SUPPLEMENT TO "RATE-OPTIMAL POSTERIOR CONTRACTION FOR SPARSE PCA"

By Chao Gao and Harrison H. Zhou

Yale University

In this text, we present proofs of Proposition 2.1, Lemma 5.1, Lemma 5.8, Lemma 5.10, Theorem 4.2, Theorem 4.3, Proposition 5.1 and Lemma 5.9 in Gao and Zhou (2013).

APPENDIX A: PROOF OF PROPOSITION 2.1

Define the concentration set $H_n = \left\{||VV^T - V_0V_0^T||_F^2 \leq M\epsilon^2\right\}$. Then, by Jensen's inequality, we have

$$\begin{aligned}
& P_\Sigma^n \left\|\mathbb{E}_\Pi\big(VV^T|X^n\big) - V_0V_0^T\right\|_F^2 \\
\leq\ & P_\Sigma^n\mathbb{E}_\Pi\Big(||VV^T - V_0V_0^T||_F^2\Big|X^n\Big) \\
=\ & P_\Sigma^n\mathbb{E}_\Pi\Big(||VV^T - V_0V_0^T||_F^2\mathbb{I}_{H_n}\Big|X^n\Big) + P_\Sigma^n\mathbb{E}_\Pi\Big(||VV^T - V_0V_0^T||_F^2\mathbb{I}_{H_n^c}\Big|X^n\Big) \\
\leq\ & M\epsilon^2 + \sup_V\Big(||VV^T - V_0V_0^T||_F^2\Big)P_\Sigma^n\Pi\big(H_n^c|X^n\big) \\
\leq\ & M\epsilon^2 + 2(p+r)\delta,
\end{aligned}$$

where $\sup_V\Big(||VV^T - V_0V_0^T||_F^2\Big) \leq 2(p+r)$ because V and V_0 are unitary matrices. Take $\sup_{\Sigma\in\mathcal{G}(p,s,r)}$ on both sides of the inequality, the proof is complete.

APPENDIX B: PROOF OF LEMMA 5.1

We renormalize the prior Π as $\tilde{\Pi} = \Pi(K_n)^{-1}\Pi$ so that $\tilde{\Pi}$ is a distribution with support within K_n. Write $\mathbb{E}_{\tilde{\Pi}}$ to be the expectation using probability $\tilde{\Pi}$. We define the random variable

$$Y_i = \int \log\frac{dP_\Gamma}{dP_\Sigma}(X_i)d\tilde{\Pi}(\Gamma) = c + \frac{1}{2}X_i^T\Big(\Sigma^{-1} - \mathbb{E}_{\tilde{\Pi}}\Gamma^{-1}\Big)X_i, \quad i = 1, ..., n.$$

Then, Y_i is a sub-exponential random variable with mean

$$\begin{aligned} -P_\Sigma Y_i &= \int D(P_\Sigma || P_\Gamma) d\tilde{\Pi}(\Gamma) \\ &= \int \left(-\frac{1}{2}\log\det\left(\Gamma^{-1/2}\Sigma\Gamma^{-1/2}\right) + \frac{1}{2}\mathrm{tr}\left(\Gamma^{-1/2}\Sigma\Gamma^{-1/2} - I\right)\right) d\tilde{\Pi}(\Gamma) \\ &\leq \frac{1}{4}\int ||\Gamma^{-1/2}\Sigma\Gamma^{-1/2} - I||_F^2 d\tilde{\Gamma} \leq \frac{1}{4}\int \frac{||\Gamma - \Sigma||_F^2}{\lambda_{\min}(\Gamma)^2} d\tilde{\Pi}(\Gamma) \\ &\leq \epsilon^2/4. \end{aligned}$$

Therefore, by Jensen's inequality, we have

$$\begin{aligned} & P_\Sigma^n\left(\int \frac{dP_\Gamma^n}{dP_\Sigma^n}(X^n) d\tilde{\Pi}(\Gamma) \leq \exp\Big(-(b+1)n\epsilon^2\Big)\right) \\ \leq\ & P_\Sigma^n\left(\frac{1}{n}\sum_{i=1}^n Y_i \leq -(b+1)\epsilon^2\right) \\ \leq\ & P_\Sigma^n\left(\frac{1}{n}\sum_{i=1}^n (Y_i - P_\Sigma Y_i) \leq -b\epsilon^2\right). \end{aligned}$$

Define Z_i through the relation $X_i = \Sigma^{1/2} Z_i$, so that $Z_1, ..., Z_n$ are i.i.d. drawn from $N(0, I)$ under P_Σ. Then Y_i can be written as

$$Y_i = c + \frac{1}{2} Z_i^T \Big(I - \mathbb{E}_{\tilde{\Pi}} \Sigma^{1/2}\Gamma^{-1}\Sigma^{1/2}\Big) Z_i.$$

Applying eigenvalue decomposition, we have

$$I - \mathbb{E}_{\tilde{\Pi}}\Sigma^{1/2}\Gamma^{-1}\Sigma^{1/2} = UDU^T,$$

where $D = \mathrm{diag}(d_1, ..., d_p)$. Denote $\tilde{Z}_i = U^T Z_i$, it is easy to see that $\tilde{Z}_i \sim N(0, 1)$ under P_Σ. Hence,

$$\begin{aligned} & P_\Sigma^n\left(\frac{1}{n}\sum_{i=1}^n (Y_i - P_\Sigma Y_i) \leq -b\epsilon^2\right) \\ =\ & \mathbb{P}\left(\sum_{i=1}^n \sum_{j=1}^p \Big(d_j \tilde{Z}_{ij}^2 - \mathbb{E} d_j \tilde{Z}_{ij}^2\Big) \leq -2bn\epsilon^2\right) \\ \leq\ & \exp\left(-C\min\left(\frac{4b^2n^2\epsilon^4}{n\sum_{j=1}^p d_j^2}, \frac{2bn\epsilon^2}{\max_j d_j}\right)\right), \end{aligned}$$

by Bernstein's inequality (Proposition 5.16 of Vershynin (2010)). Note that

$$\begin{aligned}\sum_{j=1}^{p} d_j^2 &= ||I - \mathbb{E}_{\tilde{\Pi}}\Sigma^{1/2}\Gamma^{-1}\Sigma^{1/2}||_F^2 \\ &\leq \mathbb{E}_{\tilde{\Pi}}||I - \Sigma^{1/2}\Gamma^{-1}\Sigma^{1/2}||_F^2 \\ &\leq K\mathbb{E}_{\tilde{\Pi}}\frac{||\Gamma - \Sigma||_F^2}{\lambda_{\min}(\Gamma)^2} \\ &\leq K\epsilon^2.\end{aligned}$$

By the fact that $\epsilon \to 0$, we have

$$P_\Sigma^n\left(\int \frac{dP_\Gamma^n}{dP_\Sigma^n}(X_i)d\tilde{\Pi}(\Gamma) \leq \exp\Big(-(b+1)n\epsilon^2\Big)\right) \leq \exp\Big(-4C_2b^2K^{-1}n\epsilon^2\Big).$$

The conclusion follows the fact that

$$\begin{aligned}&P_\Sigma^n\left(\int \frac{dP_\Gamma^n}{dP_\Sigma^n}(X_i)d\Pi(\Gamma) \leq \Pi(K_n)\exp\Big(-(b+1)n\epsilon^2\Big)\right) \\ \leq\ &P_\Sigma^n\left(\int \frac{dP_\Gamma^n}{dP_\Sigma^n}(X_i)d\tilde{\Pi}(\Gamma) \leq \exp\Big(-(b+1)n\epsilon^2\Big)\right).\end{aligned}$$

APPENDIX C: PROOF OF LEMMA 5.8

By the definition of spectral norm, we have

$$||\hat{\Sigma} - \bar{\Sigma}|| = \sup_{v \in S^{d-1}} v^T\big(\hat{\Sigma} - \bar{\Sigma}\big)v,$$

where S^{d-1} is the $d-1$-dimensional unit sphere. Let $S_{1/2}^{d-1}$ be a $1/2$ net of S^{d-1}. With the same calculation as in the proof of Lemma 3 in Cai, Zhang and Zhou (2010), we have

$$||\hat{\Sigma} - \bar{\Sigma}|| \leq 4 \sup_{v \in S_{1/2}^{d-1}} v^T\big(\hat{\Sigma} - \bar{\Sigma}\big)v,$$

and $|S_{1/2}^{d-1}| \leq 5^d$. Hence,

$$
\begin{aligned}
P_{\bar{\Sigma}}^n\Big(||\hat{\Sigma}-\bar{\Sigma}|| > t||\bar{\Sigma}||\Big) &\leq P_{\bar{\Sigma}}^n\Big(4\sup_{v\in S_{1/2}^{d-1}} v^T\big(\hat{\Sigma}-\bar{\Sigma}\big)v > t||\bar{\Sigma}||\Big) \\
&\leq \bigcup_{v\in S_{1/2}^{d-1}} P_{\bar{\Sigma}}^n\Big(v^T\big(\hat{\Sigma}-\bar{\Sigma}\big)v > t||\bar{\Sigma}||/4\Big) \\
&\leq \bigcup_{v\in S_{1/2}^{d-1}} \mathbb{P}\left(v^T\bar{\Sigma}v\left|\frac{1}{n}\sum_{i=1}^n Z_i^2 - 1\right| > t||\bar{\Sigma}||/4\right) \\
&\leq |S_{1/2}^{d-1}|\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i^2 - 1\right| > t/4\right) \\
&\leq \exp\Big(-C_3\big(-d+n(t\wedge t^2)\big)\Big),
\end{aligned}
$$

where $Z_1, ..., Z_i$ are i.i.d. $N(0,1)$ variables. The proof is complete.

APPENDIX D: PROOF OF LEMMA 5.10

We are going to derive an upper bound for the following metric entropy

$$
\log N\Big(R_1\epsilon, \{V : d_I(U,V) \leq R_2\epsilon\}, d_\Lambda\Big).
$$

We first prove a technical lemma, and then prove the main bound.

LEMMA D.1. *For any $U, V \in \mathcal{U}(d,r)$ with $d \geq r$, and $\Lambda = diag(\lambda_1, ..., \lambda_r)$ with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r$, we have*

$$
d_\Lambda(U,V) \leq 2\lambda_1||U-V||_F, \quad \textit{and} \quad \inf_{P,Q\in\mathcal{U}(r,r)} ||UP-VQ||_F \leq d_I(U,V).
$$

Proof. The first inequality is because

$$
\begin{aligned}
d_\Lambda(U,V) &\leq ||U\Lambda U^T - U\Lambda V^T||_F + ||U\Lambda V^T - V\Lambda V^T||_F \\
&\leq \Big(||U\Lambda|| + ||V\Lambda||\Big)||U-V||_F \\
&\leq 2\lambda_1||U-V||_F.
\end{aligned}
$$

Now we prove the second part. Choosing $P, Q \in \mathcal{U}(r,r)$ satisfying

$$
P^T U^T V Q = \Gamma = \text{diag}(\gamma_1, ..., \gamma_r).
$$

the left hand side of the above equation can be written as

$$\begin{aligned} ||UU^T - VV^T||_F^2 &= ||UPP^TU^T - VQQ^TV^T||_F^2 \\ &= 2\text{tr}\Big(I_{r\times r} - P^TU^TVQQ^TV^TUP\Big) \\ &= 2\text{tr}\Big(I_{r\times r} - \Gamma^2\Big) \\ &= 2\sum_{l=1}^r (1-\gamma_l^2). \end{aligned}$$

For the same P, Q, we also have

$$\begin{aligned} ||UP - VQ||_F^2 &= 2\text{tr}\Big(I_{r\times r} - P^TU^TVQ\Big) \\ &= 2\text{tr}\Big(I_{r\times r} - \Gamma\Big) \\ &= 2\sum_{l=1}^r (1-\gamma_l). \end{aligned}$$

Since $\max_{1\leq l\leq r}\gamma_l = ||\Gamma|| = ||P^TU^TVQ|| \leq 1$, we have

$$\sum_{l=1}^r (1-\gamma_l^2) = \sum_{l=1}^r (1-\gamma_l)(1+\gamma_l) \geq \sum_{l=1}^r (1-\gamma_l).$$

Therefore,

$$\inf_{P,Q\in\mathcal{U}(r,r)} ||UP - VQ||_F \leq ||UP - VQ||_F \leq ||UU^T - VV^T||_F.$$

∎

Proof of Lemma 5.10. Define $\rho_1(U,V) = \inf_{P,Q\in\mathcal{U}(r,r)} ||UP - VQ||_F$ and $\rho_2(U,V) = ||U-V||_F$. Then by Lemma D.1, we have

$$\rho_1(U,V) \leq d_I(U,V), \quad d_\Lambda(U,V) \leq 2\lambda_1\rho_2(U,V).$$

Therefore,

$$\begin{aligned} & N\Big(R_1\epsilon, \{V : d_I(U,V) \leq R_2\epsilon\}, d_\Lambda\Big) \\ \leq\ & N\Big((2\lambda_1)^{-1}R_1\epsilon, \{V : \rho_1(U,V) \leq R_2\epsilon\}, \rho_2\Big). \end{aligned}$$

According to the definition of ρ_1, we have

$$\{V : \rho_1(U,V) \leq R_2\epsilon\} = \bigcup_{Q\in\mathcal{U}(r,r)} \{V : ||V - UQ||_F \leq R_2\epsilon\}.$$

We first cover $\mathcal{U}(r,r)$ by $\{Q_1,...,Q_M\}\subset\mathcal{U}(r,r)$ with norm $||\cdot||_F$. Since

$$\mathcal{U}(r,r)\subset\left\{U\in\mathcal{U}(r,r):||U||_F\leq\sqrt{r}\right\},$$

the bound of M is determined by

$$\log N\Big(\epsilon,\mathcal{U}(r,r),||\cdot||_F\Big)\leq r^2\log\left(\frac{6\sqrt{r}}{\epsilon}\right).$$

Therefore, for any $Q\in\mathcal{U}(r,r)$, there exists $Q_j\in\{Q_1,...,Q_M\}$, such that

$$||V-UQ_j||_F\leq||V-UQ||_F+||U(Q-Q_j)||_F\leq||V-UQ||_F+\epsilon.$$

Hence,

$$\{V:\rho_1(U,V)\leq R_2\epsilon\}\subset\bigcup_{j=1}^M\{V:||V-UQ_j||_F\leq(R_2+1)\epsilon\}.$$

Let us cover the right hand side. Consider UQ_1. Then, there exists $\{\bar{W}_1,...,\bar{W}_N\}\subset\mathcal{U}(d,r)$, with $\log N\leq dr\log\left(\frac{6(R_2+1)}{\eta}\right)$, such that

$$\{V:||V-UQ_1||_F\leq(R_2+1)\epsilon\}\subset\bigcup_{i=1}^N\left\{V:||V-\bar{W}_i||_F\leq\eta\right\}.$$

Define $W_i=\bar{W}_iQ_1^T$ for $i=1,...,N$. Then

$$\{V:||V-UQ_1||_F\leq(R_2+1)\epsilon\}\subset\bigcup_{i=1}^N\left\{V:||V-W_iQ_1||_F\leq\eta\right\}.$$

Now consider any $j\in\{1,2,...,M\}$, we have

$$\begin{aligned}
& \{V:||V-UQ_j||_F\leq(R_2+1)\epsilon\}\\
= & \{V:||VQ_j^TQ_1-UQ_1||_F\leq(R_2+1)\epsilon\}\\
\subset & \bigcup_{j=1}^N\left\{V:||VQ_j^TQ_1-W_iQ_1||_F\leq\eta\right\}\\
= & \bigcup_{i=1}^N\{V:||V-W_iQ_j||_F\leq\eta\}.
\end{aligned}$$

Taking union over j, we have

$$\begin{aligned}
&\bigcup_{j=1}^{M}\{V : ||V - UQ_j||_F \le (R_2+1)\epsilon\} \\
\subset\ &\bigcup_{j=1}^{M}\bigcup_{i=1}^{N}\{V : ||V - W_iQ_j||_F \le \eta\} \\
=\ &\bigcup_{j=1}^{M}\bigcup_{i=1}^{N}\{V : \rho_2(V, W_iQ_j) \le \eta\},
\end{aligned}$$

which implies

$$\{V : \rho_1(U,V) \le R_2\epsilon\} \subset \bigcup_{j=1}^{M}\bigcup_{i=1}^{N}\{V : \rho_2(V, W_iQ_j) \le \eta\}.$$

We may pick η to be $\eta = (2\lambda_1)^{-1}R_1$. Since $W_i \in \mathcal{U}(d,r)$ and $Q_j \in \mathcal{U}(r,r)$, we have $W_iQ_j \in \mathcal{U}(d,r)$, and thus $\{W_iQ_j\}_{1\le i\le N, 1\le j\le M}$ is the covering set. The metric entropy is bounded by

$$\log N + \log M \le dr \log\left(\frac{12\lambda_1(R_2+1)}{R_1}\right) + r^2 \log\frac{6\sqrt{r}}{\epsilon}.$$

The proof is complete. ∎

APPENDIX E: PROOF OF THEOREM 4.2 AND THEOREM 4.3

The proofs of Theorem 4.2 and 4.3 are almost the same as the proof of Theorem 4.1. We only state the proof for Theorem 4.2. The proof of Theorem 4.3 will be sketched in the end of the section. Since we use a different prior, we need two new lemmas to replace Lemma 5.2 and Lemma 5.6.

Lemma E.1. *For any $A > 0$, we have $\Pi\big(|S| > As\big) \le 4\exp\Big(-\frac{\kappa A}{2}s\log p\Big)$.*

Proof. We write $\pi(q) = N_{\kappa,p}^{-1}\exp\Big(-\kappa q\log p\Big)$, where $N_{\kappa,p} = \sum_{q=1}^{p}\exp\Big(-\kappa q\log p\Big)$. For sufficiently large p, we have

$$\frac{1}{2}p^{-\kappa} \le N_{\kappa,p} \le 2p^{-\kappa}.$$

Therefore,

$$\Pi\Big(|S| > As\Big) \leq \sum_{q=[As]}^{p} \pi(q) \leq 2p^{\kappa} \sum_{q=[As]}^{p} \exp\big(-\kappa q \log p\big) \leq 4 \exp\Big(-\frac{\kappa A}{2} s \log p\Big).$$

∎

LEMMA E.2. *As long as $\epsilon \to 0$ and $n \leq p^m$ for some constant $m > 0$, we have* $\Pi\left(\frac{||\Gamma-\Sigma||_F}{\lambda_{\min}(\Gamma)} \leq \epsilon\right) \geq \frac{1}{2} \exp\Big(-(2m+\kappa+2)n\epsilon^2\Big)$.

Proof. The proof is similar to the proof of Lemma 5.2. Notice $\lambda_{\min}(\Gamma) = 1$, and we have

$$\Pi\left(\frac{||\Gamma - \Sigma||_F}{\lambda_{\min}(\Gamma)} \leq \epsilon\right) = \Pi\Big(||\Gamma - \Sigma||_F \leq \epsilon\Big). \tag{E.1}$$

Using conditional argument, we have

$$\Pi\Big(||\Gamma - \Sigma||_F \leq \epsilon\Big) \geq \Pi\Big(||\Gamma - \Sigma||_F \leq \epsilon|(q,S) = (s,S_0)\Big)\Pi\Big((q,S) = (s,S_0)\Big).$$

When $(q,S) = (s,S_0)$, we have $||\Gamma - \Sigma||_F = ||\eta\eta^T - \theta\theta^T||_F = ||\eta_{S_0}\eta_{S_0}^T - \theta_{S_0}\theta_{S_0}^T||_F$. Thus, the first term in the product is

$$\Pi\Big(||\Gamma - \Sigma||_F \leq \epsilon|(q,S) = (s,S_0)\Big) = \Pi\Big(||\eta_{S_0}\eta_{S_0}^T - \theta_{S_0}\theta_{S_0}^T||_F \leq \epsilon\Big).$$

Suppose $||\eta_{S_0} - \theta_{S_0}|| \leq (3K^{1/2})^{-1}\epsilon$, then we have

$$\begin{aligned}
||\eta_{S_0}\eta_{S_0}^T - \theta_{S_0}\theta_{S_0}^T||_F &= ||\eta_{S_0}\eta_{S_0}^T - \eta_{S_0}\theta_{S_0}^T + \eta_{S_0}\theta_{S_0}^T - \theta_{S_0}\theta_{S_0}^T||_F \\
&\leq \Big(||\theta_{S_0}|| + ||\eta_{S_0}||\Big)||\eta_{S_0} - \theta_{S_0}|| \\
&\leq \Big(2||\theta_{S_0}|| + ||\eta_{S_0} - \theta_{S_0}||\Big)||\eta_{S_0} - \theta_{S_0}|| \\
&\leq \Big(2K^{1/2} + (3K^{1/2})^{-1}\epsilon\Big)(3K^{1/2})^{-1}\epsilon \\
&\leq \epsilon.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pi\Big(||\eta_{S_0}\eta_{S_0}^T - \theta_{S_0}\theta_{S_0}^T||_F \le \epsilon\Big) &\ge \Pi\Big(||\eta_{S_0} - \theta_{S_0}|| \le (3K^{1/2})^{-1}\epsilon\Big) \\
&\ge \exp\Bigg(- \frac{1}{2}||\theta||^2 - s\log\frac{1}{\epsilon} - s\log\big(2\sqrt{s}/3\big)\Bigg) \\
&\ge \exp\Bigg(- \frac{1}{2}\Big(K + s\log n + s\log s\Big)\Bigg) \\
&\ge \exp\Big(- 2ms\log p\Big)
\end{aligned}$$

by Lemma F.1 and the assumption $n \le p^m$. We also have

$$\Pi\Big((q,S) = (s,S_0)\Big) = \pi(s)\frac{1}{\binom{p}{s}} \ge \frac{1}{2}\exp\Big(- (\kappa+2)s\log p\Big).$$

Hence, $\Pi\Big(||\Gamma - \Sigma||_F \le \epsilon\Big) \ge \frac{1}{2}\exp\Big(- (2m+\kappa+2)n\epsilon^2\Big)$. ∎

Proof of Theorem 4.2. Using the same method in the proof of Theorem 4.2 by Combining Lemma 5.1, Lemma E.2, Lemma E.1 and Lemma 5.4, we have

$$P_\Sigma^n \Pi\Big(||\Gamma - \Sigma|| > M'\epsilon | X^n\Big) \le \exp\Big(- Cn\epsilon^2\Big).$$

As long as $||\Gamma - \Sigma|| \le M'\epsilon$, we have $\big|||\eta||^2 - ||\theta||^2\big| \le M'\epsilon$ by Weyl's theorem. We also have $||\Gamma - \Sigma||_F \le \sqrt{2}M'\epsilon$ because $\Gamma - \Sigma = \eta\eta^T - \theta\theta^T$ is a rank-two matrix. By sin-theta theorem (Lemma **??**), $\Big\|\frac{\eta\eta^T}{||\eta||^2} - \frac{\theta\theta^T}{||\theta||^2}\Big\|_F \le \sqrt{2}KM'\epsilon$. According to Proposition 2.2 in Vu and Lei (2013),

$$\min\left\{\left\|\frac{\eta}{||\eta||} - \frac{\theta}{||\theta||}\right\|, \left\|\frac{\eta}{||\eta||} + \frac{\theta}{||\theta||}\right\|\right\} \le 2KM'\epsilon.$$

Therefore,

$$\begin{aligned}
||\eta - \theta|| &= \left\|\eta - \frac{\eta}{||\eta||}||\theta|| + \frac{\eta}{||\eta||}||\theta|| - \theta\right\| \\
&\le \big|||\eta|| - ||\theta||\big| + ||\theta||\left\|\frac{\eta}{||\eta||} - \frac{\theta}{||\theta||}\right\| \\
&= \frac{\big|||\eta||^2 - ||\theta||^2\big|}{||\eta|| + ||\theta||} + ||\theta||\left\|\frac{\eta}{||\eta||} - \frac{\theta}{||\theta||}\right\| \\
&\le (KM' + 2K^2M')\epsilon,
\end{aligned}$$

as long as $\left\|\frac{\eta}{||\eta||}-\frac{\theta}{||\theta||}\right\| \leq 2KM'\epsilon$. The same argument also works for $||\eta+\theta||$. Therefore, we have

$$||\eta-\theta|| \wedge ||\eta+\theta|| \leq (KM'+2K^2M')\epsilon.$$

Hence, we have

$$P_\Sigma^n \Pi\Big(||\eta-\theta|| \wedge ||\eta+\theta|| > M'\epsilon|X^n\Big) \leq \exp\Big(-Cn\epsilon^2\Big).$$

∎

Proof of Theorem 4.3. The only modification needed is to establish

$$\Pi\Big(\left\|A_{S_0}A_{S_0}^T - A_{0,S_0}A_{0,S_0}^T\right\|_F \leq \epsilon\Big) \geq \exp\Big(-Cs\log p\Big),$$

where $S_0 = \cup_{l=1}^r S_{0l}$, for some constant $C > 0$. This can be done similarly as in the proof of Lemma E.2. Then, combining this result with Lemma 5.1, Lemma E.1 and Lemma 5.4, we have obtained (E.1). In view of the inequality

$$||VV^T - V_0V_0^T||_F \leq C\sqrt{r}||\Gamma-\Sigma||,$$

for some constant $C > 0$, the proof is complete. ∎

APPENDIX F: PROOF OF PROPOSITION 5.1

We first present a lemma on Gaussian small ball probability.

LEMMA F.1. *For $Z \sim N(0, I_d)$ and any $\theta \in \mathbb{R}^d$, we have*

$$\mathbb{P}\Big(||Z-\theta|| \leq \epsilon\Big) \geq \exp\left(-\frac{1}{2}||\theta||^2 - d\log\frac{1}{\epsilon} - d\log\left(2\sqrt{d}/3\right)\right),$$

for any $\epsilon < 1/2$.

Proof. By Theorem 3.1 in Li and Shao (2001), we have

$$\mathbb{P}\Big(||Z-\theta|| \leq \epsilon\Big) \geq \exp\left(-||\theta||^2/2\right)\mathbb{P}\Big(||Z|| \leq \epsilon\Big).$$

For the centered small ball probability, we have

$$\begin{aligned}
\mathbb{P}\Big(||Z|| \leq \epsilon\Big) &\geq \prod_{i=1}^d \mathbb{P}\Big(Z_i^2 \leq \epsilon^2/d\Big) = \left(\int_{|z|\leq \epsilon/\sqrt{d}} (2\pi)^{-1/2} e^{-z^2/2} dz\right)^d \\
&\geq \left(\frac{2\epsilon}{\sqrt{d}}(2\pi)^{-1} e^{-\epsilon^2/2d}\right)^d \geq \left(\frac{2\epsilon}{3\sqrt{d}}\right)^d \\
&= \exp\left(-d\log\frac{1}{\epsilon} - d\log\left(2\sqrt{d}/3\right)\right).
\end{aligned}$$

∎

Proof of Proposition 5.1. We are going to lower bound $G(\mathcal{T}_l|\mathcal{T}_{l-1})$. We use the following notation

$$(u_1, ..., u_l, u_{l+1}) = (\eta_{1,S_{0,l+1}}, ..., \eta_{l,S_{0,l+1}}, \eta_{l+1,S_{0,l+1}}),$$

$$(v_1, ..., v_l, v_{l+1}) = (\theta_{1,S_{0,l+1}}, ..., \theta_{l,S_{0,l+1}}, \theta_{l+1,S_{0,l+1}}).$$

Define the projection matrix

$$H_l = \sum_{i=1}^{l} \frac{u_i u_i^T}{||u_i||^2}.$$

We also define $\tilde{u}_{l+1} = (I - H_l)u_{l+1}$ and $\tilde{v}_{l+1} = (1 - H_l)v_{l+1}$. By definition of the prior, we have $u_{l+1} = \tilde{u}_{l+1}$. We have

$$\begin{aligned}
||\eta_{l+1,S_{0,l+1}} - \theta_{l+1,S_{0,l+1}}|| &= ||\tilde{u}_{l+1} - \tilde{v}_{l+1} - H_l v_{l+1}|| \\
&\leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sum_{i=1}^{l} |u_i^T v_{l+1}| \left\| \frac{u_i}{||u_i||^2} \right\| \\
&\leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sum_{i=1}^{l} \frac{|(u_i - v_i)^T v_l|}{||u_i||} \\
&\leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sum_{i=1}^{l} \frac{||v_l||}{||u_i||} ||u_i - v_i|| \\
&\leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sqrt{2}K \sum_{i=1}^{l} ||u_i - v_i|| \\
&\leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sqrt{2}K \sum_{i=1}^{l} ||\eta_{i,S_{0i}} - \theta_{i,S_{0i}}||.
\end{aligned}$$

Conditioning on $\mathcal{T}_l$, we have

$$||\eta_{l+1,S_{0,l+1}} - \theta_{l+1,S_{0,l+1}}|| \leq ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \frac{\sqrt{2}}{\sqrt{2}+1} K^{1/2} \sum_{i=1}^{l} \epsilon_i.$$

Therefore,

$$G(\mathcal{T}_{l+1}|\mathcal{T}_l) \geq G^*_{|S_{0,l+1}|-l^*}\left((\sqrt{2}+1)K^{1/2} ||\tilde{u}_{l+1} - \tilde{v}_{l+1}|| + \sqrt{2}K \sum_{i=1}^{l} \epsilon_i \leq \epsilon_{l+1} \right).$$

Remember the sequence $\{\epsilon_i\}_{i=1}^r$ satisfies

$$K\sum_{i=1}^{l}\epsilon_i \leq \frac{1}{2}\epsilon_{l+1}, \quad \text{and} \quad \sum_{i=1}^{r}\epsilon_i \leq \epsilon.$$

Thus,

$$\begin{aligned}
G(\mathcal{T}_{l+1}|\mathcal{T}_l) &\geq G^*_{|S_{0,l+1}|-l^*}\Bigg((\sqrt{2}+1)K^{1/2}||\tilde{u}_{l+1}-\tilde{v}_{l+1}|| \leq \frac{1}{2}\epsilon_{l+1}\Bigg) \\
&= \mathbb{P}\Bigg(\left\|\frac{U_{l+1}Z_{l+1}}{||Z_{l+1}||} - T_l\tilde{v}_{l+1}\right\| \leq \frac{1}{2(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \\
&\geq \mathbb{P}\Bigg(\left\|\frac{U_{l+1}Z_{l+1}}{||Z_{l+1}||} - Z_{l+1}\right\| + \|Z_{l+1}-T_l\tilde{v}_{l+1}\| \leq \frac{1}{2(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \\
&= \mathbb{P}\Bigg(|U_{l+1}-||Z_{l+1}||| + \|Z_{l+1}-T_l\tilde{v}_{l+1}\| \leq \frac{1}{2(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \\
&\geq \mathbb{P}\Bigg(|U_{l+1}-||Z_{l+1}||| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg| \|Z_{l+1}-T_l\tilde{v}_{l+1}\| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \\
&\quad \times\mathbb{P}\Bigg(\|Z_{l+1}-T_l\tilde{v}_{l+1}\| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg),
\end{aligned}$$

where $Z_{l+1} \sim N(0, I_{|S_{0,l+1}|-l^*})$, and $U_{l+1} \sim \text{Unif}[(2K)^{-1/2}, (2K)^{1/2}]$. By Lemma F.1, we have

$$\begin{aligned}
&\mathbb{P}\Bigg(\|Z_{l+1}-T_l\tilde{v}_{l+1}\| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \\
\geq\ & \exp\big(-||\tilde{v}_{l+1}||^2/2\big)\exp\Bigg(-(s-l^*)\log\frac{4(\sqrt{2}+1)K^{1/2}}{\epsilon_{l+1}} - (s-l^*)\log\big(2\sqrt{s-l^*}/3\big)\Bigg),
\end{aligned}$$

where we have used $||\tilde{v}_{l+1}|| = ||T_l\tilde{v}_{l+1}||$. By the definition of uniform distribution, we have

$$\mathbb{P}\Bigg(|U_{l+1}-||Z_{l+1}||| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg| \|Z_{l+1}-\tilde{v}_{l+1}\| \leq \frac{1}{4(\sqrt{2}+1)K^{1/2}}\epsilon_{l+1}\Bigg) \geq \frac{\epsilon_{l+1}}{2(2+\sqrt{2})K}.$$

Hence, we have

$$G(\mathcal{T}_{l+1}|\mathcal{T}_l) \geq \frac{c(r,\epsilon)}{2(2+\sqrt{2})e^{K/2}}(3\sqrt{2}K)^{l+1}\exp\Bigg(-(s-l^*)\log\frac{(4\sqrt{2}+1)K^{1/2}}{c(r,\epsilon)} - (s-l^*)\log\big(2\sqrt{s-l^*}/3\big)\Bigg),$$

The results follows from the fact $l^* \leq s$. Similarly, $G(\mathcal{U}_1)$ can be lower bounded by the above formula with $l = 0$. ∎

APPENDIX G: PROOF OF LEMMA 5.9

For simplifying the notation, we drop the bar and write $(\Sigma, \Gamma', \Gamma)$ as their low-dimensional counterparts $(\bar{\Sigma}, \bar{\Gamma}', \bar{\Gamma})$. Consider the likelihood ratio test,

$$\phi = \mathbb{I}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i^T\Big(\Sigma^{-1} - \Gamma'^{-1}\Big)Y_i > \log\det\Big(\Sigma^{-1}\Gamma'\Big)\right\}.$$

Define $\rho = \text{tr}\Big(\Gamma'^{-1/2}\Sigma\Gamma'^{-1/2} - I\Big) - \log\det\Big(\Gamma'^{-1/2}\Sigma\Gamma'^{-1/2}\Big)$. Then because of $P_\Sigma Y_i^T(\Sigma^{-1} - \Gamma'^{-1})Y_i = \text{tr}\Big(I - \Gamma'^{-1/2}\Sigma\Gamma'^{-1/2} - I\Big)$, we have

$$\phi = \mathbb{I}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^T\Big(\Sigma^{-1} - \Gamma'^{-1}\Big)Y_i - P_\Sigma Y_i^T\Big(\Sigma^{-1} - \Gamma'^{-1}\Big)Y_i\right) > \rho\right\}.$$

Let $\{l_j\}_{j=1}^d$ be the eigenvalues of the matrix $\Gamma'^{-1/2}\Sigma\Gamma'^{-1/2}$. Since for each j, $l_j \in [(2K)^{-1}, K]$, we have

$$\rho = \sum_{j=1}^{d}\Big(l_j - 1 - \log l_j\Big) \geq \delta_K\sum_{j=1}^{d}(l_j - 1)^2 \geq \delta_K(4K^2)^{-1}||\Sigma - \Gamma'||_F^2,$$

where $\delta_K > 0$ is a constant only depending on K. Let $\{h_j\}_{j=1}^d$ be the eigenvalues of the matrix $\Sigma^{1/2}\Gamma'^{-1}\Sigma^{1/2}$ and write $Y_i = \Sigma^{1/2}\tilde{Z}_i$ so that $\tilde{Z}_i \sim N(0, I)$. Then we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^T\Big(\Sigma^{-1} - \Gamma'^{-1}\Big)Y_i - P_\Sigma Y_i^T\Big(\Sigma^{-1} - \Gamma'^{-1}\Big)Y_i\right) \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{Z}_i^T\Big(I - \Sigma^{1/2}\Gamma'^{-1}\Sigma^{1/2}\Big)\tilde{Z}_i - \mathbb{E}\tilde{Z}_i^T\Big(I - \Sigma^{1/2}\Gamma'^{-1}\Sigma^{1/2}\Big)\tilde{Z}_i\right). \end{aligned}$$

Apply SVD to the matrix $I - \Sigma^{1/2}\Gamma'^{-1}\Sigma^{1/2}$ and we have $I - \Sigma^{1/2}\Gamma'^{-1}\Sigma^{1/2} = U^T(I - H)U$, with $H = \text{diag}(h_1, ..., h_p)$. Define $Z_i = U\tilde{Z}_i \sim N(0, I)$, and the above formula can be written as

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left(Z_i^T(I - H)Z_i - \mathbb{E}Z_i^T(I - H)Z_i\right) \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}(1 - h_j)(Z_{ij}^2 - 1). \end{aligned}$$

where Z_{ij} are i.i.d. $N(0,1)$. Therefore, we have

$$\begin{aligned}
P_\Sigma^n\phi &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^d(1-h_j)(Z_{ij}^2-1)\geq\rho\right)\\
&\leq \mathbb{P}\left(\sum_{i=1}^n\sum_{j=1}^d(1-h_j)(Z_{ij}^2-1)\geq n\delta_K(4K^2)^{-1}||\Sigma-\Gamma'||_F^2\right)\\
&\leq 2\exp\left(-C_5\min\left\{\frac{n^2\delta_K^2(4K^2)^{-2}||\Sigma-\Gamma'||_F^4}{n\sum_{j=1}^d(1-h_j)^2},\frac{n\delta_K(4K^2)^{-1}||\Sigma-\Gamma'||_F^2}{\max_j|1-h_j|}\right\}\right)\\
&\leq 2\exp\left(-C_5\min\left\{\frac{n\delta_K^2(4K^2)^{-2}||\Sigma-\Gamma'||_F^2}{K},\frac{n\delta_K(4K^2)^{-1}||\Sigma-\Gamma'||_F^2}{1+K}\right\}\right)\\
&\leq 2\exp\left(-C_5\delta_K'n||\Sigma-\Gamma'||_F^2\right),
\end{aligned}$$

where we have used Bernstein's inequality (Proposition 5.16 in Vershynin (2010)) with C_5 being an absolute constant and δ_K' only depending on K. Similarly, for any Γ in the alternative set,

$$1-\phi=\mathbb{I}\left\{\frac{1}{n}\sum_{i=1}^n\left(Y_i^T\left(\Gamma'^{-1}-\Sigma^{-1}\right)Y_i-P_\Gamma Y_i^T\left(\Gamma'^{-1}-\Sigma^{-1}\right)Y_i\right)>\bar\rho\right\},$$

where

$$\begin{aligned}
\bar\rho &= \log\det\left(\Sigma\Gamma'^{-1}\right)-\mathrm{tr}\Big(\Gamma(\Gamma'^{-1}-\Sigma^{-1})\Big)\\
&= \log\det\left(\Sigma\Gamma'^{-1}\right)-\mathrm{tr}\Big(\Gamma'(\Gamma'^{-1}-\Sigma^{-1})\Big)+\mathrm{tr}\Big(\left(\Gamma'-\Gamma\right)\left(\Gamma'^{-1}-\Sigma^{-1}\right)\Big)\\
&= \mathrm{tr}\Big(\Sigma^{-1/2}\Gamma'\Sigma^{-1/2}-I\Big)-\log\det\left(\Sigma^{-1/2}\Gamma'\Sigma^{-1/2}\right)+\mathrm{tr}\Big(\left(\Gamma'-\Gamma\right)\left(\Gamma'^{-1}-\Sigma^{-1}\right)\Big)\\
&\geq \delta_K||\Sigma^{-1/2}\Gamma'\Sigma^{-1/2}-I||_F^2-||\Gamma'-\Gamma||_F||\Gamma'^{-1}-\Sigma^{-1}||_F\\
&\geq \delta_K K^{-2}||\Sigma-\Gamma'||_F^2-(2K^2)^{-1}||\Gamma'-\Gamma||_F||\Sigma-\Gamma'||_F.
\end{aligned}$$

Therefore, as long as $||\Gamma'-\Gamma||_F\leq\delta_K||\Sigma-\Gamma'||_F$, we have

$$\bar\rho\geq\frac{1}{2}\delta_K K^{-2}||\Sigma-\Gamma'||_F^2.$$

Similar argument as bounding $P_\Sigma^n\phi$ also gives

$$P_\Gamma^n(1-\phi)\leq 2\exp\left(-C_5\delta_K'n||\Sigma-\Gamma'||_F^2\right).$$

Thus, the proof is complete.

REFERENCES

[1] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics* **38** 2118–2144.

[2] GAO, C. and ZHOU, H. H. (2013). Rate-optimal posterior contraction for sparse PCA.

[3] LI, W. and SHAO, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. *Handbook of Statistics* **19** 533–597.

[4] VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

[5] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics* **41** 2905–2947.

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
E-MAIL: chao.gao@yale.edu
huibin.zhou@yale.edu