and calculations,

$$E\left(\beta^T A\gamma\right)^2 = E\left(\sum_{i,j}\beta_i A_{ij}\gamma_j\right)^2 = \sum_{i,j} E\beta_i^2 A_{ij}^2\gamma_j^2 = \sum_{i,j} A_{ij}^2 = \|A\|_F^2$$

By the first part in the proof of Lemma 7.1, we have

$$E|\beta^T A\gamma|^4 \le 9\alpha_{\mathcal{P}}^8\|A\|_F^4$$

By Hölder's inequality,

$$E|\beta^T A\gamma| \le \sqrt{E|\beta^T A\gamma|^2} = \|A\|_F$$

which gives the right of the original inequality. For the left, note that

$$E|\beta^T A\gamma|^2 \le \left(E|\beta^T A\gamma|\right)^{2/3} \cdot \left(E|\beta^T A\gamma|^4\right)^{1/3}$$

So

$$E|\beta^T A\gamma| \ge \sqrt{\frac{\left(E|\beta^T A\gamma|^2\right)^3}{E|\beta^T A\gamma|^4}} \ge \frac{\|A\|_F}{3\alpha_{\mathcal{P}}^4}. \quad \square$$

Proof of Lemma 7.3.

- We first prove the sub-Gaussian part of the lemma. The moment generating function of $|z_i|$ ($t \ge 0$) and z_i^2 ($0 \le t < 1/(2\gamma)$) satisfy

$$\begin{aligned}
Ee^{t|z_i|} &= -\int_0^\infty \exp(t\lambda)dP(|X| \ge \lambda) = 1 + \int_0^\infty P(|X| \ge \lambda)d\exp(t\lambda) \\
&\le 1 + \int_0^\infty 2t\exp(t\lambda - \lambda^2/(2\gamma^2))d\lambda \\
&\le 1 + 2t\exp(t^2\gamma^2/2)\int_0^\infty \exp(-(\lambda-\gamma^2)^2/(2\gamma^2))d\lambda \\
&\le 1 + 2t\exp\left(t^2\gamma^2/2\right)\sqrt{2\pi}\gamma \\
&\le \exp(t^2\gamma^2/2)\left(1 + 2t\sqrt{2\pi}\gamma\right) \\
&\le \exp\left(t^2\gamma^2/2 + 2\sqrt{2\pi}\gamma t\right)
\end{aligned}$$

$$\begin{aligned}
Ee^{tz_i^2} &= -\int_0^\infty \exp(t\lambda^2)dP(|X| \ge \lambda) = 1 + \int_0^\infty P(|X| \ge \lambda)d\exp(t\lambda^2) \\
&\le 1 + \int_0^\infty 4t\lambda\exp\left(-\lambda^2(1/(2\gamma^2) - t)\right)d\lambda \\
&= 1 + \frac{2t}{1/(2\gamma^2) - t}
\end{aligned}$$

Then the moment generating function of $\|z\|_1/n$ and $\|z\|_2^2/n$ satisfies

$$Ee^{t\|z\|_1/n} = \left(Ee^{t\|z_i\|/n}\right)^n \le \exp\left(t^2\gamma^2/(2n) + 2\sqrt{2\pi}t\gamma\right)$$

$$Ee^{t\|z\|_2^2/n} = \left(1 + \frac{2t/n}{1/(2\gamma^2) - t/n}\right)^n \le \exp\left(\frac{2t}{1/(2\gamma^2) - t/n}\right)$$

Hence for $C \ge 0$,

$$\begin{aligned} P\left(\|z\|_1/n \ge C\right) &\le \frac{E\exp(t\|z\|_1/n)}{\exp(tC)} \le \exp\left(t^2\gamma^2/(2n) + t\left(2\sqrt{2\pi}\gamma - C\right)\right) \\ &= \exp\left(\frac{\gamma^2}{2n}\left(t + \frac{n(2\sqrt{2\pi}\gamma - C)}{\gamma^2}\right)^2 - \frac{n(2\sqrt{2\pi}\gamma - C)^2}{2\gamma^2}\right) \end{aligned}$$

For $C > 2\sqrt{2\pi}\gamma$, we can set $t = \frac{\gamma^2}{n(C - 2\sqrt{2\pi}\gamma)}$, then

$$P\left(\|z\|_1/n \ge C\right) \le \exp\left(-\frac{n(C - 2\sqrt{2\pi}\gamma)^2}{2\gamma^2}\right)$$

Now we consider the tail bound of $\|z\|_2^2/n$. For $C > 4\gamma^2$,

$$P(\|z\|_2^2/n \ge C) = \frac{E\exp(t\|z\|_2^2/n)}{\exp(tC)} = \exp\left(\frac{2t^2\gamma^2C/n - t(C - 4\gamma^2)}{1 - 2\gamma^2 t/n}\right).$$

We set $t = \frac{C - 4\gamma^2}{4C\gamma^2/n}$,

$$P\left(\|z\|_2^2/n \ge C\right) \le \exp\left(-\frac{n(C - 4\gamma^2)^2}{8\gamma^2 C(1 - 2\gamma^2 t/n)}\right) \le \exp\left(-\frac{n(C - 4\gamma^2)^2}{8\gamma^2 C}\right)$$

Finally we consider $\|z\|_\infty$,

$$P(\|z\|_\infty \le C\sqrt{\log n}\gamma) \le 2n\exp(-C^2 \log n\gamma^2/(2\gamma^2)) = 2n^{-(C^2/2-1)}$$

- Next, we consider the Gaussian part of the lemma. The bound of $\|z\|_2$ is already given by Lemma 5.1 in [7]. For $\|z\|_1$, we can see $E|z_i|^2 = \sigma^2$,

$$E|z_i| = \frac{\sigma}{\sqrt{2\pi}}\int_0^\infty xe^{-x^2/(2)}dx = \sigma\sqrt{2/\pi}$$

Hence, $E(\|z\|_1/n) = \sigma\sqrt{2/\pi}$, $\text{Var}(\|z\|_1/n) = \text{Var}(|z_i|)/n = (1 - 2/\pi)\sigma^2/n$. By Chebyshev' inequality,

$$\begin{aligned} P(\|z\|_1/n \ge \sigma) &\le P\left(\left|\|z\|_1/n - \sigma\sqrt{2/\pi}\right| \ge \sigma(1 - \sqrt{2/\pi})\right) \\ &\le \frac{\text{Var}(\|z\|_1/n)}{\sigma^2(1 - \sqrt{2/\pi})^2} = \frac{1 + \sqrt{2/\pi}}{(1 - \sqrt{2/\pi})n} \end{aligned}$$

For the bound of $\|z\|_\infty$, we have

$$P(\|z\|_\infty \geq 2\sqrt{\log n}\sigma) \leq \sum_{i=1}^{n} P(|z_i| \leq 2\sqrt{\log n}\sigma)$$
$$\leq n \cdot \frac{2}{2\sqrt{2\pi \log n}} \exp(-\frac{1}{2} \cdot 4 \log n) = \frac{1}{n\sqrt{2\pi \log n}} \quad \square$$

Proof of Lemma 7.4. Again without confusion, we simply use α to represent $\alpha_{\mathcal{P}}$ in the proof. The proof also requires some knowledge of moment generation function and ε-net method. We'll prove by steps.

- **Moment Generation Function of $a^T \mathcal{X}^*(z) b$.** Suppose $a \in \mathbb{R}^{p_1}$, $b \in \mathbb{R}^{p_2}$ are fixed unit vectors. In order to handle the operator norm of $\mathcal{X}^*(z)$, we first consider $a^T \mathcal{X}^*(z) b$. Note that

$$a^T \mathcal{X}^*(z) b = \sum_{i=1}^{n} z_i a^T \beta^{(i)} \gamma^{(i)T} b \tag{0.22}$$

Denote $X_i = a^T \beta^{(i)}$, $Y_i = b^T \gamma^{(i)}$, then $\{X_i\}_{i=1}^{p_1}, \{Y_i\}_{i=1}^{p_2}$ are two independent sets of i.i.d. sub-Gaussian samples. Moreover by $\beta^{(i)}, \gamma^{(i)}$ are i.i.d. from symmetric distribution $\mathcal{P}$, one can show

$$E(X_i)^{2k-1} = E(\sum_j a_j \beta_j)^{2k-1} = 0,$$

$$E(X_i)^{2k} = E(\sum_{j=1}^{p_1} a_j \beta_j)^{2k} = \sum_{k_1 + \cdots k_{p_1} = k} \frac{k!}{k_1! \cdots k_{p_1}!} \left(\prod_{i=1}^{p_1} a_i^{2k_i} E(\beta_i^{2k_i}) \right)$$
$$\leq \sum_{k_1 + \cdots k_{p_1} = k} \frac{k!}{k_1! \cdots k_{p_1}!} \left(\prod_{i=1}^{p_1} a_i^{2k_i} E(x_i^{2k_i}) \alpha^{2k} \right)$$
$$= \alpha^{2k} E(\sum_{i=1}^{p_1} a_i x_i)^{2k} \leq \alpha^{2k} (2k-1)!!$$

Here $x_i \overset{iid}{\sim} N(0,1)$. Similarly, $E(Y_i)^{2k-1} = 0$, $E(Y_i)^{2k} \leq \alpha^{2k}(2k-1)!!$. Then for $|t| < 1/\alpha^2$,

(0.23)

$$\begin{aligned} E\exp(tX_iY_i) &= \sum_{k=0}^{\infty} \frac{t^k E(X_iY_i)^k}{k!} \leq \sum_{k=0}^{\infty} \frac{t^{2k}(\alpha^{2k}(2k-1)!!)^2}{(2k)!} \\ &= \sum_{k=0}^{\infty} \frac{(t\alpha^2)^{2k}(2k-1)!!}{2^k k!} = \sum_{k=0}^{\infty} (t\alpha^2)^{2k} \cdot (-1)^k \binom{-1/2}{k} \\ &= \frac{1}{\sqrt{1-t^2\alpha^4}} \end{aligned}$$

Now for fixed $z \in \mathbb{R}^n$, the logarithm of the moment generating function of $a^T\mathcal{X}^*(z)b$ satisfies

$$\begin{aligned} \log E\exp(ta^T\mathcal{X}^*(z)b) &= \sum_{i=1}^{n} \log E\exp(tz_iX_iY_i) \leq \sum_{i=1}^{n} -\frac{1}{2}\log(1-t^2z_i^2\alpha^4) \\ &\leq \sum_{i=1}^{n} \frac{t^2z_i^2\alpha^4}{2(1-t^2z_i^2\alpha^4)} \leq \frac{t^2\|z\|_2^2\alpha^4}{2(1-t^2\|z\|_\infty^2\alpha^4)} \\ &\leq \frac{t^2\|z\|_2^2\alpha^4}{2(1-|t|\|z\|_\infty\alpha^2)} \end{aligned}$$

for any $|t| < 1/(\|z\|_\infty\alpha^2)$. Here we used the fact that

$$-\log(1-x) = \sum_{i=1}^{\infty} \frac{x^i}{i} \leq \sum_{i=1}^{\infty} x^i = \frac{x}{1-x}$$

for $0 \leq x < 1$.

- **Tail Bound of $a^T\mathcal{X}^*(z)b$**
  By the proof of Lemma 1 in [28], we know for fixed $z \in \mathbb{R}^n$, $a^T\mathcal{X}^*(z)b$ has tail bound: for $x > 0$ and fixed a, b, we have

$$P\left(a^T\mathcal{X}^*(z)b \geq \alpha^2\left(\|z\|_\infty x + \|z\|_2\sqrt{2x}\right)\right) \leq \exp(-x);$$

$$P\left(a^T\mathcal{X}^*(z)b \leq -\alpha^2\left(\|z\|_\infty x + \|z\|_2\sqrt{2x}\right)\right) \leq \exp(-x).$$

Set $x = C(p_1+p_2)$, we have

(0.24)

$$\begin{aligned} &P\left(|a^T\mathcal{X}^*(z)b| \geq \alpha^2\left(\|z\|_\infty C(p_1+p_2) + \|z\|_2\sqrt{2C(p_1+p_2)}\right)\right) \\ \leq &2\exp(-C(p_1+p_2)). \end{aligned}$$

For convenience, We denote

$$T = \alpha^2 \left(\|z\|_\infty (C(p_1+p_2)) + \|z\|_2 \sqrt{2C(p_1+p_2)} \right). \tag{0.25}$$

- **ε-net and the upper bound of $\|\mathcal{X}^*(z)\|$.**
In this step, we still fix z. We use the ε-net method to derive the upper bound of $\|\mathcal{X}^*(z)\|_2$, which is given by

$$\|\mathcal{X}^*(z)\|_2 = \sup_{a\in\mathbb{R}^{p_1}, b\in\mathbb{R}^{p_2}} a^T\mathcal{X}^*(z)b$$

From Lemma 2.5 in [39], we can find an ε-net A in the unit sphere of $\mathbb{R}^{p_1}$, i.e. for all a in the unit sphere of $\mathbb{R}^{p_1}$, there exists $a' \in A$ such that $\|a'-a\|_2 \le \epsilon$. Besides, $|A| \le (1+2/\varepsilon)^{p_1}$. Similarly, there exists ϵ-net B of the unit ball of $\mathbb{R}^{p_2}$ such that $|B| \le (1+2/\varepsilon)^{p_2}$.
By (0.24), we have

$$P\left(\left|a^T\mathcal{X}^*(z)b\right| \ge T, \exists a\in A, b\in B\right) \le 2(1+2/\varepsilon)^{p_1+p_2}\exp(-C(p_1+p_2)) \tag{0.26}$$

Now we consider under the event that $\left|a^T\mathcal{X}^*(z)b\right| \le T, \forall a\in A, b\in B$. Suppose $\mu = \|\mathcal{X}^*(z)\|_2 = \max_{\|a\|_2=\|b\|_2=1} a^T\mathcal{X}^*(z)b$, $(a^*, b^*) = \arg\max_{a,b} a^T\mathcal{X}^*(z)b$, then we can find $a'\in A, b'\in B$ such that $\|a'-a^*\| \le \varepsilon$, $\|b'-b^*\| \le \varepsilon$. Then,

$$\begin{aligned}
\mu &= \left|a^{*T}\mathcal{X}^*(z)b^*\right| \\
&= \left|a'^T\mathcal{X}^*(z)b'\right| + \left|(a'-a^*)^T\mathcal{X}^*(z)b'\right| + \left|a^{*T}\mathcal{X}^*(z)(b^*-b')\right| \\
&\le T + (\|a'-a^*\|_2 + \|b'-b^*\|_2)\cdot\|\mathcal{X}^*(z)\| \le T + 2\varepsilon\mu
\end{aligned}$$

This means $\mu \le T/(1-2\varepsilon)$. Therefore, when $\left|a^T\mathcal{X}^*(z)b\right| \le T, \forall a\in A, b\in B$, we have $\|\mathcal{X}^*(z)\| \le T/(1-2\varepsilon)$.
- Finally, we set $\varepsilon = 1/3$, under the event that

$$\begin{aligned}
&\left|a^T\mathcal{X}^*(z)b\right| \le T \\
=&\alpha^2\left(\|z\|_\infty(C(p_1+p_2)) + \|z\|_2\sqrt{2C(p_1+p_2)}\right), \forall a\in A, b\in B
\end{aligned}$$

we have

$$\|\mathcal{X}^*(z)\| \le T/(1-2\varepsilon) \le 3\alpha^2\left(\|z\|_\infty C(p_1+p_2) + \|z\|_2\sqrt{2C(p_1+p_2)}\right)$$

By (0.26), the probability that all the event happen is at least

$$1 - 2\exp\left(-(C-\log 7)(p_1+p_2)\right)$$

This finished the proof of the lemma. □

Proof of Proposition 7.1.. In the proof, we will use α to represent $\alpha_{\mathcal{P}}$ without any confusion. The ideas of the proof of Proposition 7.1 follows from [36], [13].

Denote $S_{kr} = \{X \in \mathbb{R}^{p_1\times p_2} : \text{rank}(X) \le kr, \|X\|_F = 1\}$. By Lemma 3.1 in [13], for any $\varepsilon > 0$, there exists ε-net S'_{kr} such that $|S'_{kr}| \le (9/\varepsilon)^{(p_1+p_2+1)kr}$.

For given C_1, C_2 such that $C_1 < 1/(3\alpha^4)$, $C_2 > 1$, we set $C_1' = \frac{C_1+1/(3\alpha^4)}{2}$, $C_2' = \frac{C_2+1}{2}$. We can choose δ_0 small enough such that

$$\alpha^2\left(2\delta_0+\delta_0^2\right) \le \min\left(\frac{1/(3\alpha^4)-C_1}{2}, \frac{C_2-1}{2}\right)$$

then by Lemma 7.1, for any given $A \in \mathbb{R}^{p_1\times p_2}$, we have $C_1' \le \|\mathcal{X}(A)\|_1/n \le C_2'$ with probability at least $1-2\exp(-\delta_0^2 n)$. Hence,

$$P\left(C_1' \le \|\mathcal{X}(A)\|_1/n \le C_2',\ \text{for all } A \in S_r'\right) \ge 1-2(9/\varepsilon)^{kr(p_1+p_2+1)}\cdot\exp(-\delta_0^2 n)$$

Next, we'll estimate the bound of $\|\mathcal{X}(A)\|_1/n$ on the whole set S_{kr} provided that $C_1' \le \|\mathcal{X}(A)\|_1/n \le C_2'$ for all $A \in S'_{kr}$. Define

$$\kappa_1 = \inf_{A\in S_{kr}} \|\mathcal{X}(A)\|_1/n \text{ and } \kappa_2 = \sup_{A\in S_{kr}} \|\mathcal{X}(A)\|_1/n.$$

For any $A \in S_{kr}$, there exists $A' \in S'_{kr}$ such that $\|A-A'\|_F \le \varepsilon$. So

$$\|\mathcal{X}(A)\|_1/n \le \|\mathcal{X}(A')\|_1/n+\|\mathcal{X}(A-A')\|_1/n \le C_2'\|A\|+\kappa_2\|A-A'\|_F \le C_2'+\kappa_2\varepsilon$$

$$\|\mathcal{X}(A)\|_1/n \ge \|\mathcal{X}(A')\|_1/n-\|\mathcal{X}(A-A')\|_1/n \ge C_1'\|A\|-\kappa_2\|A-A'\|_F \ge C_1'-\kappa_2\varepsilon$$

which mean

$$\kappa_2 = \sup_{A\in S_{kr}} \|\mathcal{X}(A)\|_F \le C_2'+\varepsilon\kappa_2, \quad \kappa_1 = \inf_{A\in S_{kr}} \|\mathcal{X}(A)\|_F \ge C_1'-\varepsilon\kappa_2$$

namely, $\kappa_2 \le C_2'/(1-\varepsilon)$, $\kappa_1 \ge C_1'-\varepsilon\kappa_2$. We choose

$$\varepsilon \le \min\left(\frac{C_2-1}{2C_2}, \frac{1/(3\alpha^4)-C_1}{2C_2}\right),$$

by some calculations we can see $\kappa_1 \ge C_1$, $\kappa_2 \le C_2$.

To sum up, we can choose δ_0, ε only depending on C_1, C_2, α, to ensure

$$C_1 \le \kappa_1 = \inf_{A\in S_{kr}} \|\mathcal{X}(A)\|_1/n \le \sup_{A\in S_{kr}} \|\mathcal{X}(A)\|_1/n = \kappa_2 \le C_2$$

with probability at least $1-2(9/\varepsilon)^{kr(p_1+p_2+1)}\exp(-\delta_0^2 n)$. The last step is to estimate the probability above. We choose $D \geq 8k\log(9/\varepsilon)/\delta_0^2$, then for $n \geq Dr(p_1+p_2)$, we have

$$\delta_0^2 n/2 \geq 4\log(9/\varepsilon)kr(p_1+p_2) \geq \log 2 + 2\log(9/\varepsilon)kr(p_1+p_2+1),$$

$$\begin{aligned}1-2(9/\varepsilon)^{kr(p_1+p_2+1)}e^{-\delta_0^2 n} &= 1-\exp(-\delta_0^2 n + \log 2 + kr(p_1+p_2+1)\log(9/\varepsilon))\\ \geq& 1-\exp(-\delta_0^2 n/2).\end{aligned}$$

Finally, we finish the proof of the Theorem by choosing $\delta \leq \delta_0^2/2$. □

Proof of Lemma 7.7.. The idea of the proof is originated from [41, 31]. We provide the proof here for the completeness of the paper. Note $p = \min(p_1,p_2)$, suppose for any matrix B, $\sigma_i(B)$ is the i-th largest singular value of B. By Lemma 2 in [31], we have

$$\begin{aligned}&\|A_{\max(r)}\|_* + \|A_{-\max(r)}\|_*\\ =&\|A\|_* \geq \|A_*\|_* = \|A-(-R)\|_* \geq \sum_{i=1}^{p}|\sigma_i(A)-\sigma_i(-R)|\\ \geq&\sum_{i=1}^{r}(\sigma_i(A)-\sigma_i(R)) + \sum_{i=r+1}^{p}(\sigma_i(R)-\sigma_i(A))\\ =&\|A_{\max(r)}\|_* - \|A_{-\max(r)}\|_* + \|R_{-\max(r)}\|_* - \|R_{\max(r)}\|_*\end{aligned}$$

which implies (7.3). □

Proof of Lemma 7.8.. Suppose $R = A_* - A$, then we have

$$\|\mathcal{X}(R)\|_1/n \leq \lambda_1 \tag{0.27}$$

Since $\|A_*\|_* \leq \|A\|_*$. By Lemma 7.7, we must have (7.3). Suppose $p = \min(p_1,p_2)$ and R has the singular value decomposition, $R = \sum_{i=1}^{p}\sigma_i u_i v_i^T = U\mathrm{diag}(\vec{\sigma})V^T$, then $\vec{\sigma}_{-\max(kr)}$ satisfies

$$\|\vec{\sigma}_{-\max(kr)}\|_\infty \leq \sigma_{kr},$$

$$\begin{aligned}\|\vec{\sigma}_{-\max(kr)}\|_1 =&\|\vec{\sigma}_{-\max(r)}\|_1 - (\sigma_{r+1}+\cdots+\sigma_{kr})\\ \leq&\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr}\end{aligned}$$

Set

$$\theta = \max\left(\sigma_{kr}, (\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr})/(kr)\right),$$

then $\|\vec{\sigma}_{-\max(kr)}\|_\infty \leq \theta$, $\|\vec{\sigma}_{-\max(kr)}\|_1 \leq kr\theta$. Similarly to the proof of Theorem 2.1, apply Lemma 7.6, we can get $b^{(i)} \in \mathbb{R}^n$, $\lambda_i \geq 0$, $i = 1, \cdots, N$ such that $\vec{\sigma}_{-\max(kr)} = \sum_{i=1}^N \lambda_u b^{(i)}$ and (7.2). Hence,

$$\|b^{(i)}\|_2 \leq \sqrt{\|b^{(i)}\|_1 \cdot \|b^{(i)}\|_\infty} \leq \sqrt{\theta(\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr})}$$

If $\theta = \sigma_{kr}$, we can optimize over σ_{kr} in the inequality,

$$\begin{aligned}\|b^{(i)}\|_2 \leq& \sqrt{\sigma_{kr}(\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr})}\\ \leq& \frac{\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_*}{2\sqrt{r(k-1)}};\end{aligned}$$

if $\theta = (\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr})/(kr)$, we have

$$\begin{aligned}\|b^{(i)}\|_2 \leq& \frac{\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_* - (k-1)r\sigma_{kr}}{\sqrt{kr}}\\ \leq& \frac{\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_*}{\sqrt{kr}}.\end{aligned} \tag{0.28}$$

Since $k \geq 2$, we always have (0.28). Next, we define $B_i = U\mathrm{diag}(b^{(i)})V^T$, then the rank of B_i are at most kr, $\sum_{i=1}^N \lambda_i B_i = R_{-\max(kr)}$ and $\|B_i\|_F = \|b^{(i)}\|_2$. Then

$$\begin{aligned}\lambda_1 \geq& \|\mathcal{X}(R)\|_1/n \geq \|\mathcal{X}(R_{\max(kr)})\|_1/n - \|\mathcal{X}(R_{-\max(kr)})\|_1/n\\ \geq& C_1\|R_{\max(kr)}\|_F - \sum_{i=1}^N \|\mathcal{X}(\lambda_i B_i)\|_1/n\\ \geq& C_1\|R_{\max(kr)}\|_F - \sum_{i=1}^N \lambda_i C_2\|B_i\|_F\\ \geq& C_1\|R_{\max(kr)}\|_F - C_2\frac{\|R_{\max(r)}\|_* + 2\|A_{-\max(r)}\|_*}{\sqrt{kr}}\\ \geq& C_1\|R_{\max(kr)}\|_F - \frac{C_2}{\sqrt{k}}\|R_{\max(kr)}\|_F - \frac{2C_2\|A_{-\max(r)}\|_*}{\sqrt{kr}},\end{aligned} \tag{0.29}$$

where the last inequality is due to $\|R_{\max(kr)}\|_F \geq \|R_{\max(r)}\|_F \geq \sqrt{r}\|R_{\max(r)}\|_*$. Therefore,

$$\|R_{\max(kr)}\|_F \leq \frac{\lambda_1}{C_1 - C_2/\sqrt{k}} + \frac{2\|A_{-\max(r)}\|_*}{\sqrt{r}(\sqrt{k}C_1/C_2 - 1)} \tag{0.30}$$

Finally,

$$\begin{aligned}\|R_{-\max(kr)}\|_F =&\|\vec{\sigma}_{-\max(kr)}\|_2 \leq \sqrt{\|\vec{\sigma}_{-\max(kr)}\|_1 \cdot \|\vec{\sigma}_{-\max(kr)}\|_\infty} \\ \leq&\sqrt{\sigma_{kr} \cdot (\|\vec{\sigma}_{-\max(r)}\|_1 - r(k-1)\sigma_{kr})} \\ \leq&\frac{\|\vec{\sigma}_{-\max(r)}\|_1}{2\sqrt{r(k-1)}} \leq \frac{\|\vec{\sigma}_{\max(r)}\|_1 + 2\|A_{-\max(r)}\|_*}{2\sqrt{r(k-1)}} \\ \leq&\frac{\|R_{\max(r)}\|_F}{2\sqrt{k-1}} + \frac{\|A_{-\max(r)}\|_*}{\sqrt{r(k-1)}}\end{aligned}$$

Therefore,

$$\begin{aligned}\|R\|_F =&\sqrt{\|R_{\max(kr)}\|_F^2 + \|R_{-\max(kr)}\|_F^2} \\ \leq&\sqrt{\|R_{\max(kr)}\|_F^2 + \left(\frac{\|R_{\max(r)}\|_F}{2\sqrt{k-1}} + \frac{\|A_{-\max(r)}\|_*}{\sqrt{r(k-1)}}\right)^2} \\ \leq&\sqrt{1 + \frac{1}{4(k-1)}}\|R_{\max(kr)}\|_F + \frac{\|A_{-\max(r)}\|_*}{\sqrt{r(k-1)}} \\ \leq&\left(1 + \frac{1}{8(k-1)}\right)\|R_{\max(kr)}\|_F + \frac{\|A_{-\max(r)}\|_*}{\sqrt{r(k-1)}} \\ \leq&\frac{2}{C_1 - C_2/\sqrt{k}}\lambda_1 + \left(\frac{3}{\sqrt{k}C_1/C_2 - 1} + \frac{1}{\sqrt{k-1}}\right)\frac{\|A_{-\max(r)}\|_*}{\sqrt{r}}\end{aligned} \tag{0.31}$$

Proof of Lemma 7.9.. The proof of this theorem is similar to the proof of Lemma 7.8. Suppose $R = A_* - A$. In this case we have

$$\|\mathcal{X}^*\mathcal{X}(R)\| \leq \lambda_2 \tag{0.32}$$

instead of (0.27). Besides, since $\|A_*\|_* \leq \|A\|_*$ and Lemma 7.7, we still have (7.3). With the similar argument as (0.29), we have

(0.33)

$$\begin{aligned}\lambda_2\|R\|_* \geq \langle R, \mathcal{X}^*\mathcal{X}(R)\rangle = \|\mathcal{X}(R)\|_2^2 \geq \|\mathcal{X}(R)\|_1^2/n \\ \geq n\left(C_1\|R_{\max(kr)}\|_F - \frac{C_2}{\sqrt{k}}\|R_{\max(kr)}\|_F - \frac{2C_2\|A_{-\max(r)}\|_*}{\sqrt{kr}}\right)_+^2\end{aligned}$$

Here $(x)_+$ means $\max(x, 0)$. Besides,

$$\begin{aligned}\lambda_2\|R\|_* &\leq \lambda_2\left(\|R_{\max(r)}\|_* + (\|R_{\max(r)}\|_* + 2\|A_{-\max(r)}\|_*)\right) \\ &\leq 2\lambda_2\left(\sqrt{r}\|R_{\max(r)}\|_F + \|A_{-\max(r)}\|_*\right) \\ &\leq 2\lambda_2\left(\sqrt{r}\|R_{\max(kr)}\|_F + \|A_{-\max(r)}\|_*\right)\end{aligned}$$

Suppose $x = \|R_{\max(kr)}\|_F$, $y = \|A_{-\max(r)}\|_*/\sqrt{r}$. Based on the previous two inequalities, we have

$$n\left((C_1 - C_2/\sqrt{k})x - \frac{2C_2}{\sqrt{k}}y\right)_+^2 \le 2\sqrt{r}(x+y)\lambda_2$$

When $x \ge \frac{2C_2 y}{\sqrt{k}(C_1 - C_2/\sqrt{k})}$, the inequality above leads to

(0.34)
$$n(C_1 - C_2/\sqrt{k})^2 x^2 - \left(2n(C_1 - C_2/\sqrt{k})\frac{2C_2}{\sqrt{k}}y + 2\sqrt{r}\lambda_2\right)x - 2\sqrt{r}\lambda_2 y \le 0.$$

Note that for second order inequality $ax^2 - bx - c \le 0$, $a > 0$, $b, c \ge 0$, we have $x \le \frac{b+\sqrt{b^2+4ac}}{2a} \le b/a + \sqrt{c/a}$. Hence we can get an upper bound of x from (0.34).

$$\begin{aligned} x &\le \frac{2\sqrt{r}\lambda_2}{n(C_1 - C_2/\sqrt{k})^2} + \frac{4C_2/\sqrt{k}y}{(C_1 - C_2/\sqrt{k})} + \frac{\sqrt{2\sqrt{r}\lambda_2 y}}{\sqrt{n}(C_1 - C_2/\sqrt{k})} \\ &\le \frac{2\sqrt{r}\lambda_2}{n(C_1 - C_2/\sqrt{k})^2} + \frac{4C_2/\sqrt{k}y}{(C_1 - C_2/\sqrt{k})} + \frac{\sqrt{r}\lambda_2}{n(C_1 - C_2/\sqrt{k})^2} + \frac{1}{2}y. \end{aligned}$$

Hence whenever $x \ge \frac{2C_2 y}{\sqrt{k}(C_1 - C_2/\sqrt{k})}$ or not,

(0.35)
$$\begin{aligned} &\|R_{\max(kr)}\|_F = x \\ \le &\max\left\{\frac{2C_2 y}{\sqrt{k}(C_1 - C_2/\sqrt{k})}, \frac{3\sqrt{r}\lambda_2}{n(C_1 - C_2/\sqrt{k})^2} + \left(\frac{4C_2/\sqrt{k}}{(C_1 - C_2/\sqrt{k})} + \frac{1}{2}\right)y\right\} \\ \le &\frac{3\sqrt{r}\lambda_2}{n(C_1 - C_2/\sqrt{k})^2} + \left(\frac{4C_2/\sqrt{k}}{(C_1 - C_2/\sqrt{k})} + \frac{1}{2}\right)\frac{\|A_{-\max(r)}\|_*}{\sqrt{r}}. \end{aligned}$$

Finally, similarly to (0.31) in Lemma 7.8, we can get the upper bound of $\|R\|_F$,

$$\begin{aligned} \|R\|_F &\le \left(1 + \frac{1}{8k-8}\right)\|R_{\max(kr)}\|_F + \frac{\|A_{-\max(r)}\|_*}{\sqrt{r(k-1)}} \\ &\le \frac{4}{(C_1 - C_2/\sqrt{k})^2} \cdot \frac{\sqrt{r}(\varepsilon + \eta)}{n} + \left(\frac{5}{\sqrt{k}C_1/C_2 - 1} + \frac{1}{\sqrt{k-1}} + 1\right)\frac{\|A_{-\max(r)}\|_*}{\sqrt{r}} \end{aligned}$$

which finished the proof of lemma 7.9. □

0.7. *Proof of Propositions 2.2, 2.3 and 2.4.*. We first show Proposition 2.2. Denote $\mathcal{X}_1, \mathcal{X}_2 : \mathbb{R}^{p\times p} \to \mathbb{R}^{\lfloor \frac{n}{2} \rfloor}$ such that

$$[\mathcal{X}_1]_i(B) = (\frac{\beta^{(2i-1)} + \beta^{(2i)}}{\sqrt{2}})^T B(\frac{\beta^{(2i-1)} - \beta^{(2i)}}{\sqrt{2}}), \quad i = 1, \cdots, \lfloor \frac{n}{2} \rfloor \tag{0.36}$$

$$[\mathcal{X}_2](B) = (\frac{\beta^{(2i-1)} - \beta^{(2i)}}{\sqrt{2}})^T B(\frac{\beta^{(2i-1)} + \beta^{(2i)}}{\sqrt{2}}), \quad i = 1, \cdots, \lfloor \frac{n}{2} \rfloor \tag{0.37}$$

Note that $\frac{1}{\sqrt{2}}\left(\beta^{(2i-1)} + \beta^{(2i)}\right)$ and $\frac{1}{\sqrt{2}}\left(\beta^{(2i-1)} - \beta^{(2i)}\right)$ are independent i.i.d. standard normal samples, so both $\mathcal{X}_1$ and $\mathcal{X}_2$ are ROP design (see (1.6)). By Corollary 2.1, we know there exists uniform constant C such that whenever $\lfloor \frac{n}{2} \rfloor \geq Cr \cdot 2p$, $\mathcal{X}_1$ satisfies the following property with probability at least $1 - \exp(-n\delta)$,

$$\forall A \in \{A \in \mathbb{R}^{p\times p} : rank(A) \leq r\}, \quad A = \arg\min_{B\in\mathbb{R}^{p\times p}} \|B\|_* \quad \text{subject to} \quad \mathcal{X}_1(B) = \mathcal{X}_1(A). \tag{0.38}$$

Now we consider the event that (0.38) holds. We note that for any symmetric matrix B,

$$[\mathcal{X}_1]_i(B) = \frac{1}{2}\beta^{(2i-1)T}B\beta^{(2i-1)} - \frac{1}{2}\beta^{(2i)T}B\beta^{(2i)} = \frac{1}{2}\left([\mathcal{X}]_{2i-1}(B) - [\mathcal{X}]_{2i}(B)\right)$$

So $\mathcal{X}(B) = \mathcal{X}(A)$ implies $\mathcal{X}_1(B) = \mathcal{X}_1(A)$ for symmetric A and B. Also, since A is feasible in programming (2.13), we have

$$\begin{aligned}
\|A\|_* \geq & \min_{B\in\mathbb{S}^p} \|B\|_* \quad \text{subject to} \quad \mathcal{X}(B) = \mathcal{X}(A)\\
\geq & \min_{B\in\mathbb{R}^{p\times p}} \|B\|_* \quad \text{subject to} \quad \mathcal{X}_1(B) = \mathcal{X}_1(A)\\
= & \|A\|_*,
\end{aligned}$$

So we can conclude that A can be exactly recovered by (2.13) given (0.38) holds. In summary, for $n \geq 6Crp$, with probability at least $1-\exp(-n\delta)$, $\mathcal{X}$ satisfies (0.38), then programming (2.13) can recover all $A \in \mathbb{S}^p$ of rank at most r. $\square$

Next, we consider Proposition 2.3. The idea of the proof is similar to Proposition 2.1. Define $\tilde{z} \in \mathbb{R}^{\lfloor \frac{n}{2} \rfloor}$ such that

$$\tilde{z}_i = z_{2i-1} - z_{2i}, \quad i = 1, \cdots, \lfloor \frac{n}{2} \rfloor. \tag{0.39}$$

Then $\tilde{z} \overset{iid}{\sim} N(0,2)$. We shall also point out two facts, $\tilde{z} = \tilde{y} - \tilde{\mathcal{X}}(A)$ and $\mathcal{X}^* = \mathcal{X}_1^* + \mathcal{X}_2^*$ (defined as (0.36), (0.37)). By Lemma 7.3, we know

$$P(\|z\|_1/n > \sigma) \leq \frac{9}{n}$$

$$P(\|\tilde{z}\|_2 > \sigma\sqrt{2n}) \leq \frac{1}{\lfloor n/2 \rfloor}$$

$$P(\|\tilde{z}\|_\infty > 2\sigma\sqrt{2\log n}) \leq \frac{1}{\lfloor n/2 \rfloor}$$

Hence,

$$\begin{aligned}
&P\left(A \text{ is NOT in the feasible set of programming (2.17)}\right) \\
=&P\left(\|z\|_1/n > \sigma, \text{ or } \|\tilde{\mathcal{X}}^*(\tilde{z})\| > \eta\right) \\
=&P\left(\|z\|_1/n > \sigma, \text{ or } \|\tilde{\mathcal{X}}^*(\tilde{z})\| > 24\sigma\sqrt{pn} + 48\sigma p\sqrt{2\log n}\right) \\
\leq&P(\|z\|_1/n > \sigma) + P_{\mathcal{X}}(\|\tilde{z}\|_2 > \sigma\sqrt{2n}) + P(\|\tilde{z}\|_\infty > 2\sigma\sqrt{2\log n}) \\
&+ P_{\mathcal{X}}\left(\|\tilde{\mathcal{X}}^*(\tilde{z})\| > 24p\|\tilde{z}\|_\infty + 12\sqrt{2p}\|\tilde{z}\|_2\right) \\
\leq&\frac{9}{n} + \frac{2}{\lfloor n/2 \rfloor} + P_{\mathcal{X}}\left(\|\mathcal{X}_1(\tilde{z})\| > 12p\|\tilde{z}\|_\infty + 6\sqrt{2p}\|\tilde{z}\|_2\right) \\
&+ P_{\mathcal{X}}\left(\|\mathcal{X}_2(\tilde{z})\| > 12p\|\tilde{z}\|_\infty + 6\sqrt{2p}\|\tilde{z}\|_2\right) \\
\leq&\frac{15}{n} + 4\exp(-2p(2-\log 7))
\end{aligned}$$

When A is in the feasible set of programming (2.17), we have $\|\hat{A}\|_* \leq \|A\|_*$ and

$$\begin{aligned}
\|\tilde{\mathcal{X}}(\hat{A}-A)\|_1 &= \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left|[\mathcal{X}]_{2i-1}(\hat{A}-A) - [\mathcal{X}]_{2i}(\hat{A}-A)\right| \\
&\leq \|\mathcal{X}(\hat{A}-A)\|_1 \leq \|y - \mathcal{X}(\hat{A})\|_1 + \|\mathcal{X}(A) - y\|_1 \\
&\leq \|y - \mathcal{X}(\hat{A})\|_1 + \|z\|_1 \leq 2n\sigma
\end{aligned} \tag{0.40}$$

$$\begin{aligned}
&\|\tilde{\mathcal{X}}^*\tilde{\mathcal{X}}(\hat{A}-A)\| \leq \|\tilde{\mathcal{X}}^*(\tilde{y} - \tilde{\mathcal{X}}(\hat{A}))\| + \|\tilde{\mathcal{X}}^*(\tilde{\mathcal{X}}(A) - \tilde{y})\| \\
=&\|\tilde{\mathcal{X}}^*(\tilde{y} - \tilde{\mathcal{X}}(\hat{A}))\| + \|\tilde{\mathcal{X}}^*(\tilde{z})\| \leq 2\eta
\end{aligned} \tag{0.41}$$

Similarly as the proof to Proposition 2.1, by Theorem 2.2, there exists constant D, δ' such that if $n \geq Drp$, $\mathcal{X}_1$ satisfies RUB of order $10r$ with constants C_1, C_2 such that $C_2/C_1 < \sqrt{10}$ with probability at least $1 - \exp(-n\delta')$. Now we suppose the following two events happen,

1. $\mathcal{X}_1$ satisfies RUB of order $10r$ and constants C_1, C_2 satisfying $C_2/C_1 < \sqrt{10}$,
2. A is feasible in the programming (2.17).

Since $\tilde{\mathcal{X}}(B) = 2\mathcal{X}_1(B)$ for any symmetric matrix B, by $\mathcal{X}_1$ satisfies RUB condition, we have $\tilde{\mathcal{X}}$ satisfies RUB for symmetric matrices of order $10r$ and constants $2C_1, 2C_2$ satisfying $(2C_2)/(2C_1) < \sqrt{10}$. We note that the proof of Lemmas 7.8 and 7.9 still apply for $\tilde{\mathcal{X}}$ in the symmetric matrices class, so we can get (2.18) based on (0.40) and (0.41) under those two events happen. Finally the probability that these events happen is at least $1 - 15/n - 4\exp(-p\delta) - \exp(-n\delta)$ for $\delta < \min(2(2-\log 7), \delta')$, which finished the proof of Proposition 2.3. □

Finally we consider Proposition 2.4. Denote $p' = \lfloor p/2 \rfloor$, $r' = \lfloor r/2 \rfloor$. By $r, p \geq 2$, we have $r' \geq r/3, p' \geq p/3$. Define a sub-class of the class rank-r symmetric matrices,

$$\mathcal{G} = \left\{ A \in \mathbb{S}^p : A = \begin{array}{c} \begin{array}{cc} p' & p-p' \end{array} \\ \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \end{array} \begin{array}{c} p' \\ p-p' \end{array}, B \in \mathbb{R}^{p' \times (p-p')}, \mathrm{rank}(B) \leq r' \right\}$$

we can see $\forall A \in \mathcal{G}$,

$$[\mathcal{X}(A)]_i = \beta^{(i)T} A \beta^{(i)} = 2(\beta_1^{(i)}, \cdots, \beta_{p'}^{(i)}) B (\beta_{p'+1}^{(i)}, \cdots, \beta_p^{(i)})^T,$$

so in $\mathcal{G}$ the SROP model becomes

$$\frac{y_i}{2} = (\beta_1^{(i)}, \cdots, \beta_{p'}^{(i)}) B (\beta_{p'+1}^{(i)}, \cdots, \beta_p^{(i)})^T + \frac{z_i}{2}, \quad \frac{z}{2} \overset{iid}{\sim} N(0, \sigma^2/4)$$

which is an ROP model which we already discussed in section 2. We omit the rest of the proof as it can be followed by the proof of Theorem 2.4. □

Proof of Proposition 3.1.. The proof can follow from the proof of Proposition 2.3 and Theorem 3.1 once we can prove that in high probability, $\mathcal{X}_1$ (defined in (0.36)) satisfies RUB condition with C_1, C_2 such that C_2/C_1 is bounded. This can be proved similarly as Proposition 7.1, where we only need to edit the proof that we use the following Lemma 0.11 instead of Lemma 7.1.

LEMMA 0.11. *Suppose $A \in \mathbb{R}^{p \times p}$ is a fixed matrix (not necessarily symmetric) and $\mathcal{X}_1$ is given by (0.36). $\beta^{(i)}$ is a set of p-dimensional vectors such* *that* $\overset{iid}{\sim} \mathcal{P}$, *where* $\mathcal{P}$ *is some symmetric variance 1 sub-Gaussian distribution except Rademacher* ± 1 *distribution. Then for* $\delta > 0$, *we have*

$$\left(\frac{\min^{3/2}(\mathrm{Var}(\mathcal{P}^2)/2, 1)}{3(2\alpha_{\mathcal{P}})^4} - 8\alpha_{\mathcal{P}}^2\delta - 4\alpha_{\mathcal{P}}^2\delta^2\right)\|A\|_F \le \frac{\|\mathcal{X}_1(A)\|_1}{\lfloor n/2 \rfloor} \le \left(\sqrt{3/2}\alpha_{\mathcal{P}}^2 + 8\alpha_{\mathcal{P}}^2\delta + 4\alpha_{\mathcal{P}}^2\delta^2\right)\|A\|_F. \tag{0.42}$$

with probability at least $1 - 2\exp(-\delta^2\lfloor n/2 \rfloor)$.

The proof of Lemma 0.11 is in the Appendix right after this paragraph. Note that provided $\mathcal{P}$ is symmetric and with variance 1, $\mathrm{Var}(\mathcal{P}^2) = 0$ if and only $\mathcal{P}$ is Rademacher ± 1 and A is diagonal, in which the lower bound of (0.42) becomes meaningless. So we only exclude Rademacher ± 1 distribution from the result. □

Proof of Lemma 0.11.. The proof of Lemma 0.11 is basically the same to Lemma 7.1. We only need to redo two parts of the proof, where there are major differences.

1. Part 1. "Step 1. Even moments of $|\frac{1}{2}(\beta^{(1)} + \beta^{(2)})^T A(\beta^{(1)} - \beta^{(2)})|$.". First, based on $\mathcal{P}$ is symmetric and with variance 1, we have $E\mathcal{P}^{2k+1} = 0$, $E\mathcal{P}^{2k} \le \alpha^{2k}Ex^{2k} = \alpha^{2k}(2k-1)!!$.Then we can calculate that

$$E(\beta_i^{(1)} + \beta_i^{(2)})^{2k+1} = E(\beta_i^{(1)} - \beta_i^{(2)})^{2k+1} = 0$$

$$\begin{aligned}&E(\beta_i^{(1)} + \beta_i^{(2)})^{2k}\\ =&\sum_{l=0}^{k}\binom{2k}{2l}E(\beta_i^{(1)})^{2l}E(\beta_i^{(2)})^{2(k-l)} \le \alpha^{2k}\sum_{l=0}^{k}\binom{2k}{2l}(2l-1)!!(2(k-l)-1)!!\\ =&\alpha^{2k}\sum_{l=0}^{k}\frac{(2k-1)!!2^k k!}{2^l l! 2^{k-l}(k-l)!} = \alpha^{2k}(2k-1)!!\sum_{l=0}^{k}\binom{k}{l} = 2^k\alpha^{2k}(2k-1)!!\end{aligned}$$

Similarly, $E(\beta_i^{(1)} - \beta_i^{(2)})^{2k} \le 2^k\alpha^{2k}(2k-1)!!$. Next, we can similarly consider the expansion of $E\left(\frac{1}{2}(\beta^{(1)} + \beta^{(2)})A(\beta^{(1)} - \beta^{(2)})\right)^{2k}$, where the non-zero terms can be written as

$$\frac{1}{2^{2k}}\prod_{l=1}^{2k}A_{i_l,j_l}\prod_{i=1}^{p}E(\beta_i^{(1)} + \beta_i^{(2)})^{2s_i}\prod_{j=1}^{p}E(\beta_j^{(1)} - \beta_j^{(2)})^{2t_j}$$

Here $s_1 + \cdots + s_p = t_1 + \cdots + t_p = k$. This term can be bounded as

$$\begin{aligned}
&\left| \frac{1}{2^{2k}} \prod_{l=1}^{2k} A_{i_l,j_l} \prod_{i=1}^{p} E(\beta_i^{(1)} + \beta_i^{(2)})^{2s_i} (\beta_i^{(1)} - \beta_i^{(2)})^{2t_i} \right| \\
\leq & \frac{1}{2^{2k}} \prod_{l=1}^{2k} |A_{i_l,j_l}| \cdot \prod_{i=1}^{p} \left(\frac{s_i}{s_i+t_i} E\left(\beta_i^{(1)} + \beta_i^{(2)}\right)^{2s_i+2t_i} + \frac{t_i}{s_i+t_i} E\left(\beta_i^{(1)} - \beta_i^{(2)}\right)^{2s_i+2t_i} \right) \\
\leq & \frac{1}{2^{2k}} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} 2^{s_i+t_i} \alpha^{2(s_i+t_i)} (2(s_i+t_i)-1)!! \\
= & \alpha^{4k} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} \frac{(2(s_i+t_i))!}{2^{s_i+t_i}(s_i+t_i)!} \leq \alpha^{4k} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} \frac{((2s_i+2t_i)!!)^2}{2^{s_i+t_i} s_i! t_i!} \\
\leq & \alpha^{4k} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} \frac{2^{2(s_i+t_i)}\,((s_i+t_i)!)^2}{2^{s_i+t_i} s_i! t_i!} \leq \alpha^{4k} 2^{4p} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} \frac{(2s_i)!(2t_i)!}{(2s_i)!!(2t_i)!!} \\
= & \alpha^{4k} 2^{4p} \prod_{l=1}^{2k} |A_{i_l,j_l}| \prod_{i=1}^{p} E x_i^{2s_i} \cdot \prod_{i=1}^{p} E y_i^{2t_i}
\end{aligned}$$

Here we assume that $x_i, y_i \overset{iid}{\sim} N(0,1)$. The right hand side of the inequality above is exactly the term in the expansion of $(2\alpha)^{4k} E(x^T A_{abs} y)^{2k}$, where A_{abs} is the element-wise absolute value of A. Therefore, we have

$$E\left(\frac{1}{2}\left(\beta^{(1)} + \beta^{(2)}\right)^T A \left(\beta^{(1)} - \beta^{(2)}\right)\right)^{2k} \leq (2\alpha)^{4k} E[x^T A_{abs} y]^{2k}$$

Now we follow the same argument of the rest part of Step 1 in the proof of Lemma 7.1, we can prove that

(0.43)
$$E\left(\frac{1}{2}\left(\beta^{(1)} + \beta^{(2)}\right)^T A \left(\beta^{(1)} - \beta^{(2)}\right)\right)^{2k} \leq (2\alpha)^{4k} ((2k-1)!!)^2 \|A\|_F^{2k}$$

2. Part 2. The upper and lower bound of $\mu = E\left|\frac{1}{2}\left(\beta^{(1)} + \beta^{(2)}\right)^T A \left(\beta^{(1)} - \beta^{(2)}\right)\right|$. To follow the argument of the proof of Step 3 in Lemma 7.1, we need to derive a new bound for $\mu = E|\frac{1}{2}\left(\beta^{(1)} + \beta^{(2)}\right)^T A \left(\beta^{(1)} - \beta^{(2)}\right)|$. First,

we denote $M = |\frac{1}{2}\left(\beta^{(1)}+\beta^{(2)}\right)^T A\left(\beta^{(1)}-\beta^{(2)}\right)|$, then

$$\begin{aligned}\mu =& EM \le \sqrt{EM^2}\\ =&\sqrt{E\left(\frac{1}{2}(\beta^{(1)}+\beta^{(2)})^T A(\beta^{(1)}-\beta^{(2)})\right)^2}\\ =&\sqrt{\frac{1}{4}\sum_{i,j}E(\beta_i^{(1)}+\beta_i^{(2)})^2 A_{ij}^2(\beta_j^{(1)}-\beta_j^{(2)})^2}\\ =&\sqrt{\sum_{i\neq j}A_{ij}^2+\frac{1}{2}\mathrm{Var}(\mathcal{P}^2)\sum_i A_{ii}^2} \le \sqrt{\sum_{i\neq j}A_{ij}^2+\frac{3}{2}\alpha_{\mathcal{P}}^4\sum_i A_{ii}^2}\\ \le&\sqrt{\frac{3}{2}}\alpha_{\mathcal{P}}^2\|A\|_F\end{aligned}$$

On the other hand,

$$\begin{aligned}EM^2 =& E\left(\frac{1}{2}(\beta^{(1)}+\beta^{(2)})^T A(\beta^{(1)}-\beta^{(2)})\right)^2\\ =&\sum_{i\neq j}A_{ij}^2+\frac{1}{2}\mathrm{Var}(\mathcal{P}^2)\sum_i A_{ii}^2 \ge \min(\frac{1}{2}\mathrm{Var}(\mathcal{P}^2),1)\|A\|_F^2,\end{aligned}$$

By (0.43), we also have $EM^4 \le 9(2\alpha)^8\|A\|_F^{2k}$. By Hölder's inequality, $EM^2 \le (EM)^{2/3}(EM^4)^{1/3}$. Hence,

$$\mu \ge \sqrt{\frac{(EM^2)^3}{EM^4}} \ge \frac{\min^{3/2}(\mathrm{Var}(\mathcal{P}^2)/2,1)\|A\|_F}{3(2\alpha)^4}$$

To sum up, instead of Lemma 7.2, we have the bound of μ as follows,

$$\frac{\min^{3/2}(\mathrm{Var}(\mathcal{P}^2)/2,1)\|A\|_F}{3(2\alpha_{\mathcal{P}})^4} \le \mu \le \sqrt{3/2}\alpha_{\mathcal{P}}^2\|A\|_F \tag{0.44}$$

The rest of the proof follows the proof of Lemma 7.1 with modifications of constants, which we do not go into details. □

Proof of Lemma 7.10.. Note that $\xi_i\Big|\beta^{(i)} \sim N(0,\beta^{(i)T}\Sigma\beta^{(i)})$, we can assume that

$$\xi_i^2 = \beta^{(i)T}\Sigma\beta^{(i)}\cdot Z_i, \tag{0.45}$$

where $Z_i \overset{iid}{\sim} (N(0,1))^2$ and $Z_i, \beta^{(i)}$ are independent. Based on the definition of z in (4.3), we have

(0.46)

$$z_i = y_i - [\mathcal{X}(\Sigma_0)]_i = \xi_i^2 - \beta^{(i)\intercal}\Sigma\beta^{(i)} = \beta^{(i)T}\Sigma\beta^{(i)}\left(Z_i-1\right), \quad i=1,\cdots,n.$$

We also denote

$$Q_1 = \frac{C_1}{n}\sum_{i=1}^{n}\xi_i^2, \quad Q_2 = \frac{C_2^2}{n}\sum_{i=1}^{n}\xi_i^4, \quad Q_3 = C_3 \cdot \log n \max_{1\le i\le n}\xi_i^2.$$

- We'll first consider the former part of (7.8), (7.9) and (7.10). Suppose $Z \sim (N(0,1))^2$. It is well known that the non-central m-th moment of Z is $(2m-1)!!$, so we have

$$E\left(C_1 Z - |Z-1|\right) \ge C_1 - \sqrt{E\left|Z-1\right|^2} = C_1 - \sqrt{2} \tag{0.47}$$

$$E\left(C_1 Z - |Z-1|\right)^2 \le E\left(C_1 Z\right)^2 + E\left(Z-1\right)^2 = 3C_1^2 + 2 \tag{0.48}$$

$$E\left(C_2^2 Z^2 - (Z-1)^2\right) = 3C_2^2 - 2 \tag{0.49}$$

$$E\left(C_2^2 Z^2 - (Z-1)^2\right)^2 \le C_2^4 E\left(Z\right)^4 + E\left(Z-1\right)^4 = 105C_2^4 + 60 \tag{0.50}$$

Next we consider the random quadratic form of Σ. Suppose $\beta = (\beta_1, \cdots, \beta_p) \overset{iid}{\sim} N(0,1)$, $X_1, \cdots \overset{iid}{\sim} N(0,1)$, $\lambda_1(\Sigma), \cdots, \lambda_p(\Sigma)$ are the eigenvalues of Σ. Since Σ is positive definite, we have

$$E\beta^T\Sigma\beta = \mathrm{tr}(\Sigma) = \|\Sigma\|_* \tag{0.51}$$

$$\begin{aligned} E(\beta^T\Sigma\beta)^2 =& E(\sum_i \beta_i^2\Sigma_{ii} + 2\sum_{i<j}\Sigma_{ij}\beta_i\beta_j)^2 \\ =& \sum_i \Sigma_{ii}^2 E\beta_i^4 + 2\sum_{i<j}\Sigma_{ii}\Sigma_{jj}E\beta_i^2\beta_j^2 + \sum_{i<j}4\Sigma_{ij}^2 E\beta_i^2\beta_j^2 \\ =& 2(\sum_{i,j}\Sigma_{ij}^2) + (\sum_{ii}\Sigma_i)^2 = 2\|\Sigma\|_F^2 + \|\Sigma\|_*^2 \end{aligned}$$

Hence,

$$\|\Sigma\|_*^2 \le E(\beta^T\Sigma\beta)^2 \le 3\|\Sigma\|_*^2 \tag{0.52}$$

$$\begin{aligned} E(\beta^T\Sigma\beta)^4 &= E(\sum_{i=1}^{p}\lambda_i(\Sigma)X_i^2)^4 \\ &= \sum_{1\le i,j,s,t\le p}\lambda_i(\Sigma)\lambda_j(\Sigma)\lambda_s(\Sigma)\lambda_t(\Sigma)EX_i^2X_j^2X_s^2X_t^2 \\ &\le \sum_{1\le i,j,s,t\le p}\lambda_i(\Sigma)\lambda_j(\Sigma)\lambda_s(\Sigma)\lambda_t(\Sigma)7!! = 105\|\Sigma\|_*^4 \end{aligned} \tag{0.53}$$

Then we consider $C_1\xi_i^2 - |z_i|$ and $C_2^2\xi_i^4 - z_i^2$. By (0.45) and (0.46), we have

$$C_1\xi_i^2 - |z_i| = \beta^{(i)T}\Sigma\beta^{(i)} \cdot (C_1 Z_i - |Z_i - 1|)\,,$$

$$C_2^2\xi_i^4 - z_i^2 = \left(\beta^{(i)T}\Sigma\beta^{(i)}\right)^2 \left(C_2^2 Z^2 - (Z-1)^2\right),$$

while $\beta^{(i)}$ and Z_i are independent in the equation above. By (0.47)-(0.53), we obtain an estimation of the first and second moment of these two quantities as

$$E\left(C_1\xi_i^2 - |z_i|\right) \geq (C_1 - \sqrt{2})\|\Sigma\|_* \tag{0.54}$$

$$\begin{aligned} &\mathrm{Var}\left(C_1\xi_i^2 - |z_i|\right) \leq E\left(C_1\xi_i^2 - |z_i|\right)^2 \\ =&E\left(C_1 Z_i - |Z_i - 1|\right)^2 E\left(\beta^{(i)T}\Sigma\beta^{(i)}\right)^2 \leq \left(9C_1^2 + 6\right)\|\Sigma\|_*^2 \end{aligned} \tag{0.55}$$

$$E\left(C_2^2\xi_i^4 - z_i^2\right) \geq \left(3C_2^2 - 2\right) \cdot \|\Sigma\|_*^2 \tag{0.56}$$

$$\mathrm{Var}\left(C_2^2\xi_i^4 - z_i^2\right) \leq E\left(C_2^2\xi_i^4 - z_i^2\right)^2 \leq 105(105C_2^4 + 60)\|\Sigma\|_*^4 \tag{0.57}$$

We note that $Q_1 - \|z\|_1/n$ and $Q_2 - \|z\|_2^2/n$ are the average of n i.i.d. copy of $C_1\xi_i^2 - |z_i|$ and $C_2^2\xi_i^4 - z_i^2$. We can immediately get an estimation of the mean and variance of $Q_1 - \|z\|_1/n$ and $Q_2 - \|z\|_2^2/n$ based on (0.54)-(0.57). Finally, by Chebyshev's inequality,

$$P\left(Q_1 \leq \|z\|_1/n\right) \leq \frac{\mathrm{Var}(Q_1 - \|z\|_1/n)}{\left(E\left(Q_1 - \|z\|_1/n\right)\right)^2} \leq \frac{9C_1^2 + 6}{n(C_1 - \sqrt{2})^2}$$

$$P\left(Q_2 \leq \frac{\|z\|_2^2}{n}\right) \leq \frac{\mathrm{Var}\left(Q_2 - \|z\|_2^2/n\right)}{\left(E\left(Q_2 - \|z\|_2^2/n\right)\right)^2} \leq \frac{105\left(105C_2^4 + 60\right)}{n(3C_2^2 - 2)^2}$$

Since $C_3 > 1$ and $n \geq 3$, we know $C_3 \log nZ \geq Z - 1$ with probability 1. Suppose $i_0 = \arg\max_i \beta^{(i)T}\Sigma\beta^{(i)}$, then

$$\begin{aligned} &P\left(Q_3 \leq \|z\|_\infty\right) \\ \leq &P\left(\max_i\left(C_3 \log n(\beta^{(i)T}\Sigma\beta^{(i)})Z_i\right) \leq \max_i\left((\beta^{(i)T}\Sigma\beta^{(i)})(Z_i - 1)\right)\right) \\ &+ P\left(\max_i\left(C_3 \log n(\beta^{(i)T}\Sigma\beta^{(i)})Z_i\right) \leq \max_i\left((\beta^{(i)T}\Sigma\beta^{(i)})(1 - Z_i)\right)\right) \\ \leq &0 + P\left(\max_i\left(C_3 \log n(\beta^{(i)T}\Sigma\beta^{(i)})Z_i\right) \leq \max_i\left(\beta^{(i)T}\Sigma\beta^{(i)}\right)\right) \\ \leq &P\left(C_3 \log n\beta^{(i_0)T}\Sigma\beta^{(i_0)}Z_{i_0} \leq \beta^{(i_0)T}\Sigma\beta^{(i_0)}\right) \\ \leq &P\left(Z_{i_0} \leq \frac{1}{C_3 \log n}\right) = P\left(|N(0,1)| \leq \frac{1}{\sqrt{C_3 \log n}}\right) \leq \frac{2}{\sqrt{2\pi C_3 \log n}} \end{aligned}$$

- Then we consider the latter part of (7.8)-(7.10). We can do similar calculations as the first part of the proof and get

$$EQ_1 = C_1 E\xi_i^2 = C_1 E(\beta^{(i)T}\Sigma\beta^{(i)}) = C_1\|\Sigma\|_*$$

$$\mathrm{Var}(Q_1) = \frac{C_1^2}{n}\mathrm{Var}\xi_i^2 \le \frac{C_1^2}{n}E\xi_i^4 = \frac{C_1^2}{n}E\left(\beta^{(i)T}\Sigma\beta^{(i)}\right)^2 EZ^2 \le \frac{9C_1^2}{n}\|\Sigma\|_*^2$$

$$EQ_2 = C_2^2 EZ^2 \cdot E\left(\beta^{(i)T}\Sigma\beta^{(i)}\right)^2 \le 9C_2^2\|\Sigma\|_*^2$$

$$\begin{aligned}\mathrm{Var}(Q_2) &= \frac{1}{n}\mathrm{Var}\left(C_2^2\xi_i^4\right) \le \frac{C_2^4}{n}E\xi_i^8 \\ &= \frac{C_2^4}{n}EZ^4 \cdot E\left(\beta^{(i)T}\Sigma\beta^{(i)}\right)^4 \le \frac{105^2 C_2^4}{n}\|\Sigma\|_*^4\end{aligned}$$

So by Chebyshev's inequality,

$$P\left(Q_1 \ge M_1C_1\|\Sigma\|_*\right) \le P\left(Q_1 - EQ_1 \ge (M_1-1)C_1\|\Sigma\|_*\right) \le \frac{9}{(M_1-1)^2 n}$$

$$\begin{aligned}&P\left(Q_2 \ge M_2C_2^2\|\Sigma\|_*^2\right) \le P\left(Q_2 - EQ_2 \ge (M_2-9)C_2^2\|\Sigma\|_*^2\right) \\ \le&\frac{\mathrm{Var}(Q_2)}{(M_2-9)^2C_2^4\|\Sigma\|_*^4} \le \frac{105^2}{n(M_1-9)^2}\end{aligned}$$

which provide the latter part of (7.8) and (7.9). Finally we note that $\xi_i^2 = (\beta^{(i)T}\Sigma\beta^{(i)})Z_i$. By Lemma 1 in [28] and the fact that $\|\Sigma\|_* = \sum_i \lambda_i(\Sigma)$, $\|\Sigma\| = \max_i \lambda_i(\Sigma)$, $\|\Sigma\|_F = \sqrt{\sum_i \lambda_i^2(\Sigma)}$, $\|\Sigma\|_F \le \sqrt{\|\Sigma\|_* \cdot \|\Sigma\|}$, we have

(0.58)

$$\begin{aligned}&P\left(\beta^T\Sigma\beta \ge \left(\sqrt{\|\Sigma\|_*} + \sqrt{2M_3\log n\|\Sigma\|}\right)^2\right) \\ =&P\left(\sum_{i=1}^n \lambda_i(\Sigma)X_i^2 \ge \|\Sigma\|_* + 2\sqrt{2M_3\log n\|\Sigma\|_*\|\Sigma\|} + 2M_3\log n\|\Sigma\|\right) \\ \le&P\left(\sum_{i=1}^n \lambda_i(\Sigma)(X_i^2-1) \ge 2\sqrt{2M_3\log n\sum_{i=1}^n\lambda_i^2(\Sigma) + 2M_3\log n\max_i\lambda_i(\Sigma)}\right) \\ \le&n^{-M_3}\end{aligned}$$

and

$$P\left(Z \ge 2M_3\log n\right) \le \exp(-2M_3\log n/2) = n^{-M_3}.$$

Hence,

$$P\left(C_3 \log n \xi_i^2 \geq 2C_3 M_3 \log^2 n \left(\sqrt{\|\Sigma\|_*} + \sqrt{2M_3 \log n \|\Sigma\|}\right)^2\right) \leq 2n^{-M_3},$$

and consequently

$$P\left(C_3 \log n \max_{1\leq i\leq n} \xi_i^2 \geq 2C_3 M_3 \log^2 n \left(\sqrt{\|\Sigma\|_*} + \sqrt{2M_3 \log n \|\Sigma\|}\right)^2\right) \leq 2n^{-M_3+1},$$

which gives the right side of (7.10). □

Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA, 19104.
E-mail: tcai@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~tcai/
E-mail: anrzhang@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~anrzhang/