Supplementary Material to Sparse Linear Isotonic Models

Sheng Chen **Arindam Banerjee**
Department of Computer Science and Engineering, University of Minnesota, Twin Cities
{shengc, banerjee}@cs.umn.edu

1 Proof of Lemma 2

Statement of Lemma: *The transformed population Kendall's tau correlation vector $\boldsymbol{\beta}$ satisfies*

$$\boldsymbol{\beta} = \frac{\tilde{\boldsymbol{\beta}}}{\sigma_y} = \frac{\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\theta}}}{\sigma_y} , \tag{S.1}$$

where σ_y^2 is the variance of y. The transformed sample Kendall's tau correlation vector $\hat{\boldsymbol{\beta}}$, with probability at least $1 - \frac{2}{p}$, satisfies

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty \leq 2\pi\sqrt{\frac{\log p}{n}} \tag{S.2}$$

Proof: By definition, $\tilde{\boldsymbol{\beta}} = \mathbb{E}[y\tilde{\mathbf{x}}] = \mathbb{E}_{\tilde{\mathbf{x}}}[\tilde{\mathbf{x}} \cdot \mathbb{E}_y[y|\tilde{\mathbf{x}}]] = \mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\tilde{\boldsymbol{\theta}}] = \tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\theta}}$. Given that $\lambda_{\min} > 0$ and the properties of elliptical distribution (15), we have $\mathbb{E}[\tilde{\mathbf{x}}] = \mathbf{0}$, $\text{rank}(\mathbf{A}) = \text{rank}(\tilde{\boldsymbol{\Sigma}}) = p$ and $\text{Cov}[\tilde{\mathbf{x}}] = \tilde{\boldsymbol{\Sigma}}$. Since $\tilde{\mathbf{x}}$, y are jointly elliptical and $\boldsymbol{\beta}$ is invariant to $\mathbf{f}$, using Theorem 2 in [3], we have for each β_j,

$$\beta_j = \frac{\mathbb{E}[y\tilde{x}_j] - \mathbb{E}[y]\mathbb{E}[\tilde{x}_j]}{\sqrt{\text{Var}[y]}\sqrt{\text{Var}[\tilde{x}_j]}} = \frac{\mathbb{E}\left[\langle\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{x}}\rangle \cdot \tilde{x}_j\right]}{\sqrt{\text{Var}[y]}} = \frac{\langle\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}_j\rangle}{\sigma_y} ,$$

which implies (S.1). Using Hoeffding's inequality for U-statistics [2], we have for each β_j and $\hat{\beta}_j$

$$\mathbb{P}\left(\left|\beta_j - \hat{\beta}_j\right| \geq \epsilon\right) \leq \mathbb{P}\left(\left|b_j - \hat{b}_j\right| \geq \frac{2}{\pi}\epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2\pi^2}\right) .$$

Letting $\epsilon = 2\pi\sqrt{\frac{\log p}{n}}$ and taking union bound, we obtain

$$\mathbb{P}\left(\left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\|_\infty \geq 2\pi\sqrt{\frac{\log p}{n}}\right) \leq \frac{2}{p} ,$$

which completes the proof. ■

2 Proof of Lemma 3

Statement of Lemma: *Define the descent cone for any s-sparse vector $\boldsymbol{\theta}^* \in \mathbb{R}^p$,*

$$\mathcal{C} = \{\mathbf{v} \in \mathbb{R}^p \mid \|\boldsymbol{\theta}^* + \mathbf{v}\|_1 \leq \|\boldsymbol{\theta}^*\|_1\} . \tag{S.3}$$

If $\mathbf{x} \sim TE(\tilde{\boldsymbol{\Sigma}}, \xi, \mathbf{f})$ and $n \geq \left(\frac{24\pi}{\lambda_{\min}}\right)^2 s^2 \log p = O(s^2 \log p)$, with probability at least $1 - p^{-2.5}$, the following RE condition holds for $\hat{\boldsymbol{\Sigma}}$ in $\mathcal{C}$,

$$\inf_{\mathbf{v} \in \mathcal{C} \cap \mathbb{S}^{p-1}} \mathbf{v}^T\hat{\boldsymbol{\Sigma}}\mathbf{v} \geq \frac{\lambda_{\min}}{2} , \tag{S.4}$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\tilde{\boldsymbol{\Sigma}}$.

Proof: Let $\mathcal{S}$ be the support of $\boldsymbol{\theta}^*$, then we have

$$\mathbf{v} \in \mathcal{C} \cap \mathbb{S}^{p-1} \implies \|\boldsymbol{\theta}^*_\mathcal{S} + \mathbf{v}_\mathcal{S}\|_1 + \|\mathbf{v}_{\mathcal{S}^c}\|_1 \leq \|\boldsymbol{\theta}^*\|_1 \implies \|\boldsymbol{\theta}^*_\mathcal{S}\|_1 - \|\mathbf{v}_\mathcal{S}\|_1 + \|\mathbf{v}_{\mathcal{S}^c}\|_1 \leq \|\boldsymbol{\theta}^*\|_1 \implies \|\mathbf{v}_{\mathcal{S}^c}\|_1 \leq \|\mathbf{v}_\mathcal{S}\|_1 \implies \|\mathbf{v}\|_1 \leq 2\|\mathbf{v}_\mathcal{S}\|_1 \leq 2\sqrt{s}\|\mathbf{v}_\mathcal{S}\|_2 \leq 2\sqrt{s}$$

With probability at least $1 - p^{-2.5}$, we have for any $\mathbf{v} \in \mathcal{C} \cap \mathbb{S}^{p-1}$

$$\mathbf{v}^T\hat{\boldsymbol{\Sigma}}\mathbf{v} \geq \mathbf{v}^T\tilde{\boldsymbol{\Sigma}}\mathbf{v} - \left|\mathbf{v}^T\left(\hat{\boldsymbol{\Sigma}} - \tilde{\boldsymbol{\Sigma}}\right)\mathbf{v}\right|$$
$$\geq \lambda_{\min} - \left|\sum_{1 \leq i,j \leq p} v_i v_j \left(\hat{\sigma}_{ij} - \tilde{\sigma}_{ij}\right)\right|$$
$$\geq \lambda_{\min} - \|\mathbf{v}\|_1^2 \left\|\hat{\boldsymbol{\Sigma}} - \tilde{\boldsymbol{\Sigma}}\right\|_{\max} \geq \lambda_{\min} - 12\pi\sqrt{\frac{s^2 \log p}{n}} ,$$

where we use Lemma 1 and the fact that $\|\mathbf{v}\|_1 \leq 2\sqrt{s}$. Since we choose $n \geq \left(\frac{24\pi}{\lambda_{\min}}\right)^2 s^2 \log p$, we have

$$\mathbf{v}^T\hat{\boldsymbol{\Sigma}}\mathbf{v} \geq \lambda_{\min} - 12\pi\sqrt{\frac{s^2 \log p}{n}} \geq \lambda_{\min} - \frac{\lambda_{\min}}{2} = \frac{\lambda_{\min}}{2} ,$$

which completes the proof.

3 Proof of Theorem 2

Statement of Theorem: *Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$ be i.i.d. samples of $\mathbf{x} \sim TE(\tilde{\boldsymbol{\Sigma}}, \xi, \mathbf{f})$ for which the sign sub-*

Gaussian condition holds with constant κ. Define the constant

$$c_0 = \max\left\{\frac{320\kappa\pi^4\|\tilde{\mathbf{\Sigma}}\|_2^2}{\lambda_{\min}^2}, \frac{\pi^2}{\lambda_{\min}}\right\} ,$$

in which $\lambda_{\min}$ is the smallest eigenvalue $\tilde{\mathbf{\Sigma}}$. If $n \geq \frac{128c_0}{\lambda_{\min}} s\log p = O(s\log p)$, with probability at least $1 - \frac{2}{p} - \frac{1}{p^2}$, $\hat{\mathbf{\Sigma}}$ satisfies the following RE condition,

$$\inf_{\mathbf{v}\in\mathcal{C}\cap\mathbb{S}^{p-1}} \mathbf{v}^T\hat{\mathbf{\Sigma}}\mathbf{v} \geq \frac{\lambda_{\min}}{2} , \tag{S.5}$$

where $\mathcal{C}$ is defined in (21).

To prove Theorem 2, we first formally state below the convergence result for $\hat{\mathbf{\Sigma}}$ and $\tilde{\mathbf{\Sigma}}$ in [1].

Lemma A (Theorem 4.10 in [1]) *Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$ be i.i.d. samples of $\mathbf{x} \sim TE(\tilde{\mathbf{\Sigma}}, \xi, \mathbf{f})$ for which the sign sub-Gaussian condition holds with constant κ. With probability at least $1 - 2\alpha - \alpha^2$, $\hat{\mathbf{\Sigma}}$ constructed from $\mathbf{X}$ satisfies*

$$\|\hat{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}\|_{2,s_0} \leq \pi^2\left(\frac{s_0\log p}{n} + 2\sqrt{2\kappa}\|\tilde{\mathbf{\Sigma}}\|_2\sqrt{\frac{s_0\left(3+\log(p/s_0)\right)+\log(1/\alpha)}{n}}\right) , \tag{S.6}$$

where $\|\mathbf{A}\|_{2,s_0} \triangleq \sup_{\mathbf{v}\in\mathbb{S}^{p-1}, \|\mathbf{v}\|_0\leq s_0} \mathbf{v}^T\mathbf{A}\mathbf{v}$.

The next step for showing Theorem 2 is to extend the RE condition on all s_0-sparse unit vectors (s_0 needs to be appropriately specified) to all unit vectors inside the targeted descent cone $\mathcal{C}$. Lemma B accomplishes this goal.

Lemma B *Given $\hat{\mathbf{\Sigma}}$ constructed from $\mathbf{X}$ whose rows are generated from $\mathbf{x} \sim TE(\tilde{\mathbf{\Sigma}}, \xi, \mathbf{f})$, we assume that for every s_0-sparse unit vector $\mathbf{v}$, the condition $\mathbf{v}^T\hat{\mathbf{\Sigma}}\mathbf{v} \geq \mu$ is satisfied. Then we have for any $\mathbf{u} \in \mathcal{C}\cap\mathbb{S}^{p-1}$,*

$$\mathbf{u}^T\hat{\mathbf{\Sigma}}\mathbf{u} \geq \mu - \frac{4s}{s_0 - 1}(1-\mu) . \tag{S.7}$$

Proof: For any $\mathbf{u} \in \mathcal{C}\cap\mathbb{S}^{p-1}$, let $\mathbf{z} \in \mathbb{R}^p$ be a random vector defined by

$$\mathbb{P}\left(\mathbf{z} = \|\mathbf{u}\|_1 \text{sign}(u_i)\cdot\mathbf{e}_i\right) = \frac{|u_i|}{\|\mathbf{u}\|_1} , \tag{S.8}$$

where $\{\mathbf{e}_i\}_{i=1}^p$ is the canonical basis of $\mathbb{R}^p$. Therefore, $\mathbb{E}[\mathbf{z}] = \mathbf{u}$. Let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{s_0}$ be independent copies of $\mathbf{z}$ and set $\bar{\mathbf{z}} = \frac{1}{s_0}\sum_{i=1}^{s_0}\mathbf{z}_i$. Therefore $\bar{\mathbf{z}}$ is an s_0-sparse vector, and by our assumption on quadratic forms on s_0-sparse vectors

$$\bar{\mathbf{z}}^T\hat{\mathbf{\Sigma}}\bar{\mathbf{z}} \geq \mu\|\bar{\mathbf{z}}\|_2^2 \implies \mathbb{E}\left[\bar{\mathbf{z}}^T\hat{\mathbf{\Sigma}}\bar{\mathbf{z}}\right] \geq \mu\mathbb{E}\left[\|\bar{\mathbf{z}}\|_2^2\right] , \tag{S.9}$$

where the expectation is taken w.r.t $\bar{\mathbf{z}}$. Since $\bar{\mathbf{z}} = \frac{1}{s_0}\sum_{i=1}^{s_0}\mathbf{z}_i$, we have

$$\begin{aligned}
\mathbb{E}\left[\bar{\mathbf{z}}^T\hat{\mathbf{\Sigma}}\bar{\mathbf{z}}\right] &= \frac{1}{s_0^2}\sum_{1\leq i,j\leq s_0}\mathbb{E}\left[\mathbf{z}_i^T\hat{\mathbf{\Sigma}}\mathbf{z}_j\right] \\
&= \frac{1}{s_0^2}\sum_{\substack{1\leq i,j\leq s_0\\ i\neq j}}\mathbb{E}\left[\mathbf{z}_i^T\hat{\mathbf{\Sigma}}\mathbf{z}_j\right] + \frac{1}{s_0^2}\sum_{1\leq i\leq s_0}\mathbb{E}\left[\mathbf{z}_i^T\hat{\mathbf{\Sigma}}\mathbf{z}_i\right] \\
&= \frac{s_0(s_0-1)}{s_0^2}\mathbf{u}^T\hat{\mathbf{\Sigma}}\mathbf{u} + \frac{s_0}{s_0^2}\sum_{i=1}^p\frac{|u_i|}{\|\mathbf{u}\|_1}\|\mathbf{u}\|_1^2\hat{\sigma}_{ii} \\
&= \frac{s_0-1}{s_0}\mathbf{u}^T\hat{\mathbf{\Sigma}}\mathbf{u} + \frac{\|\mathbf{u}\|_1^2}{s_0} ,
\end{aligned}$$

since $\hat{\sigma}_{ii} = 1$, and $\sum_{i=1}^p \frac{|u_i|}{\|\mathbf{u}\|_1} = 1$. Replacing $\hat{\mathbf{\Sigma}}$ in the above expression by the identity matrix $\mathbf{I} \in \mathbb{R}^{p\times p}$, we have

$$\mathbb{E}\|\bar{\mathbf{z}}\|_2^2 = \frac{s_0-1}{s_0}\|\mathbf{u}\|_2^2 + \frac{\|\mathbf{u}\|_1^2}{s_0} .$$

Plugging both these expressions back in (S.9), we have

$$\begin{aligned}
&\frac{s_0-1}{s_0}\mathbf{u}^T\hat{\mathbf{\Sigma}}\mathbf{u} + \frac{\|\mathbf{u}\|_1^2}{s_0} \geq \mu\frac{s_0-1}{s_0}\|\mathbf{u}\|_2^2 + \mu\frac{\|\mathbf{u}\|_1^2}{s_0} \implies \\
&\mathbf{u}^T\hat{\mathbf{\Sigma}}\mathbf{u} \geq \mu\|\mathbf{u}\|_2^2 - \frac{\|\mathbf{u}\|_1^2}{s_0-1}(1-\mu) \geq \mu - \frac{4s}{s_0-1}(1-\mu) ,
\end{aligned}$$

where we use the facts that $\|\mathbf{u}\|_2 = 1$ and $\|\mathbf{u}\|_1 \leq 2\sqrt{s}$. That completes the proof. ■

Equipped with Lemma A and B, we present the proof of Theorem 2.

Proof of Theorem 2: For Lemma A, we set $\alpha = \frac{1}{p}$, $s_0 = \frac{16s}{\lambda_{\min}}$, and let $c_0 = \max\{\frac{320\kappa\pi^4\|\tilde{\mathbf{\Sigma}}\|_2^2}{\lambda_{\min}^2}, \frac{\pi^2}{\lambda_{\min}}\}$. When $n \geq \frac{128c_0}{\lambda_{\min}}s\log p = 8c_0 s_0\log p$, by Lemma A, we have

$$\begin{aligned}
\|\hat{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}\|_{2,s_0} &\leq \pi^2\Bigg(\frac{s_0\log p}{n} \\
&\qquad + 2\sqrt{2\kappa}\|\tilde{\mathbf{\Sigma}}\|_2\sqrt{\frac{s_0(3+\log(p/s_0))+\log p}{n}}\Bigg) \\
&\leq \pi^2\Bigg(\frac{s_0\log p}{\frac{\pi^2}{\lambda_{\min}}\cdot 8s_0\log p} \\
&\qquad + 2\sqrt{2\kappa}\|\tilde{\mathbf{\Sigma}}\|_2\sqrt{\frac{s_0(3+\log(p/s_0))+\log p}{\frac{320\kappa\pi^4\|\tilde{\mathbf{\Sigma}}\|_2^2}{\lambda_{\min}^2}\cdot 8s_0\log p}}\Bigg) \\
&\leq \pi^2\left(\frac{\lambda_{\min}}{\pi^2}\sqrt{\frac{5s_0\log p}{320s_0\log p}} + \frac{\lambda_{\min}}{\pi^2}\frac{s_0\log p}{8s_0\log p}\right) \\
&\leq \frac{\lambda_{\min}}{8} + \frac{\lambda_{\min}}{8} = \frac{\lambda_{\min}}{4} ,
\end{aligned}$$

with probability at least $1 - \frac{2}{p} - \frac{1}{p^2}$. It follows that for any s_0-sparse unit vector $\mathbf{v}$,

$$\mathbf{v}^T\hat{\boldsymbol{\Sigma}}\mathbf{v} \geq \mathbf{v}^T\tilde{\boldsymbol{\Sigma}}\mathbf{v} - \left|\mathbf{v}^T\left(\hat{\boldsymbol{\Sigma}} - \tilde{\boldsymbol{\Sigma}}\right)\mathbf{v}\right| \geq \lambda_{\min} - \|\hat{\boldsymbol{\Sigma}} - \tilde{\boldsymbol{\Sigma}}\|_{2,s_0} \geq \frac{3}{4}\lambda_{\min} ,$$

which satisfies the assumption in Lemma B with $\mu = \frac{3}{4}\lambda_{\min}$. With the same $s_0 = \frac{16s}{\lambda_{\min}}$, by Lemma B, we have for any $\mathbf{v} \in \mathcal{C} \cap \mathbb{S}^{p-1}$,

$$\begin{aligned}\mathbf{v}^T\hat{\boldsymbol{\Sigma}}\mathbf{v} &\geq \frac{3}{4}\lambda_{\min} - \frac{4s}{\frac{16s}{\lambda_{\min}} - 1}\left(1 - \frac{3}{4}\lambda_{\min}\right) \\ &\geq \frac{3}{4}\lambda_{\min} - \frac{4s}{\frac{16s}{\lambda_{\min}} - 12s}\left(1 - \frac{3}{4}\lambda_{\min}\right) \\ &= \frac{3}{4}\lambda_{\min} - \frac{4s}{\frac{16s}{\lambda_{\min}}(1 - \frac{3}{4}\lambda_{\min})}\left(1 - \frac{3}{4}\lambda_{\min}\right) \\ &= \frac{3}{4}\lambda_{\min} - \frac{\lambda_{\min}}{4} = \frac{\lambda_{\min}}{2} ,\end{aligned}$$

which completes the proof. ■

4 Proof of Theorem 3

Statement of Theorem: *Given any monotone cone $\mathcal{M}$, the following equality holds*

$$P_{\mathcal{M}\cap\mathcal{L}\cap\mathcal{B}}(\cdot) = P_{\mathcal{B}}(P_{\mathcal{L}}(P_{\mathcal{M}}(\cdot))) , \qquad \text{(S.10)}$$

where $P_{\mathcal{L}}(\mathbf{z}) = \mathbf{z} - \frac{\mathbf{1}^T\mathbf{z}}{n}\cdot\mathbf{1}$ *and* $P_{\mathcal{B}}(\mathbf{z}) = \min\{\frac{\sqrt{n}}{\|\mathbf{z}\|_2}, 1\}\cdot\mathbf{z}$.

Proof: It is easy to verify the the analytic expression for $P_{\mathcal{L}}(\cdot)$ and $P_{\mathcal{B}}(\cdot)$. To show (S.10), we let $\mathbf{x}^* = P_{\mathcal{M}}(\mathbf{z})$ and $\tilde{\mathbf{x}}^* = P_{\mathcal{M}\cap\mathcal{L}\cap\mathcal{B}}(\mathbf{z})$. We assume w.l.o.g. that the monotone cone is $\mathcal{M} = \{\mathbf{x} \mid x_1 \geq x_2 \geq \ldots \geq x_n\}$. By introducing the Lagrange multipliers $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{n-1}]^T$, the isotonic regression $P_{\mathcal{M}}(\mathbf{z})$ can be casted as

$$\max_{\boldsymbol{\lambda}\preceq\mathbf{0}} \min_{\mathbf{x}} \; g(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \sum_{i=1}^{n-1}\lambda_i(x_i - x_{i+1}) ,$$

where we use the strong duality. The optimum $\mathbf{x}^*$ has to satisfy the stationarity $\nabla_{\mathbf{x}}\, g(\mathbf{x}, \boldsymbol{\lambda}) = 0$, i.e.,

$$\begin{aligned} x_1^* - z_1 + \lambda_1 = 0 , \\ x_2^* - z_2 - \lambda_1 + \lambda_2 = 0 , \\ \vdots \\ x_{n-1}^* - z_{n-1} - \lambda_{n-2} + \lambda_{n-1} = 0 , \\ x_n^* - z_n - \lambda_{n-1} = 0 . \end{aligned} \qquad \text{(S.11)}$$

Using (S.11) to express $\mathbf{x}^*$ in terms of $\boldsymbol{\lambda}$, we denote $\min_{\mathbf{x}} g(\mathbf{x}, \boldsymbol{\lambda})$ by another function $h(\boldsymbol{\lambda})$, and the optimal dual variables $\boldsymbol{\lambda}^*$ satisfies

$$\boldsymbol{\lambda}^* = \operatorname*{argmax}_{\boldsymbol{\lambda}\preceq\mathbf{0}} \; h(\boldsymbol{\lambda}) .$$

For the standardized isotonic regression $P_{\mathcal{M}\cap\mathcal{L}\cap\mathcal{B}}(\mathbf{z})$, we can also introduce the Lagrange multipliers $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{n-1}]^T$, β and γ, and obtain the following optimization problem

$$\max_{\boldsymbol{\lambda}\preceq\mathbf{0},\gamma\leq 0,\beta} \min_{\mathbf{x}} \; \tilde{g}(\mathbf{x}, \boldsymbol{\lambda}, \beta, \gamma) = \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \sum_{i=1}^{n-1}\lambda_i(x_i - x_{i+1}) + \beta\sum_{i=1}^{n} x_i + \gamma(n - \|\mathbf{x}\|_2^2) . \qquad \text{(S.12)}$$

Again the optimum $\tilde{\mathbf{x}}^*$ has to satisfy $\nabla_{\mathbf{x}}\, \tilde{g}(\tilde{\mathbf{x}}^*, \boldsymbol{\lambda}, \beta, \gamma)$,

$$\begin{aligned} (1-2\gamma)\tilde{x}_1^* - z_1 + \beta + \lambda_1 = 0 , \\ (1-2\gamma)\tilde{x}_2^* - z_2 + \beta - \lambda_1 + \lambda_2 = 0 , \\ \vdots \\ (1-2\gamma)\tilde{x}_{n-1}^* - z_{n-1} + \beta - \lambda_{n-2} + \lambda_{n-1} = 0 , \\ (1-2\gamma)\tilde{x}_n^* - z_n + \beta - \lambda_{n-1} = 0 . \end{aligned} \qquad \text{(S.13)}$$

By substituting $\tilde{\mathbf{x}}^*$ for $\boldsymbol{\lambda}$, β and γ, we have

$$\begin{aligned} \min_{\mathbf{x}} \tilde{g}(\mathbf{x}, \boldsymbol{\lambda}, \beta, \gamma) &= \frac{1-2\gamma}{2}\sum_{i=1}^{n}\left(\tilde{x}_i^* - \frac{z_i - \beta}{1-2\gamma}\right)^2 \\ &+ \sum_{i=1}^{n-1}\lambda_i(\tilde{x}_i^* - \tilde{x}_{i+1}^*) + \frac{\|\mathbf{z}\|_2^2}{2} - \frac{\sum_{i=1}^{n}(z_i - \beta)^2}{2(1-2\gamma)} + \gamma n \\ &= \frac{h(\boldsymbol{\lambda})}{1-2\gamma} + \frac{\|\mathbf{z}\|_2^2}{2} - \frac{\sum_{i=1}^{n}(z_i - \beta)^2}{2(1-2\gamma)} + \gamma n , \end{aligned}$$

in which we note that the last three terms are free of $\boldsymbol{\lambda}$. Hence the optimal $\boldsymbol{\lambda}$ for standardized isotonic regression,

$$\begin{aligned} \tilde{\boldsymbol{\lambda}}^* &= \operatorname*{argmax}_{\boldsymbol{\lambda}\preceq\mathbf{0}} \; \frac{h(\boldsymbol{\lambda})}{1-2\gamma} + \frac{\|\mathbf{z}\|_2^2}{2} - \frac{\sum_{i=1}^{n}(z_i - \beta)^2}{2(1-2\gamma)} + \gamma n \\ &= \operatorname*{argmax}_{\boldsymbol{\lambda}\preceq\mathbf{0}} \; h(\boldsymbol{\lambda}) \end{aligned}$$

is the same as the one for isotonic regression. Thus, combining (S.11) and (S.13), we have

$$\tilde{\mathbf{x}}^* = \frac{\mathbf{x}^* - \beta\cdot\mathbf{1}}{1-2\gamma} . \qquad \text{(S.14)}$$

On the other hand, by summing up the equations respectively in (S.11) and (S.13) and using the primal feasibility $\sum_{i=1}^{n}\tilde{x}_i^* = 0$, we have

$$\sum_{i=1}^{n} x_i^* = \sum_{i=1}^{n} z_i, \;\; \sum_{i=1}^{n} z_i = n\beta \implies \beta = \frac{\mathbf{1}^T\mathbf{x}^*}{n} ,$$

which implies that

$$\mathbf{x}^* - \beta\cdot\mathbf{1} = P_{\mathcal{L}}(\mathbf{x}^*) = P_{\mathcal{L}}(P_{\mathcal{M}}(\mathbf{z})) . \qquad \text{(S.15)}$$

Denoting $\mathbf{x}^* - \beta\cdot\mathbf{1}$ by $\hat{\mathbf{x}}^*$, we now show that scaling $\hat{\mathbf{x}}^*$ by $\frac{1}{1-2\gamma}$ is exactly the projection onto $\mathcal{B}$. If $\|\hat{\mathbf{x}}^*\|_2 > \sqrt{n}$, then $\gamma < 0$ due to (S.14) and primal feasibility $\|\tilde{\mathbf{x}}^*\|_2 \leq \sqrt{n}$. By complementary slackness $\gamma(n - \|\tilde{\mathbf{x}}^*\|_2^2) = 0$, we have $\|\tilde{\mathbf{x}}^*\|_2 = \sqrt{n}$. If $\|\hat{\mathbf{x}}^*\|_2 < \sqrt{n}$, then $\|\tilde{\mathbf{x}}^*\| < \sqrt{n}$ due to (S.14) and dual feasibility $\gamma \leq 0$. It follows from complementary slackness that $\gamma = 0$, which result in $\tilde{\mathbf{x}}^* = \hat{\mathbf{x}}^*$. If $\|\hat{\mathbf{x}}^*\|_2 = \sqrt{n}$, by similar argument, we have $\tilde{\mathbf{x}}^* = \hat{\mathbf{x}}^*$ as well. In a word, we have

$$\tilde{\mathbf{x}}^* = \begin{cases} \hat{\mathbf{x}}^*, & \text{if } \|\hat{\mathbf{x}}^*\|_2 \leq \sqrt{n} \\ \frac{\sqrt{n}}{\|\hat{\mathbf{x}}^*\|_2}\hat{\mathbf{x}}^*, & \text{if } \|\hat{\mathbf{x}}^*\|_2 > \sqrt{n} \end{cases},$$

which matches the expression for $P_{\mathcal{B}}(\cdot)$. Thus we complete the proof by noting $\tilde{\mathbf{x}}^* = P_{\mathcal{B}}(\hat{\mathbf{x}}^*) = P_{\mathcal{B}}(P_{\mathcal{L}}(P_{\mathcal{M}}(\mathbf{z})))$. ■

References

[1] F. Han and H. Liu. Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distribution. *arXiv:1305.6916*, 2013.

[2] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[3] F. Lindskog, A. McNeil, and U. Schmock. Kendalls tau for elliptical distributions. In *Credit Risk*, Contributions to Economics, pages 149–156. Physica-Verlag HD, 2003.