The Effect of L_1 Penalization on Condition Number Constrained Estimation of Precision Matrix

Xiao Guo and Chunming Zhang

University of Science and Technology of China and University of Wisconsin-Madison

Supplementary Material

The supplementary material here includes the detailed proofs of Theorems 1–4 and Proposition 1 in the paper.

S1 Detailed proofs

Proof of Theorem 1. Without loss of generality, we assume $\boldsymbol{\mu}_0 = \mathbf{0}$.

Condition A1 indicates that $\mathrm{P}\{|\mathbf{S}_n(i,j) - \boldsymbol{\Sigma}_0(i,j)| \geq \delta\} \leq C\exp(-c\delta^2 n)$ for $i,j = 1,\ldots,p_n$ with an arbitrarily small constant $\delta \in (0,\infty)$ (see, for example, (11) and Lemma A.3 of Bickel and Levina (2008)), and hence

$$\|\widehat{W}^2 - W_0^2\|_2^2 = O_{\mathrm{P}}\{\log(p_n)/n\} = \|\widehat{W}^{-1} - W_0^{-1}\|_2^2. \tag{S1.1}$$

Under Condition A2, we have $\mathrm{P}\{|\mathbf{S}_n(i,j) - \boldsymbol{\Sigma}_0(i,j)| \geq \delta\} \leq Cn^{-\beta/4}\delta^{-\beta/2}$ for $i,j = 1,\ldots,p_n$ with a constant $\delta \in (0,\infty)$ (see, for example, Lemma 2 of Ravikumar et al. (2011)), which implies

$$\|\widehat{W}^2 - W_0^2\|_2^2 = O_{\mathrm{P}}(p_n^{4/\beta}/n) = \|\widehat{W}^{-1} - W_0^{-1}\|_2^2. \tag{S1.2}$$

It's easy to see that Condition A3 implies

$$\|\widehat{W}^2 - W_0^2\|_2^2 = O_{\mathrm{P}}(p_n/n) = \|\widehat{W}^{-1} - W_0^{-1}\|_2^2.$$

Therefore, under either Condition A1 or A2 or A3,

$$\|\widehat{W}^2 - W_0^2\|_2^2 = o_{\mathrm{P}}(1) = \|\widehat{W}^{-1} - W_0^{-1}\|_2^2.$$

To prove $\|\widehat{\boldsymbol{\Theta}}_{\mathrm{prop}-1} - \boldsymbol{\Theta}_0\|_2 \xrightarrow{\mathrm{P}} 0$, it suffices to show that $\|\widetilde{\Omega}_{\kappa_n} - \Omega_0\|_2 \xrightarrow{\mathrm{P}} 0$.

Under Condition A1 or A2, $\boldsymbol{\Sigma}_0$ being diagonal induces $\Gamma_0 = \mathbf{I}_{p_n}$. From (2.2), $\widetilde{\Omega}_{\kappa_n} = \{p_n/\mathrm{tr}(\mathbf{R}_n)\}\mathbf{I}_{p_n} = \mathbf{I}_{p_n}$ due to $\kappa_n = 1$. Hence, $\|\widetilde{\Omega}_{\kappa_n} - \Omega_0\|_2 \xrightarrow{\mathrm{P}} 0$.

Under Condition A3, we first prove $\|\mathbf{S}_n - \mathbf{\Sigma}_0\|_2 \xrightarrow{\text{P}} 0$ for $\mathbf{\Sigma}_0 = \mathbf{I}_{p_n}$. For $i = 1, \ldots, n$, define $\boldsymbol{X}_i^* = (\boldsymbol{X}_i^T, \boldsymbol{Y}_i^T)^T \in \mathbb{R}^{p_n^*}$ with $p_n^* > p_n$ an integer and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \in \mathbb{R}^{p_n^* - p_n}$ i.i.d. random vectors, such that $\{\boldsymbol{e}_{j,p_n^*}^T \boldsymbol{X}_i^* : i = 1, \ldots, n; j = 1, \ldots, p_n^*\}$ are i.i.d. random variables. Let $\mathbf{S}_n^* = n^{-1} \sum_{i=1}^n \boldsymbol{X}_i^* \boldsymbol{X}_i^{*T}$.

From Theorem 2 of Bai and Yin (1993), if $\lim_{n\to\infty} p_n^*/n = y$ with a constant $y \in (0, 1)$, then $\lambda_{\max}(\mathbf{S}_n^*) \xrightarrow{\text{P}} (1 + \sqrt{y})^2$ and $\lambda_{\min}(\mathbf{S}_n^*) \xrightarrow{\text{P}} (1 - \sqrt{y})^2$. We know $\lambda_{\min}(\mathbf{S}_n^*) \le \lambda_{\min}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) \le \lambda_{\max}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) \le \lambda_{\max}(\mathbf{S}_n^*)$. Thus, if y is arbitrarily close to 0, then we have $\lambda_{\max}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) \xrightarrow{\text{P}} 1$ and $\lambda_{\min}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) \xrightarrow{\text{P}} 1$. From

$$\|\mathbf{S}_n - \mathbf{\Sigma}_0\|_2 \le \Big\| n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T - \mathbf{\Sigma}_0 \Big\|_2 + \|\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}^T\|_2 = \text{I} + \text{II},$$

$\text{I} = \max\{|\lambda_{\max}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) - 1|, |\lambda_{\min}(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T) - 1|\} \xrightarrow{\text{P}} 0$ and $\text{II} = \overline{\boldsymbol{X}}^T \overline{\boldsymbol{X}} \xrightarrow{\text{P}} 0$, we have $\|\mathbf{S}_n - \mathbf{\Sigma}_0\|_2 \xrightarrow{\text{P}} 0$.

For $\mathbf{\Sigma}_0$ not necessarily equal to $\mathbf{I}_{p_n}$, $\|\mathbf{S}_n - \mathbf{\Sigma}_0\|_2 \le \|\mathbf{\Sigma}_0^{1/2}\|_2 \|\mathbf{\Sigma}_0^{-1/2} \mathbf{S}_n \mathbf{\Sigma}_0^{-1/2} - \mathbf{I}_{p_n}\|_2 \|\mathbf{\Sigma}_0^{1/2}\|_2 \xrightarrow{\text{P}} 0$, since $\mathbf{\Sigma}_0^{-1/2} \mathbf{S}_n \mathbf{\Sigma}_0^{-1/2}$ is the sample covariance matrix of $\{\mathbf{\Sigma}_0^{-1/2} \boldsymbol{X}_1, \ldots, \mathbf{\Sigma}_0^{-1/2} \boldsymbol{X}_n\}$ which are i.i.d. with covariance matrix $\mathbf{I}_{p_n}$ and

$$\begin{aligned}
&\{E(|\boldsymbol{e}_{1,p_n}^T \mathbf{\Sigma}_0^{-1/2} \boldsymbol{X}_1|^4)\}^{1/4} \le \sum_{i=1}^{p_n} \{E(|\boldsymbol{e}_{1,p_n}^T \mathbf{\Sigma}_0^{-1/2} \boldsymbol{e}_{i,p_n} X_{1,i}|^4)\}^{1/4} \\
\le\ & \max_{1\le i\le p_n} \{E(|X_{1,i}|^4)\}^{1/4} \sum_{i=1}^{p_n} |\boldsymbol{e}_{1,p_n}^T \mathbf{\Sigma}_0^{-1/2} \boldsymbol{e}_{i,p_n}| = \max_{1\le i\le p_n} \{E(|X_{1,i}|^4)\}^{1/4} \|\mathbf{\Sigma}_0^{-1/2}\|_\infty \\
<\ & C < \infty.
\end{aligned}$$

Therefore, $\|\mathbf{R}_n - \Gamma_0\|_2 \xrightarrow{\text{P}} 0$, which implies that $\|\widetilde{\Omega}_{\kappa_n} - \Omega_0\|_2 \xrightarrow{\text{P}} 0$ since $\liminf_{n\to\infty}\{\kappa_n - \text{cond}(\Omega_0)\} > 0$.

The result $\|\widehat{\mathbf{\Theta}}_{\text{prop}-1}^{-1} - \mathbf{\Sigma}_0\|_2 \xrightarrow{\text{P}} 0$ comes from $\|\widehat{\mathbf{\Theta}}_{\text{prop}-1}^{-1} - \mathbf{\Sigma}_0\|_2 = O_{\text{P}}(\|\widehat{\mathbf{\Theta}}_{\text{prop}-1} - \mathbf{\Theta}_0\|_2)$. ■

Proof of Theorem 2. Suppose the eigendecomposition of $\mathbf{R}_n$ is $Q\text{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{p_n})Q^T$, where $\widehat{\lambda}_1 \ge \cdots \ge \widehat{\lambda}_{p_n}$ are the eigenvalues of $\mathbf{R}_n$. From Won et al. (2013), $\widetilde{\Omega}_{\kappa_n}^{-1} = Q\text{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{p_n})Q^T$, where $\widetilde{\lambda}_i = \min\{\max(\tau^*, \widehat{\lambda}_i), \kappa_n \tau^*\}$ with $\tau^* \in (0, \infty)$ depending on $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{p_n}$ and κ_n. Hence, $\widetilde{\Omega}_{\kappa_n}^{-1}$ truncates the eigenvalues of $\mathbf{R}_n$. From Stewart and Sun (1990) (Corollary 4.10, p. 203),

$$\max_{1\le i\le p_n} |\widetilde{\lambda}_i - \lambda_i| \le \|\widetilde{\Omega}_{\kappa_n}^{-1} - \Gamma_0\|_2,$$

where $\lambda_1 \geq \cdots \geq \lambda_{p_n}$ are the eigenvalues of Γ_0. If $\|\widetilde{\Omega}_{\kappa_n}^{-1} - \Gamma_0\|_2 \xrightarrow{\mathrm{P}} 0$, then $\max_{1\leq i\leq p_n} |\widetilde{\lambda}_i - \lambda_i| \xrightarrow{\mathrm{P}} 0$ which implies that $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ converges weakly to F_0 in probability. Therefore, to prove $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 \nrightarrow 0$ in probability, we only need to show $\|\widetilde{\Omega}_{\kappa_n}^{-1} - \Gamma_0\|_2 \nrightarrow 0$ in probability, and it suffices to show that $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ doesn't converge weakly to F_0 in probability.

Under Condition B1, if $\lim_{n\to\infty} p_n/n = \infty$, then the rank of $\mathbf{R}_n$ is at most n when $p_n > n$, and hence, the proportion of the 0 eigenvalues among $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{p_n}$ is at least $(p_n - n)/p_n$ which converges to 1 as $n \to \infty$. Therefore, $\mathbb{F}^{\mathbf{R}_n}$ will converge weakly to $\mathrm{I}_{[0,\infty)}$ in probability. Since $\widetilde{\lambda}_i = \min\{\max(\tau^*, \widehat{\lambda}_i), \kappa_n\tau^*\}$, if $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ converges weakly in probability, then the limit is $\mathrm{I}_{[c,\infty)}$ for some $c \in [0,\infty)$. Since $F_0 \neq \mathrm{I}_{[C,\infty)}$ for any $C \in [0,\infty)$, $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ doesn't converge weakly to F_0 in probability. Therefore, $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 \nrightarrow 0$ in probability.

Under Condition B2, we will show that $|\text{cond}(\widetilde{\Omega}_{\kappa_n}^{-1}) - \text{cond}(\Gamma_0)| \nrightarrow 0$ in probability which implies that $\|\widetilde{\Omega}_{\kappa_n}^{-1} - \Gamma_0\|_2 \nrightarrow 0$ in probability. From Theorem 1 in Won et al. (2013), $\text{cond}(\widetilde{\Omega}_{\kappa_n}^{-1}) = \min\{\kappa_n, \text{cond}(\mathbf{R}_n)\}$. Since $|\min\{\kappa_n, \text{cond}(\mathbf{R}_n)\} - \text{cond}(\Gamma_0)| \nrightarrow 0$ in probability, we have $|\text{cond}(\widetilde{\Omega}_{\kappa_n}^{-1}) - \text{cond}(\Gamma_0)| \nrightarrow 0$ in probability as $n \to \infty$.

Under Condition B3, we will show that $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ does not converge weakly to F_0 in probability. We truncate $\mathbb{F}^{\mathbf{R}_n}$ in order to obtain $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$, i.e., $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}} = \mathbb{F}^{\mathbf{R}_n}\,\mathrm{I}_{[\tau^*,\kappa_n\tau^*)} + \mathrm{I}_{[\kappa_n\tau^*,\infty)}$. If $\mathbb{F}^{\widetilde{\Omega}_{\kappa_n}^{-1}}$ converges weakly to F_0 in probability, then $F_0 = F\,\mathrm{I}_{[l_{\min},l_{\max})} + \mathrm{I}_{[l_{\max},\infty)}$, which contradicts Condition B3.

Therefore, we have demonstrated that $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 \nrightarrow 0$ in probability under either Condition B1 or B2 or B3. Next, we will show that $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1}^{-1} - \boldsymbol{\Sigma}_0\|_2 \nrightarrow 0$ in probability. If $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1}^{-1} - \boldsymbol{\Sigma}_0\|_2 = o_{\mathrm{P}}(1)$, then $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 = o_{\mathrm{P}}(1)$ because $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 = O_{\mathrm{P}}(\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1}^{-1} - \boldsymbol{\Sigma}_0\|_2)$. Since $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1} - \boldsymbol{\Theta}_0\|_2 \nrightarrow 0$ in probability, we can claim $\|\widehat{\boldsymbol{\Theta}}_{\text{prop}-1}^{-1} - \boldsymbol{\Sigma}_0\|_2 \nrightarrow 0$ in probability. ■

Proof of Theorem 3. Following the proofs of Corollaries 1 and 2 in Ravikumar et al. (2011), we have that, with probability tending to 1,

$$|\widehat{\Omega}_{\text{RBLZ}} - \Omega_0|_\infty^2 \leq C r_n^*$$

with $r_n^* = \log(p_n)/n$ under Condition C1, and $r_n^* = p_n^{4\tau/\beta}/n$ under Condition C2, where $|\cdot|_\infty$ is the matrix elementwise L_∞ norm defined as $|A|_\infty = \max_{i,j} |A(i,j)|$ for a generic matrix A. The proof of Theorem 1 in Ravikumar et al. (2011) indicates that $\{(i,j) :$

$\widehat{\Omega}_{\mathrm{RBLZ}}(i,j) \neq 0\} \subseteq \{(i,j) : \Omega_0(i,j) \neq 0\}$ with probability tending to 1.

For $n = 1, 2, \ldots$, let $\mathcal{A}_n$ denote the event that $|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0|_\infty^2 \le C r_n^*$ and $\{(i,j) : \widehat{\Omega}_{\mathrm{RBLZ}}(i,j) \neq 0\} \subseteq \{(i,j) : \Omega_0(i,j) \neq 0\}$. Hence, $\lim_{n\to\infty} \mathrm{P}(\mathcal{A}_n) = 1$. Then, conditional on event $\mathcal{A}_n$, we have

$$\begin{aligned}
& \|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 \le \|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_F^2 \\
= \ & \sum_{i=1}^{p_n} |\widehat{\Omega}_{\mathrm{RBLZ}}(i,i) - \Omega_0(i,i)|^2 + \sum\sum_{i\neq j:\Omega_0(i,j)\neq 0} |\widehat{\Omega}_{\mathrm{RBLZ}}(i,j) - \Omega_0(i,j)|^2
\end{aligned} \tag{S1.3}$$

and

$$\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 \le \|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_\infty^2 \le t_n^2 |\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0|_\infty^2. \tag{S1.4}$$

If $p_n \le s_n$, then (S1.3) and (S1.4) indicate $\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 \le C \min(p_n + s_n, t_n^2) r_n^* \le C r_n$ under Condition C1 or C2. Next, we consider the case $p_n > s_n$. Define $t_n^i = |\{j = 1, \ldots, p_n : \mathbf{\Theta}_0(i,j) \neq 0\}|$. For any $i \in \{1, \ldots, p_n\}$ such that $t_n^i = 1$, we know $\Omega_0 \boldsymbol{e}_{i,p_n} = \boldsymbol{e}_{i,p_n}$, which means that the diagonal element is the only nonzero element in the ith column of Ω_0. Since $p_n > s_n$, we have $|\{i = 1, \ldots, p_n : t_n^i = 1\}| \ge p_n - s_n$. Because $\{(i,j) : \widehat{\Omega}_{\mathrm{RBLZ}}(i,j) \neq 0\} \subseteq \{(i,j) : \Omega_0(i,j) \neq 0\}$, from the definition of $\widehat{\Omega}_{\mathrm{RBLZ}}$, we have $\widehat{\Omega}_{\mathrm{RBLZ}} \boldsymbol{e}_{i,p_n} = \boldsymbol{e}_{i,p_n}$ for any $i \in \{1, \ldots, p_n\}$ with $\Omega_0 \boldsymbol{e}_{i,p_n} = \boldsymbol{e}_{i,p_n}$. Hence, $|\widehat{\Omega}_{\mathrm{RBLZ}}(i,i) - \Omega_0(i,i)| = 0$ for $i \in \{1, \ldots, p_n\}$ with $t_n^i = 1$. Therefore, (S1.3) indicates $\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 \le C(1 + s_n) r_n^*$, which together with (S1.4) implies that $\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 \le C \min(1 + s_n, t_n^2) r_n^* \le C r_n$.

Hence, under Condition C1 or C2, $\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 = O_{\mathrm{P}}(r_n)$. Therefore, from (S1.1) and (S1.2),

$$\begin{aligned}
& \|\widehat{\mathbf{\Theta}}_{\mathrm{RBLZ}} - \mathbf{\Theta}_0\|_2 = \|\widehat{W}^{-1} \widehat{\Omega}_{\mathrm{RBLZ}} \widehat{W}^{-1} - W_0^{-1} \Omega_0 W_0^{-1}\|_2 \\
\le \ & \|\widehat{W}^{-1} - W_0^{-1}\|_2 \|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2 \|\widehat{W}^{-1} - W_0^{-1}\|_2 \\
& + \|\widehat{W}^{-1} - W_0^{-1}\|_2 (\|\widehat{\Omega}_{\mathrm{RBLZ}}\|_2 \|W_0^{-1}\|_2 + \|\widehat{W}^{-1}\|_2 \|\Omega_0\|_2) \\
& + \|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2 \|\widehat{W}^{-1}\|_2 \|W_0^{-1}\|_2 = O_{\mathrm{P}}(r_n^{1/2}).
\end{aligned}$$

We obtain $\|\widehat{\mathbf{\Theta}}_{\mathrm{RBLZ}}^{-1} - \mathbf{\Sigma}_0\|_2^2 = O_{\mathrm{P}}(r_n)$, since $\|\widehat{\mathbf{\Theta}}_{\mathrm{RBLZ}}^{-1} - \mathbf{\Sigma}_0\|_2^2 = O_{\mathrm{P}}(\|\widehat{\mathbf{\Theta}}_{\mathrm{RBLZ}} - \mathbf{\Theta}_0\|_2^2)$. ■

Proof of Theorem 4. Following the proof of Theorem 3, under Condition C1 or C2, $\|\widehat{\Omega}_{\mathrm{RBLZ}} - \Omega_0\|_2^2 = O_{\mathrm{P}}(r_n)$. Now that $\mathrm{cond}(\widehat{\Omega}_{\mathrm{RBLZ}}) - \mathrm{cond}(\Omega_0) = o_{\mathrm{P}}(1)$, from

$\liminf_{n\to\infty}\{\kappa_n - \mathrm{cond}(\Omega_0)\} > 0$, we have $\mathrm{cond}(\widehat{\Omega}_{\mathrm{RBLZ}}) \le \kappa_n$ with probability tending to 1, which means that $\widehat{\Omega}_{\mu_n,\kappa_n} = \widehat{\Omega}_{\mathrm{RBLZ}}$ with probability tending to 1, and hence $\lim_{n\to\infty} \mathrm{P}(\widehat{\mathbf{\Theta}}_{\mathrm{prop}-2} = \widehat{\mathbf{\Theta}}_{\mathrm{RBLZ}}) = 1$. Therefore, from the conclusion in Theorem 3, $\|\widehat{\mathbf{\Theta}}_{\mathrm{prop}-2} - \mathbf{\Theta}_0\|_2^2 = O_{\mathrm{P}}(r_n) = \|\widehat{\mathbf{\Theta}}_{\mathrm{prop}-2}^{-1} - \mathbf{\Sigma}_0\|_2^2$. ■

Proof of Proposition 1. From (3.2) and (3.3), suppose the eigendecomposition of variable Ω is RMR^T, where R is orthogonal and $M = \mathrm{diag}(m_1, \ldots, m_{p_n})$ with $m_1 \le \cdots \le m_{p_n}$. For Step 1 in Section 3,

$$
\begin{aligned}
& \operatorname*{arg\,min}_{\Omega\succ 0,\,\mathrm{cond}(\Omega)\le\kappa_n} L_\rho(\Omega, Z^{(i-1)}; U^{(i-1)}) \\
= & \operatorname*{arg\,min}_{\Omega\succ 0,\,\mathrm{cond}(\Omega)\le\kappa_n} -\log\{\det(\Omega)\} + \mathrm{tr}(\mathbf{R}_n\Omega) + \frac{\rho}{2}\|\Omega - Z^{(i-1)} + U^{(i-1)}\|_F^2 \\
= & \operatorname*{arg\,min}_{\Omega\succ 0,\,\mathrm{cond}(\Omega)\le\kappa_n} -\log\{\det(\Omega)\} + \mathrm{tr}(\mathbf{R}_n\Omega) + \frac{\rho}{2}\mathrm{tr}\{\Omega\Omega^T + 2(-Z^{(i-1)} + U^{(i-1)})\Omega^T\} \\
= & \operatorname*{arg\,min}_{\Omega\succ 0,\,\mathrm{cond}(\Omega)\le\kappa_n} -\log\{\det(\Omega)\} + \frac{\rho}{2}\mathrm{tr}(\Omega\Omega^T) + \rho\mathrm{tr}\{(\mathbf{R}_n/\rho - Z^{(i-1)} + U^{(i-1)})\Omega^T\} \\
= & \operatorname*{arg\,min}_{\Omega\succ 0,\,\mathrm{cond}(\Omega)\le\kappa_n} -\log\{\det(\Omega)\} + \frac{\rho}{2}\mathrm{tr}(\Omega\Omega^T) + \rho\mathrm{tr}\{(VDV^T)\Omega^T\} \\
= & \operatorname*{arg\,min}_{\Omega=RMR^T:M\succ 0,\,\mathrm{cond}(M)\le\kappa_n} -\log\{\det(M)\} + \frac{\rho}{2}\mathrm{tr}(MM^T) + \rho\mathrm{tr}\{(VDV^T)(RMR^T)^T\} \\
= & \operatorname*{arg\,min}_{\Omega=RMR^T:R=V,\,M\succ 0,\,\mathrm{cond}(M)\le\kappa_n} -\log\{\det(M)\} + \frac{\rho}{2}\mathrm{tr}(MM^T) + \rho\mathrm{tr}(DM^T). \quad \text{(S1.5)}
\end{aligned}
$$

The last equation in (S1.5) is true since $\mathrm{tr}\{(VDV^T)(RMR^T)^T\} \ge \mathrm{tr}(DM^T)$ with equality if $R = V$ (Theorem 14.3.2 in Farrell (1985)). Therefore, to prove $\Omega^{(i)} = V\widetilde{D}V^T$, it suffices to show that

$$
\widetilde{D} = \operatorname*{arg\,min}_{M:M\succ 0,\,\mathrm{cond}(M)\le\kappa_n} -\log\{\det(M)\} + \frac{\rho}{2}\mathrm{tr}(MM^T) + \rho\mathrm{tr}(DM^T),
$$

which is equivalent to

$$
\begin{aligned}
\widetilde{D} &= \operatorname*{arg\,min}_{M:\,0<m_1\le\cdots\le m_{p_n},\,m_{p_n}/m_1\le\kappa_n} \Big\{ -\sum_{j=1}^{p_n}\log(m_j) + \frac{\rho}{2}\sum_{j=1}^{p_n} m_j^2 + \rho\sum_{j=1}^{p_n} d_j m_j \Big\} \\
&= \operatorname*{arg\,min}_{M:\,\exists\,\tau,\,0<\tau\le m_1\le\cdots\le m_{p_n}\le\kappa_n\tau} \sum_{j=1}^{p_n} \Big\{ -\log(m_j) + \frac{\rho}{2}(m_j + d_j)^2 \Big\}. \quad \text{(S1.6)}
\end{aligned}
$$

Define

$$
g(m_j; d_j) = -\log(m_j) + \frac{\rho}{2}(m_j + d_j)^2.
$$

Then, $g(m_j; d_j)$ is strictly convex in $m_j \in (0, \infty)$ for any $j = 1, \ldots, p_n$, and has a unique minimizer $\delta_j = -d_j/2 + \sqrt{d_j^2/4 + 1/\rho}$. Noting that $0 < \delta_1 \le \cdots \le \delta_{p_n}$, if

$\delta_{p_n}/\delta_1 \le \kappa_n$, then $\widetilde{D} = \mathrm{diag}(\delta_1, \ldots, \delta_{p_n})$ coincides with the solution to problem (S1.6) with any $\tau \in [\delta_{p_n}/\kappa_n, \delta_1]$.

For case $\delta_{p_n}/\delta_1 > \kappa_n$, we first consider minimizing the objective function in (S1.6) with respect to $m_1, \ldots, m_{p_n}$ separately. For any $\tau > 0$ and $j = 1, \ldots, p_n$, it follows that

$$\begin{aligned} m_j^*(\tau) \quad &:= \quad \operatorname*{arg\,min}_{\tau \le m_j \le \kappa_n \tau} \sum_{k=1}^{p_n} g(m_k; d_k) = \operatorname*{arg\,min}_{\tau \le m_j \le \kappa_n \tau} g(m_j; d_j) = \min\{\max(\tau, \delta_j), \kappa_n \tau\} \\ &= \begin{cases} \tau, & \text{if } \delta_j < \tau, \\ \delta_j, & \text{if } \tau \le \delta_j \le \kappa_n \tau, \\ \kappa_n \tau, & \text{if } \delta_j > \kappa_n \tau. \end{cases} \end{aligned}$$

Since $\tau \le m_1^*(\tau) \le \cdots \le m_{p_n}^*(\tau) \le \kappa_n \tau$ for any $\tau > 0$, problem (S1.6) amounts to

$$\operatorname*{arg\,min}_{M:\, \exists\, \tau > 0,\, m_j = m_j^*(\tau)} \sum_{j=1}^{p_n} g(m_j; d_j) = \operatorname*{arg\,min}_{M:\, \exists\, \tau > 0,\, m_j = m_j^*(\tau)} \sum_{j=1}^{p_n} g(m_j^*(\tau); d_j).$$

Therefore, to prove that $\widetilde{D}$ is the solution to the optimization problem in (S1.6), we only need to show that τ_0 is the minimizer of

$$f(\tau) := \sum_{j=1}^{p_n} g(m_j^*(\tau); d_j) = \sum_{j:\delta_j < \tau} g(\tau; d_j) + \sum_{j:\tau \le \delta_j \le \kappa_n \tau} g(\delta_j; d_j) + \sum_{j:\delta_j > \kappa_n \tau} g(\kappa_n \tau; d_j).$$

We can verify that $g(m_j^*(\tau); d_j)$ is a convex function of $\tau \in (0, \infty)$ and has a continuous first-order derivative with respect to $\tau \in (0, \infty)$, for any $j = 1, \ldots, p_n$. Therefore, $f(\tau)$ is convex and continuously differentiable for $\tau \in (0, \infty)$. For $\alpha \in \{1, \ldots, p_n\}$ and $\beta \in \{1, \ldots, p_n\}$ such that $\beta - 1 \ge \alpha$, define

$$\begin{aligned} R_{\alpha,\beta} \quad &= \quad \{\tau : \delta_\alpha < \tau \le \delta_{\alpha+1} \text{ and } \delta_{\beta-1} \le \kappa_n \tau < \delta_\beta\}, \\ f_{\alpha,\beta}(\tau) \quad &= \quad \sum_{j=1}^{\alpha} g(\tau; d_j) + \sum_{j=\alpha+1}^{\beta-1} g(\delta_j; d_j) + \sum_{j=\beta}^{p_n} g(\kappa_n \tau; d_j). \end{aligned}$$

Then, $f(\tau) = f_{\alpha,\beta}(\tau)$ for $\tau \in R_{\alpha,\beta}$. Since $f''_{\alpha,\beta}(\tau) > 0$ for $\tau \in R_{\alpha,\beta}$, we know $f'(\tau)$ is strictly monotone increasing on $[\delta_1, \delta_{p_n}/\kappa_n]$. It's also easy to see that $f(\tau)$ is decreasing for $\tau \in (0, \delta_1]$ and increasing for $\tau \in [\delta_{p_n}/\kappa_n, \infty)$. Then, the unique minimizer of $f(\tau)$ is the value of $\tau \in [\delta_1, \delta_{p_n}/\kappa_n]$ such that $f'(\tau) = 0$.

The solution to $f'_{\alpha,\beta}(\tau) = 0$ for $\tau \in (0, \infty)$ is

$$\tau_{\alpha,\beta} \quad = \quad \Big[-\rho\Big(\sum_{j=1}^{\alpha} d_j + \kappa_n \sum_{j=\beta}^{p_n} d_j\Big) + \Big\{\rho^2 \Big(\sum_{j=1}^{\alpha} d_j + \kappa_n \sum_{j=\beta}^{p_n} d_j\Big)^2 + 4\rho(\alpha + {\kappa_n}^2 p_n$$

$$-{\kappa_n}^2\beta+{\kappa_n}^2)(\alpha+p_n-\beta+1)\Big\}^{1/2}\Big]\Big/\{2\rho(\alpha+{\kappa_n}^2p_n-{\kappa_n}^2\beta+{\kappa_n}^2)\}.$$

Then, $\tau_{\alpha,\beta}$ is also the solution to $f'(\tau)=0$ if and only if $\tau_{\alpha,\beta}\in R_{\alpha,\beta}$. This value of $\tau_{\alpha,\beta}$ is the same as τ_0.

In practice, we can search over $\{R_{\alpha,\beta}:\alpha,\,\beta=1,\ldots,p_n\}$ to find α_0 and β_0 such that $\tau_{\alpha_0,\beta_0}\in R_{\alpha_0,\beta_0}$. Start the selection procedure from (α^*,β^*), where $\alpha^*=1$ and β^* is the smallest index in $\{1,\ldots,p_n\}$ such that $\delta_{\beta^*}>\kappa_n\delta_{\alpha^*}$. If $\tau_{\alpha^*,\beta^*}\notin R_{\alpha^*,\beta^*}$, then move on to R_{α^*+1,β^*}, R_{α^*+1,β^*+1} or R_{α^*,β^*+1} for the selection of α_0 and β_0. Specifically, if $\kappa_n\delta_{\alpha^*+1}<\delta_{\beta^*}$, then move on to R_{α^*+1,β^*}; if $\kappa_n\delta_{\alpha^*+1}>\delta_{\beta^*}$, then go to R_{α^*,β^*+1}; otherwise, continue searching α_0 and β_0 within R_{α^*+1,β^*+1}. Repeat the above procedure until condition $\tau_{\alpha,\beta}\in R_{\alpha,\beta}$ is satisfied. The procedure requires $O(p_n)$ operations. ■