Supplementary Materials

Proof of Theorem 2

The proof given below is a slight modification of the proof of Theorem 3.1 in Davanloo et al. (2015) to obtain tighter bounds. For the sake of completeness, we provide the proof. Through the change of variables $\Delta := P - P^*$, we can write (2.1) in terms of Δ as

$$\hat{\Delta} = \operatorname{argmin}\{F(\Delta) := \langle S, \Delta + P^* \rangle - \log\det(\Delta + P^*) + \alpha \left\langle G \otimes (\mathbf{1}_p \mathbf{1}_p^\top), |\Delta + P^*| \right\rangle : \Delta \in \mathcal{F}\},$$

where $\mathcal{F} := \{\Delta \in \mathbb{R}^{np \times np} : \ \Delta = \Delta^\top, \ a^*\mathbf{I} \preceq \Delta + P^* \preceq b^*\mathbf{I}\}$. Note that $\hat{\Delta} = \hat{P} - P^*$. Define $g(\Delta) := -\log\det(\Delta + P^*)$ on $\mathcal{F}$. $g(.)$ is strongly convex over $\mathcal{F}$ with modulus $1/{b^*}^2$; hence, for any $\Delta \in \mathcal{F}$, it follows that $g(\Delta) - g(\mathbf{0}) \geq -\left\langle {P^*}^{-1}, \Delta \right\rangle + \frac{1}{2{b^*}^2}\|\Delta\|_F^2$. Let $H(\Delta) := F(\Delta) - F(\mathbf{0})$ and $S_\Delta := \{\Delta \in \mathcal{F} : \|\Delta\|_F > 2{b^*}^2 p(n + \|G\|_F)\alpha\}$. Under probability event $\Omega = \{\|\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty \leq \alpha, \ \forall (i,j) \in \mathcal{I} \times \mathcal{I}\}$, for any $\Delta \in S_\Delta \subset \mathcal{F}$,

$$\begin{aligned} H(\Delta) &\geq \langle S, \Delta \rangle - \left\langle {P^*}^{-1}, \Delta \right\rangle + \frac{1}{2{b^*}^2}\|\Delta\|_F^2 + \alpha \left\langle G \otimes (\mathbf{1}_p \mathbf{1}_p^\top), |\Delta + P^*| \right\rangle - \alpha \langle G, |P^*| \rangle \\ &\geq \frac{1}{2{b^*}^2}\|\Delta\|_F^2 + \langle \Delta, S - C^* \rangle - \alpha \left\langle G \otimes (\mathbf{1}_p \mathbf{1}_p^\top), |\Delta| \right\rangle \\ &\geq \frac{1}{2{b^*}^2}\|\Delta\|_F^2 - \alpha p(n + \|G\|_F)\|\Delta\|_F > 0, \end{aligned}$$

where the second inequality follows from the triangle inequality, the third one holds under the probability event Ω and follows from the Cauchy-Schwarz inequality, and the final strict one follows from the definition of S_Δ. Since $F(\mathbf{0})$ is a constant, $\hat{\Delta} = \operatorname{argmin}\{H(\Delta) : \Delta \in \mathcal{F}\}$. Hence, $H(\hat{\Delta}) \leq H(\mathbf{0}) = 0$. Therefore, $\hat{\Delta} \notin S_\Delta$ under the probability event Ω. It is important to note that $\hat{\Delta}$ satisfies the first two conditions given in the definition of S_Δ. This implies $\|\hat{\Delta}\|_F \leq 2{b^*}^2 p(n + \|G\|_F)\alpha$ whenever the probability event Ω is true. Hence,

$$\begin{aligned}\Pr\left(\|\hat{P} - P^*\|_F \leq 2{b^*}^2 p(n + \|G\|_F)\alpha\right) &\geq \Pr\left(\|\,\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty \leq \alpha,\ \forall (i,j) \in \mathcal{I} \times \mathcal{I}\right)\\ &= 1 - \Pr\left(\max_{i,j\in\mathcal{I}} \|\,\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty > \alpha\right)\\ &\geq 1 - \sum_{i,j\in\mathcal{I}} \Pr\left(\|\,\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty > \alpha\right).\end{aligned}$$

Recall that $S = \frac{1}{N}\sum_{r=1}^N \mathbf{y}^{(r)}{\mathbf{y}^{(r)}}^\top$ and $\mathbf{y}^{(r)} = [y_i^{(r)}]_{i\in\mathcal{I}}$ for $r = 1,\ldots,N$. Note $\Sigma^{ii} = \Gamma^*$ for $i \in \mathcal{I}$; hence, $y_i^{(r)} \sim \mathcal{N}(\mathbf{0}, \Gamma^*)$, i.e., multivariate Gaussian with mean $\mathbf{0}$ and covariance matrix Γ^*, for all i and r. Therefore, Lemma 1 in Ravikumar et al. (2011) implies $\Pr\left(\|\,\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty > \alpha\right) \leq B_\alpha$ for $\alpha \in (0, 40\max_i \Gamma^*_{ii})$, where $B_\alpha := 4p^2 \exp\left(\frac{-N}{2}\left(\frac{\alpha}{40\max_i \Gamma^*_{ii}}\right)^2\right)$. Hence, given any $M > 0$, by requiring $N \geq \left(\frac{40\max_i \Gamma^*_{ii}}{\alpha}\right)^2 N_0$, we get $B_\alpha \leq \frac{1}{n^2}(np)^{-M}$. Thus, for any $N \geq N_0$, we have $\sum_{i,j\in\mathcal{I}} \Pr\left(\|\,\mathbf{vec}(S^{ij} - \Sigma^{ij})\|_\infty > \alpha\right) \leq (np)^{-M}$ for all $40\max_i \Gamma^*_{ii}\sqrt{\frac{N_0}{N}} \leq \alpha \leq 40\max_i \Gamma^*_{ii}$. □

Proof of Theorem 4

For the sake of simplicity of the notation let $\Phi = (\Gamma, C) \in \mathbb{S}^n \times \mathbb{S}^{np}$, and define $\|(\Gamma, C)\|_a := \max\{\|\Gamma\|_2, \|C\|_2\}$ over the product vector space $\mathbb{S}^n \times \mathbb{S}^{np}$; also let $\Psi = (\boldsymbol{\theta}, \Gamma, C) \in \mathbb{R}^q \times \mathbb{S}^n \times \mathbb{S}^{np}$, and define $\|(\boldsymbol{\theta}, \Gamma, C)\|_b := \|\boldsymbol{\theta}\| + \|(\Gamma, C)\|_a$ over the product vector space $\mathbb{R}^q \times \mathbb{S}^n \times \mathbb{S}^{np}$. Throughout the proof $\hat{\Phi} := (\hat{\Gamma}, \hat{C})$, $\Phi^* := (\Gamma^*, C^*)$, and $\hat{\Psi} := (\hat{\boldsymbol{\theta}}, \hat{\Phi})$, $\Psi^* := (\boldsymbol{\theta}^*, \Phi^*)$.

As $\boldsymbol{\theta}^* \in \mathbf{int}(\Theta)$, there exists $\delta_1 > 0$ such that $\mathcal{B}_{\|.\|_2}(\boldsymbol{\theta}^*, \delta_1) \subset \Theta$. Moreover, since $\rho(\mathbf{x},\mathbf{x}';\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ over Θ for all $\mathbf{x},\mathbf{x}' \in \mathcal{X}$, $R:\Theta \to \mathbb{S}^n$ is also twice continuously differentiable. Hence, from (3.12), it follows that $\nabla^2 f(\boldsymbol{\theta};\Gamma,C)$ is continuous in $\Psi = (\boldsymbol{\theta},\Gamma,C)$; and since eigenvalues of a matrix are continuous functions of matrix entries, $\lambda_{\min}\left(\nabla^2 f(\boldsymbol{\theta};\Gamma,C)\right)$ is continuous in Ψ on $\mathcal{B}_{\|.\|_b}(\Psi^*,\delta_1)$ as well. Therefore, it follows from Lemma 1 that there exists $0<\delta_2\leq\delta_1$ such that $\nabla^2_{\boldsymbol{\theta}} f(\boldsymbol{\theta};\Gamma,C) \succeq \frac{\gamma^*}{2} I$ for all $\Psi = (\boldsymbol{\theta},\Gamma,C)\in\mathcal{B}_{\|.\|_b}(\Psi^*,\delta_2)$.

Let $\mathcal{Q} := \bar{\mathcal{B}}_{\|.\|_a}(\Phi^*,\frac{1}{2}\delta_2)$ and $\Theta' := \Theta \cap \bar{\mathcal{B}}_{\|.\|_2}(\boldsymbol{\theta}^*,\frac{1}{2}\delta_2)$, i.e.,

$$\mathcal{Q} = \{(\Gamma,C):\ \max\{\|\Gamma-\Gamma^*\|_2,\ \|C-C^*\|_2\} \leq \tfrac{1}{2}\delta_2\}, \tag{5.1}$$

$$\Theta' = \{\boldsymbol{\theta}\in\Theta:\ \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \leq \tfrac{1}{2}\delta_2\}. \tag{5.2}$$

Clearly f is strongly convex in $\boldsymbol{\theta}$ over Θ' with convexity modulus $\frac{\gamma^*}{2}$ for all $(\Gamma,C)\in\mathcal{Q}$. Define the unique minimizer over Θ':

$$\boldsymbol{\theta}(\Gamma,C) := \operatorname*{argmin}_{\boldsymbol{\theta}\in\Theta'} f(\boldsymbol{\theta};\Gamma,C). \tag{5.3}$$

Since Θ' is a convex compact set and $f(\boldsymbol{\theta};\Gamma,C)$ is jointly continuous in $\Psi = (\boldsymbol{\theta},\Gamma,C)$ on $\Theta'\times\mathcal{Q}$, from Berge's Maximum Theorem – see Ok (2007), $\boldsymbol{\theta}(\Gamma,C)$ is continuous at (Γ^*,C^*) and $\boldsymbol{\theta}(\Gamma^*,C^*) = \boldsymbol{\theta}^*$. Therefore, for any $0<\epsilon\leq\frac{1}{2}\delta_2$, there exists $\delta(\epsilon)>0$ such that $\delta(\epsilon)\leq\frac{1}{2}\delta_2$ and $\|\boldsymbol{\theta}(\Gamma,C)-\boldsymbol{\theta}^*\| < \epsilon$ for all $\Phi=(\Gamma,C)$ satisfying $\|\Phi-\Phi^*\|_a<\delta(\epsilon)$.

Fix some arbitrary $\epsilon\in(0,\frac{1}{2}\delta_2]$. Let $\hat{P}(\epsilon)$ be computed as in (3.6) with

$\alpha(\epsilon) = 40 \max_{i=1,\ldots,p}(\Gamma^*_{ii})\sqrt{\frac{N_0}{N(\epsilon)}}$ where sample size $N(\epsilon)$ denotes the number of process realizations (chosen depending on $\epsilon > 0$). Hence, Theorem 3 implies that by choosing $N(\epsilon)$ sufficiently large, we can guarantee that $\hat{C}(\epsilon) = \hat{P}(\epsilon)^{-1}$, and $\hat{\Gamma}(\epsilon)$ defined as in (3.8) satisfy

$$\max\{\|\hat{C}(\epsilon) - C^*\|_2,\ \|\hat{\Gamma}(\epsilon) - \Gamma^*\|_2\} < \delta(\epsilon) \leq \tfrac{1}{2}\delta_2, \tag{5.4}$$

i.e., $\|\hat{\Phi} - \Phi^*\|_a < \delta(\epsilon)$, with high probability. In the rest of the proof, for the sake of notational simplicity, we do not explicitly show the dependence on the fixed tolerance ϵ; instead we simply write $\hat{P}$, $\hat{C}$, and $\hat{\Gamma}$.

Note that due to the parametric continuity discussed above, (5.4) implies that $\|\boldsymbol{\theta}(\hat{\Gamma}, \hat{C}) - \boldsymbol{\theta}^*\| < \epsilon \leq \frac{1}{2}\delta_2$. Hence, the norm-ball constraint in the definition of Θ' will not be tight when $f(\boldsymbol{\theta}; \hat{\Gamma}, \hat{C})$ is minimized over $\boldsymbol{\theta} \in \Theta'$, i.e., $\boldsymbol{\theta}(\hat{\Gamma}, \hat{C}) = \operatorname{argmin}_{\boldsymbol{\theta}\in\Theta'} f(\boldsymbol{\theta}; \hat{\Gamma}, \hat{C}) = \operatorname{argmin}_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{\theta}; \hat{\Gamma}, \hat{C}) =: \hat{\boldsymbol{\theta}}$ – see (3.9) for the definition of $\hat{\boldsymbol{\theta}}$. Therefore, $\|\hat{\Psi} - \Psi^*\|_b < \delta_2 \leq \delta_1$, i.e.,

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| + \|(\hat{\Gamma}, \hat{C}) - (\Gamma^*, C^*)\|_a < \delta_2 \leq \delta_1. \tag{5.5}$$

This implies that $\hat{\boldsymbol{\theta}} \in \mathbf{int}\,\Theta$; thus, $\nabla_{\boldsymbol{\theta}} f(\hat{\boldsymbol{\theta}}; \hat{\Gamma}, \hat{C}) = \mathbf{0}$.

Although one can establish a direct relation between $\delta(\epsilon)$ and ϵ by showing that $\boldsymbol{\theta}(\Gamma, C)$ is Lipschitz continuous around $\boldsymbol{\theta}^*$, we will show a more specific result by upper bounding the error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|$ using $\|\hat{\Phi} - \Phi^*\|_a$. Indeed, since $(\hat{\Gamma}, \hat{C}) \in \mathcal{Q}$, $f(\boldsymbol{\theta}; \hat{\Gamma}, \hat{C})$ is strongly convex in $\boldsymbol{\theta} \in \Theta'$ with modulus $\frac{1}{2}\gamma^*$; hence, $\boldsymbol{\theta}^* \in \Theta'$ and $\hat{\boldsymbol{\theta}} \in \Theta'$ imply that

$$\begin{aligned}\frac{\gamma^*}{2}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 &\leq \left\langle \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \hat{\Gamma}, \hat{C}) - \nabla_{\boldsymbol{\theta}} f(\hat{\boldsymbol{\theta}}; \hat{\Gamma}, \hat{C}),\ \boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}} \right\rangle \\ &= \left\langle \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \hat{\Gamma}, \hat{C}) - \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \Gamma^*, C^*),\ \boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}} \right\rangle, \qquad (5.6)\end{aligned}$$

where the equality follows from the fact that $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \Gamma^*, C^*) = \nabla_{\boldsymbol{\theta}} f(\hat{\boldsymbol{\theta}}; \hat{\Gamma}, \hat{C}) = \mathbf{0}$. Next, from (3.11) it follows that

$$\begin{aligned}\Delta_k &:= \left| \frac{\partial}{\partial \theta_k} f(\boldsymbol{\theta}^*; \hat{\Gamma}, \hat{C}) - \frac{\partial}{\partial \theta_k} f(\boldsymbol{\theta}^*; \Gamma^*, C^*) \right| \\ &\leq \left| (\|\hat{\Gamma}\|_F^2 - \|\Gamma^*\|_F^2) \langle R_k'(\boldsymbol{\theta}^*), R(\boldsymbol{\theta}^*) \rangle + \langle C^*, R_k'(\boldsymbol{\theta}^*) \otimes \Gamma^* \rangle - \langle \hat{C}, R_k'(\boldsymbol{\theta}^*) \otimes \hat{\Gamma} \rangle \right| \\ &\leq \left(\|\hat{\Gamma} + \Gamma^*\|_* \|R(\boldsymbol{\theta}^*)\|_* + \|\hat{C}\|_* \right) \|R_k'(\boldsymbol{\theta}^*)\|_2 \|\hat{\Gamma} - \Gamma^*\|_2 + n\|\Gamma^*\|_* \|R_k'(\boldsymbol{\theta}^*)\|_2 \|\hat{C} - C^*\|_2,\end{aligned}$$

where the second inequality uses the following basic inequalities and identities: Given $X, Y, V, W \in \mathbb{R}^{m \times n}$ **i)** $\langle X, Y \rangle \leq \|X\|_2 \|Y\|_*$, **ii)** $\|X\|_F^2 - \|Y\|_F^2 = \langle X + Y, X - Y \rangle$, **iii)** $\langle X, Y \rangle - \langle V, W \rangle = \langle X, Y - W \rangle + \langle W, X - V \rangle$; given $X \in \mathbb{S}^p$, $Y \in \mathbb{S}^n$ **iv)** $\|X \otimes Y\|_2 = \|X\|_2 \|Y\|_2$, **v)** $\|X \otimes Y\|_* \leq \min\{p\|X\|_2\|Y\|_*,\ n\|X\|_*\|Y\|_2\}$. Note that since $R(\boldsymbol{\theta}^*) \in \mathbb{S}^n_{++}$, $\|R(\boldsymbol{\theta}^*)\|_* = \mathbf{Tr}(R(\boldsymbol{\theta}^*)) = n$. Moreover, (5.4) implies that $\|\hat{\Gamma}\|_* \leq \|\Gamma^*\|_* + \frac{p}{2}\delta_2$, and $\|\hat{C}\|_* \leq \|C^*\|_* + \frac{np}{2}\delta_2$. Hence,

$$\Delta_k \leq \left(3n\|\Gamma^*\|_* + \|C^*\|_* + \frac{(np+1)}{2}\delta_2 \right) \|R_k'(\boldsymbol{\theta}^*)\|_2 \|(\hat{\Gamma}, \hat{C}) - (\Gamma^*, C^*)\|_a.$$

Therefore, for $\kappa := \left(3n\|\Gamma^*\|_* + \|C^*\|_* + \frac{(np+1)}{2}\delta_2 \right) \left(\sum_{k=1}^q \|R_k'(\boldsymbol{\theta}^*)\|_2^2\right)^{\frac{1}{2}}$

$$\|\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \hat{\Gamma}, \hat{C}) - \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^*; \Gamma^*, C^*)\|_2 \leq \kappa\ \|(\hat{\Gamma}, \hat{C}) - (\Gamma^*, C^*)\|_a$$

Applying Cauchy Schwarz inequality to (5.6), we have

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \leq 2\frac{\kappa}{\gamma^*} \ \|(\hat{\Gamma}, \hat{C}) - (\Gamma^*, C^*)\|_a. \tag{5.7}$$

Thus, choosing $N(\epsilon) \geq N_0 := \left\lceil 2\left[(M+2)\ln(np) + \ln 4\right]\right\rceil$ such that

$$\sqrt{\frac{N(\epsilon)}{N_0}} \geq 160 \max_{i=1,\ldots,p}(\Gamma^*_{ii})\frac{\kappa}{\gamma^*}\left(\frac{b^*}{a^*}\right)^2 p(n + \|G\|_F) \ \frac{1}{\epsilon},$$

i.e., $N(\epsilon) = \mathcal{O}(\frac{1}{\epsilon^2})$, implies that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \leq \epsilon$, and $\|\hat{\Gamma} - \Gamma^*\|_2 \leq \frac{\gamma^*}{2\kappa}\epsilon$ with probability at least $1 - (np)^{-M}$. □