Supplementary Materials for: The Log-Shift Penalty for Adaptive Estimation of Multiple Gaussian Graphical Models

Yuancheng Zhu
University of Chicago

Rina Foygel Barber
University of Chicago

A Proofs

Proof of Theorem 1. For any $\Omega^{(k)} \succeq 0$ with $\|\Omega^{(k)}\|_{\mathsf{op}} \leq b_k$,

$$\begin{aligned}
&\lambda_{\min}\left[\nabla^2_{\Omega^{(k)}}\left(\langle\Omega^{(k)}, S^{(k)}\rangle - \log\det(\Omega^{(k)})\right)\right] \\
&= \lambda_{\min}\left[\Omega^{(k)-1}\otimes\Omega^{(k)-1}\right] \\
&= \left[\lambda_{\min}(\Omega^{(k)-1})\right]^2 = \frac{1}{\|\Omega^{(k)}\|^2_{\mathsf{op}}} \geq \frac{1}{b_k^2}\ .
\end{aligned}$$

Then

$$\mathbf{\Omega} \mapsto \sum_{k=1}^K \frac{n_k}{2}\left(\langle\Omega^{(k)}, S^{(k)}\rangle - \log\det(\Omega^{(k)}) - \frac{\|\Omega^{(k)}\|_F^2}{b_k^2}\right)$$

is a convex function over $\mathbf{\Omega} \in \mathcal{S}_p(b)$. Applying Lemma 1 given below, the function

$$x \mapsto \log(1 + f(x)/\beta) + \frac{L^2}{2\beta^2}\|x\|_2^2$$

is convex, and so the following is a convex function over $\mathbf{\Omega} \in \mathcal{S}_p(b)$:

$$\begin{aligned}
\mathbf{\Omega} \mapsto &\sum_{k=1}^K \frac{n_k}{2}\left(\langle\Omega^{(k)}, S^{(k)}\rangle - \log\det(\Omega^{(k)}) - \frac{\|\Omega^{(k)}\|_F^2}{b_k^2}\right) \\
&+ \gamma\sum_{i<j}\beta\left[\log\left(1 + f(\mathbf{\Omega}_{ij})/\beta\right) + \frac{L^2}{2\beta^2}\|\mathbf{\Omega}_{ij}\|_2^2\right] \\
&= F(\mathbf{\Omega}) - \sum_{k=1}^K \frac{n_k}{2}\cdot\frac{\|\Omega^{(k)}\|_F^2}{b_k^2} + \gamma\sum_{i<j}\frac{L^2}{2\beta}\|\mathbf{\Omega}_{ij}\|_2^2 \\
&= F(\mathbf{\Omega}) - \sum_{k=1}^K \|\Omega^{(k)}\|_F^2\cdot\left(\frac{n_k}{2b_k^2} - \frac{\gamma L^2}{4\beta}\right)\ ,
\end{aligned}$$

where the switch from a 2 to a 4 in the last step comes from the fact that $\mathbf{\Omega}_{ij}$ is penalized for $i<j$ but not $i>j$.

This proves that $F(\mathbf{\Omega})$ is convex over $\mathcal{S}_p(b)$ as long as $\frac{n_k}{2b_k^2} \geq \frac{\gamma L^2}{4\beta}$ for all k, which is equivalent to the condition in the theorem. If this inequality is strictly satisfied, then this implies strict convexity of $F(\mathbf{\Omega})$. □

Lemma 1. *Let $f: \mathbb{R}^p \to \mathbb{R}$ be a L-Lipschitz convex non-negative function and fix any $\beta > 0$. Then*

$$x \mapsto \log\left(1 + f(x)/\beta\right) + \frac{L^2}{2\beta^2}\|x\|_2^2$$

is a convex function.

Proof. Take any $x, y \in \mathbb{R}^p$, and any $t \in [0,1]$. Then, using the convexity of $f(\cdot)$,

$$\begin{aligned}
&\log\left(1 + f(t\cdot x + (1-t)\cdot y)/\beta\right) \\
&\leq \log\left(1 + t\cdot f(x)/\beta + (1-t)\cdot f(y)/\beta\right) \\
&= \log\left(t\cdot(1 + f(x)/\beta) + (1-t)\cdot(1 + f(y)/\beta)\right)
\end{aligned}$$

Since $\frac{\partial^2}{\partial z^2}\log(z) \in [-1, 0]$ for all $z \geq 1$,

$$\begin{aligned}
&\leq t\cdot\log\left(1 + f(x)/\beta\right) + (1-t)\cdot\log\left(1 + f(y)/\beta\right) \\
&\quad + \frac{t(1-t)}{2\beta^2}\cdot(f(y) - f(x))^2 \\
&\leq t\cdot\log\left(1 + f(x)/\beta\right) + (1-t)\cdot\log\left(1 + f(y)/\beta\right) \\
&\quad + \frac{t(1-t)}{2\beta^2}\cdot L^2\|x - y\|_2^2\ .
\end{aligned}$$

Then

$$\begin{aligned}
&\log\left(1 + f(t\cdot x + (1-t)\cdot y)/\beta\right) \\
&\quad + \frac{L^2}{2\beta^2}\|t\cdot x + (1-t)\cdot y\|_2^2 \\
&\leq t\cdot\log\left(1 + f(x)/\beta\right) + (1-t)\cdot\log\left(1 + f(y)/\beta\right) \\
&\quad + \frac{t(1-t)}{2\beta^2}\cdot L^2\|x-y\|_2^2 + \frac{L^2}{2\beta^2}\|t\cdot x + (1-t)\cdot y\|_2^2 \\
&\leq t\cdot\left[\log\left(1 + f(x)/\beta\right) + \frac{L^2}{2\beta^2}\|x\|_2^2\right] \\
&\quad + (1-t)\cdot\left[\log\left(1 + f(y)/\beta\right) + \frac{L^2}{2\beta^2}\|y\|_2^2\right]\ ,
\end{aligned}$$

proving convexity of the function as desired. □

Proof of Theorem 2. Define

$$\mathcal{S}_p(b;\mathcal{A}) = \left\{\boldsymbol{\Omega} \in \mathcal{S}_p(b) : \Omega^{(k)}_{ij} = 0,\ \forall\, k \text{ and } \forall\, i \not\sim_{\mathcal{A}} j\right\} \subset \mathcal{S}_p(b)$$

and let

$$\widehat{\boldsymbol{\Omega}} \in \underset{\boldsymbol{\Omega}\in\mathcal{S}_p(b;\mathcal{A})}{\arg\min}\ F(\boldsymbol{\Omega})\ .$$

We will show that $\widehat{\boldsymbol{\Omega}}$ is a minimizer of $F(\boldsymbol{\Omega})$ over the larger set $\mathcal{S}_p(b)$.

Take any $\Delta = (\Delta^{(1)},\dots,\Delta^{(K)})$ with $\Delta^{(k)} \in \mathbb{R}^{d\times d}$ for each k. Let D and E be the block-diagonal and off-block-diagonal parts of Δ; that is,

$$D^{(k)}_{ij} = \begin{cases} \Delta^{(k)}_{ij} & \text{if } i \sim_{\mathcal{A}} j \\ 0 & \text{if } i \not\sim_{\mathcal{A}} j \end{cases}$$

$$\text{and } E^{(k)}_{ij} = \begin{cases} 0 & \text{if } i \sim_{\mathcal{A}} j \\ \Delta^{(k)}_{ij} & \text{if } i \not\sim_{\mathcal{A}} j \end{cases}\ .$$

Suppose that $\widehat{\boldsymbol{\Omega}} + \Delta \in \mathcal{S}_p(b)$. Then

$$b_k \ge \|\widehat{\Omega}^{(k)} + \Delta^{(k)}\| \ge \|\left(\widehat{\Omega}^{(k)} + \Delta^{(k)}\right)_{A_r,A_r}\| = \|\left(\widehat{\Omega}^{(k)} + D^{(k)}\right)_{A_r,A_r}\|\ ,$$

and so

$$\|\widehat{\Omega}^{(k)} + D^{(k)}\| = \max_r \|\left(\widehat{\Omega}^{(k)} + D^{(k)}\right)_{A_r,A_r}\| \le b_k\ ,$$

proving that $\widehat{\boldsymbol{\Omega}} + D \in \mathcal{S}_p(b;\mathcal{A})$. Then by optimality of $\widehat{\boldsymbol{\Omega}}$ over the set $\mathcal{S}_p(b;\mathcal{A})$,

$$F(\widehat{\boldsymbol{\Omega}} + D) \ge F(\widehat{\boldsymbol{\Omega}})\ .$$

Then

$$\begin{aligned} F(\widehat{\boldsymbol{\Omega}} + \Delta) &= F(\widehat{\boldsymbol{\Omega}} + D + E) \\ &= \sum_{k=1}^K \frac{n_k}{2}\Big(\langle \widehat{\Omega}^{(k)} + D^{(k)} + E^{(k)}, S^{(k)}\rangle \\ &\quad - \log\det(\widehat{\Omega}^{(k)} + D^{(k)} + E^{(k)})\Big) \\ &\quad + \gamma \sum_{i<j} \beta \log\left(1 + f(\widehat{\boldsymbol{\Omega}}_{ij} + D_{ij} + E_{ij})/\beta\right) \end{aligned}$$

Since D_{ij} and $\widehat{\boldsymbol{\Omega}}_{ij}$ are nonzero only when i,j are in the same block, and E_{ij} is nonzero only if i,j are in different blocks,

$$\begin{aligned} &= \sum_{k=1}^K \frac{n_k}{2}\left(\langle \widehat{\Omega}^{(k)} + D^{(k)}, S^{(k)}\rangle - \log\det(\widehat{\Omega}^{(k)} + D^{(k)})\right) \\ &\quad + \gamma \sum_{i<j} \beta \log\left(1 + f(\widehat{\boldsymbol{\Omega}}_{ij} + D_{ij})/\beta\right) \\ &\quad + \sum_{k=1}^K \frac{n_k}{2}\Big(\langle E^{(k)}, S^{(k)}\rangle - \log\det(\widehat{\Omega}^{(k)} + D^{(k)} + E^{(k)}) \\ &\quad + \log\det(\widehat{\Omega}^{(k)} + D^{(k)})\Big) \\ &\quad + \gamma \sum_{i\not\sim_{\mathcal{A}} j, i<j} \beta \log\left(1 + f(E_{ij})/\beta\right) \end{aligned}$$

Letting $M = \#\{(i,j) : i \not\sim_{\mathcal{A}} j, i<j\}$,

$$\begin{aligned} &= F(\widehat{\boldsymbol{\Omega}} + D) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^K \frac{n_k}{2}\Big(\langle E^{(k)}, S^{(k)}\rangle - \log\det(\widehat{\Omega}^{(k)} + D^{(k)} + E^{(k)}) \\ &\quad + \log\det(\widehat{\Omega}^{(k)} + D^{(k)})\Big) \\ &\quad + \gamma \sum_{i\not\sim_{\mathcal{A}} j, i<j} \beta \log\left(1 + f(E_{ij})/\beta\right) \end{aligned}$$

By optimality of $\boldsymbol{\Omega}$ over $\mathcal{S}_p(b;\mathcal{A})$,

$$\begin{aligned} &\ge F(\widehat{\boldsymbol{\Omega}}) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^K \frac{n_k}{2}\Big(\langle E^{(k)}, S^{(k)}\rangle - \log\det(\widehat{\Omega}^{(k)} + D^{(k)} + E^{(k)}) \\ &\quad + \log\det(\widehat{\Omega}^{(k)} + D^{(k)})\Big) \\ &\quad + \gamma \sum_{i\not\sim_{\mathcal{A}} j, i<j} \beta \log\left(1 + f(E_{ij})/\beta\right) \end{aligned}$$

Since $\frac{\partial}{\partial \boldsymbol{\Omega}} \log\det(\boldsymbol{\Omega}) = -\boldsymbol{\Omega}^{-1}$ and $\|\nabla^2 \log\det(\boldsymbol{\Omega}')\|$ is bounded for $\boldsymbol{\Omega}'$ near $\boldsymbol{\Omega}$, we can apply a Taylor expansion to the difference of $\log\det(\cdot)$ terms:

$$\begin{aligned} &= F(\widehat{\boldsymbol{\Omega}}) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^K \frac{n_k}{2}\Big(\langle E^{(k)}, S^{(k)}\rangle - \langle E^{(k)}, -(\Omega^{(k)} + D^{(k)})^{-1}\rangle \\ &\quad - \mathcal{O}\left(\|E^{(k)}\|_F^2\right)\Big) \\ &\quad + \gamma \sum_{i\not\sim_{\mathcal{A}} j, i<j} \beta \log\left(1 + f(E_{ij})/\beta\right) \end{aligned}$$

Since $\widehat{\Omega}^{(k)} + D^{(k)}$ is block-diagonal and therefore so is $\left(\widehat{\Omega}^{(k)} + D^{(k)}\right)^{-1}$, while $E^{(k)}$ is supported off of the diagonal blocks,

$$\begin{aligned} &= F(\widehat{\mathbf{\Omega}}) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^{K} \frac{n_k}{2} \left(\langle E^{(k)}, S^{(k)} \rangle - \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) \right) \\ &\quad + \gamma \sum_{i \not\sim_{\mathcal{A}} j, i<j} \beta \log\left(1 + f(E_{ij})/\beta\right) \end{aligned}$$

Applying Taylor expansion to the terms $\log(1+f(E_{ij})/\beta)$, for E_{ij} sufficiently close to 0,

$$\begin{aligned} &= F(\widehat{\mathbf{\Omega}}) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^{K} \frac{n_k}{2} \left(\langle E^{(k)}, S^{(k)} \rangle - \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) \right) \\ &\quad + \gamma \sum_{i \not\sim_{\mathcal{A}} j, i<j} \beta \Big[\log(1 + f(\mathbf{0})/\beta) + \frac{f(E_{ij}) - f(\mathbf{0})}{\beta} \\ &\quad - \mathcal{O}\left((f(E_{ij}) - f(\mathbf{0}))^2 \right) \Big] \end{aligned}$$

Since f is L-Lipschitz,

$$\begin{aligned} &= F(\widehat{\mathbf{\Omega}}) - M\gamma \cdot \beta \log(1 + f(\mathbf{0})/\beta) \\ &\quad + \sum_{k=1}^{K} \frac{n_k}{2} \left(\langle E^{(k)}, S^{(k)} \rangle - \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) \right) \\ &\quad + \gamma \sum_{i \not\sim_{\mathcal{A}} j, i<j} \beta \Big[\log(1 + f(\mathbf{0})/\beta) + \frac{f(E_{ij}) - f(\mathbf{0})}{\beta} \\ &\quad - L^2 \cdot \mathcal{O}\left(\|E_{ij}\|_2^2 \right) \Big] \end{aligned}$$

Simplifying,

$$\begin{aligned} &= F(\widehat{\mathbf{\Omega}}) + \sum_{k=1}^{K} \frac{n_k}{2} \left(\langle E^{(k)}, S^{(k)} \rangle - \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) \right) \\ &\quad + \gamma \sum_{i \not\sim_{\mathcal{A}} j, i<j} \beta \left[\frac{f(E_{ij}) - f(\mathbf{0})}{\beta} - L^2 \cdot \mathcal{O}\left(\|E_{ij}\|_2^2 \right) \right] \end{aligned}$$

If we assume that $-\gamma^{-1} \cdot \mathrm{diag}\{n_1, \ldots, n_K\} \cdot \mathbf{S}_{ij} \in \partial f(\mathbf{0})$ for all $i \not\sim_{\mathcal{A}} j$,

$$\begin{aligned} &\geq F(\widehat{\mathbf{\Omega}}) + \sum_{k=1}^{K} \frac{n_k}{2} \left(\langle E^{(k)}, S^{(k)} \rangle - \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) \right) \\ &\quad + \gamma \sum_{i \not\sim_{\mathcal{A}} j, i<j} \beta \Big[\frac{\langle E_{ij}, -\gamma^{-1} \cdot \mathrm{diag}\{n_1, \ldots, n_K\} \cdot \mathbf{S}_{ij} \rangle}{\beta} \\ &\quad - L^2 \cdot \mathcal{O}\left(\|E_{ij}\|_2^2 \right) \Big] \end{aligned}$$

Cancelling out the terms that are linear in E,

$$\begin{aligned} &= F(\widehat{\mathbf{\Omega}}) - \sum_{k=1}^{K} \frac{n_k}{2} \mathcal{O}\left(\|E^{(k)}\|_F^2 \right) - \alpha \sum_{i \not\sim_{\mathcal{A}} j, i<j} L^2 \mathcal{O}\left(\|E_{ij}\|_2^2 \right) \\ &= F(\widehat{\mathbf{\Omega}}) - \mathcal{O}\left(\sum_{i,j,k} E_{ij}^{(k)2} \right) \geq F(\widehat{\mathbf{\Omega}}) - \mathcal{O}\left(\sum_{i,j,k} \Delta_{ij}^{(k)2} \right) . \end{aligned}$$

Since F is convex over $\mathcal{S}_p(b)$ which is itself a convex set, this is sufficient to prove that $\widehat{\mathbf{\Omega}}$ is a minimizer of $F(\cdot)$ over $\mathcal{S}_p(b)$. □