Supplementary material: Sparsity and the truncated ℓ^2-norm

Lee H. Dicker
Department of Statistics and Biostatistics
Rutgers University
ldicker@stat.rutgers.edu

S1 PROOFS OF RESULTS FROM THE MAIN TEXT

S1.1 Proof of Theorem 1

We require a basic lemma on soft-threhsolding before proceeding with the proof of Theorem 1. Recall the soft-thresholding function $s_\lambda(x) = \mathrm{sign}(x)\{(|x| - \lambda) \vee 0\}$ ($x \in \mathbb{R}$, $\lambda \geq 0$) and define

$$r(\lambda; \theta) = E_\theta\left[\{s_\lambda(x) - \theta\}^2\right], \quad \theta \in \mathbb{R}, \ \lambda \geq 0$$

to be the risk of soft-thresholding in the 1-dimensional problem, where $x \sim N(\theta, 1)$. The following result is essentially contained in (Johnstone, 2013).

Lemma S1. *Let* $0 < \eta < 1$ *and* $\lambda_\eta = \{2\log(\eta^{-1})\}^{1/2}$. *Then*

$$r(\lambda_\eta; \theta) \leq \eta + \left[\theta^2 \wedge \{1 + 2\log(\eta^{-1})\}\right] \leq \left\{ \begin{array}{ll} \eta + \{1 + 2\log(\eta^{-1})\}^{1-p/2}|\theta|^p & \text{for all } 0 < p < 2, \\ \eta + \{1 + 2\log(\eta^{-1})\}(\theta^2 \wedge 1). & \end{array} \right. \tag{S1}$$

Proof of Lemma S1. The first inequality in (S1) follows immediately from Equations (8.7) and (8.12) in (Johnstone, 2013). For $0 < p < 2$, we have

$$\theta^2 \wedge \{1 + 2\log(\eta^{-1})\} = \{1 + 2\log(\eta^{-1})\}\left[\left\{\frac{\theta^2}{1 + 2\log(\eta^{-1})}\right\} \wedge 1\right] \leq \{1 + 2\log(\eta^{-1})\}^{1-p/2}|\theta|^p,$$

which yields the first part of the second inequality in (S1); the second part of the second inequality is obvious. □

Returning to the proof of Theorem 1, by (12) in the main text, it suffices to show that if $n \to \infty$ and $\eta \to 0$, then

$$R\{\hat{\boldsymbol{\theta}}_{\lambda_\eta}; B_n^t(\eta)\} \lesssim 2\eta\log(\eta^{-1}). \tag{S2}$$

Suppose that $\boldsymbol{\theta} \in B_n^t(\eta)$. Then, by Lemma S1,

$$R(\hat{\boldsymbol{\theta}}_{\lambda_\eta}; \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n r(\lambda_\eta; \theta_i) \leq \eta + \{1 + 2\log(\eta^{-1})\}\frac{1}{n}||\boldsymbol{\theta}||_t^2 \leq 2\eta\{1 + \log(\eta^{-1})\}. \tag{S3}$$

The (asymptotic) inequality (S2) follows, which proves the theorem.

S1.2 Proof of Theorem 2

Suppose that $\boldsymbol{\theta} \in \mathbb{R}^n$ and let $\eta_t(\boldsymbol{\theta}) = n^{-1}||\boldsymbol{\theta}||_t^2$. Then

$$\begin{aligned} R(\hat{\boldsymbol{\theta}}_{\hat{\lambda}}; \boldsymbol{\theta}) &= \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\hat{\lambda}} - \hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}}||^2 \right\} + \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}} - \boldsymbol{\theta}||^2 \right\} \\ &\quad + \frac{2}{n} E_{\boldsymbol{\theta}} \left[\left\{ \hat{\boldsymbol{\theta}}_{\hat{\lambda}} - \hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}} \right\}^\top \left\{ \hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}} - \boldsymbol{\theta} \right\} \right] \\ &\leq I_1(\boldsymbol{\theta}) + I_2(\boldsymbol{\theta}) + 2\{I_1(\boldsymbol{\theta}) I_2(\boldsymbol{\theta})\}^{1/2}, \end{aligned}$$

where

$$\begin{aligned} I_1(\boldsymbol{\theta}) &= \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\hat{\lambda}} - \hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}}||^2 \right\}, \\ I_2(\boldsymbol{\theta}) &= \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}} - \boldsymbol{\theta}||^2 \right\}. \end{aligned}$$

By Proposition 1 and Lemma S1,

$$\sup_{\boldsymbol{\theta} \in B_n^p(\eta)} I_2(\boldsymbol{\theta}) \lesssim \begin{cases} 2\eta \log(\eta^{-1}) & \text{if } p = t, \\ \eta\{2\log(\eta^{-1})\}^{1-p/2} & \text{if } 0 \leq p < 2. \end{cases}$$

Thus, by Theorem 1 and (9)–(10) in the main text, it suffices to prove

$$\sup_{\boldsymbol{\theta} \in B_n^p(\eta)} I_1(\boldsymbol{\theta}) = O(\eta), \quad p \in [0, 2) \cup \{t\} \tag{S4}$$

in order to prove Theorem 2.

Focusing on $I_1(\boldsymbol{\theta})$, we have the further decomposition,

$$I_1(\boldsymbol{\theta}) \leq J_1(\boldsymbol{\theta}) + J_2(\boldsymbol{\theta}) + 2\{J_1(\boldsymbol{\theta}) J_2(\boldsymbol{\theta})\}^{1/2}, \tag{S5}$$

where

$$\begin{aligned} J_1(\boldsymbol{\theta}) &= \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\hat{\lambda}} - \hat{\boldsymbol{\theta}}_{\lambda_{\kappa_t(\boldsymbol{\theta})}}||^2 \right\}, \\ J_2(\boldsymbol{\theta}) &= \frac{1}{n} E_{\boldsymbol{\theta}} \left\{ ||\hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}} - \hat{\boldsymbol{\theta}}_{\lambda_{\kappa_t(\boldsymbol{\theta})}}||^2 \right\} \end{aligned}$$

and $\kappa_t(\boldsymbol{\theta}) = 1 - n^{-1} \sum_{i=1}^n e^{-\theta_i^2/4}$. The quantities $J_1(\boldsymbol{\theta})$ and $J_2(\boldsymbol{\theta})$ both involve the difference of soft-thresholding estimators. Consider the following basic property of the soft-thresholding function $s_\lambda(x)$: if $0 \leq \lambda, \lambda'$ and $x \in \mathbb{R}$, then $|s_\lambda(x) - s_{\lambda'}(x)| \leq |\lambda' - \lambda|$; if additionally $x \leq \lambda \wedge \lambda'$, then $|s_\lambda(x) - s_{\lambda'}(x)| = 0$. Now define the set $A_{\boldsymbol{\theta}}(\rho) = \{i;\ |\theta_i| \geq \rho\}$, for $\rho \geq 0$, and let $A^c_{\boldsymbol{\theta}}(\rho) = \{1, ..., n\} \setminus A_{\boldsymbol{\theta}}(\rho)$. Define $a_{\boldsymbol{\theta}}(\rho) = |A_{\boldsymbol{\theta}}(\rho)|$ to be the number of elements in the set $A_{\boldsymbol{\theta}}(\rho)$. Focusing on $J_2(\boldsymbol{\theta})$ for the moment, we have

$$\begin{aligned} J_2(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}} \left[\{s_{\lambda_{\eta_t(\boldsymbol{\theta})}}(x_i) - s_{\lambda_{\kappa_t(\boldsymbol{\theta})}}(x_i)\}^2 \right] \\ &= \frac{1}{n} \sum_{i \in A_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}} \left[\{s_{\lambda_{\eta_t(\boldsymbol{\theta})}}(x_i) - s_{\lambda_{\kappa_t(\boldsymbol{\theta})}}(x_i)\}^2 \right] + \frac{1}{n} \sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}} \left[\{s_{\lambda_{\eta_t(\boldsymbol{\theta})}}(x_i) - s_{\lambda_{\kappa_t(\boldsymbol{\theta})}}(x_i)\}^2 \right] \\ &\leq \{\lambda_{\eta_t(\boldsymbol{\theta})} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2 \left[\frac{a_{\boldsymbol{\theta}}(\rho)}{n} + \frac{1}{n} \sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\eta_t(\boldsymbol{\theta})}\} \right] \\ &\leq 2 \left(\frac{3e+1}{e-1} \right)^2 \frac{1}{\log\{\eta_t(\boldsymbol{\theta})^{-1}\}} \left[\frac{\eta_t(\boldsymbol{\theta})}{\rho^2 \wedge 1} + \frac{1}{n} \sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\eta_t(\boldsymbol{\theta})}\} \right]. \end{aligned}$$

If $0 < \rho < \lambda$ and $i \in A^c_{\boldsymbol{\theta}}(\rho)$, then

$$P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda\} \leq \sqrt{\frac{2}{\pi}} \int_{\lambda-\rho}^{\infty} e^{-z^2/2}\, dz \leq \sqrt{\frac{2}{\pi}} \frac{e^{-\{\lambda-\rho\}^2/2}}{\lambda-\rho}. \tag{S6}$$

Taking $\lambda = \lambda_{\eta_t(\boldsymbol{\theta})}$ and $\rho = \log\{\lambda_{\eta_t(\boldsymbol{\theta})}\}/\lambda_{\eta_t(\boldsymbol{\theta})}$ above, we obtain

$$J_2(\boldsymbol{\theta}) \leq 2\left(\frac{3e+1}{e-1}\right)^2 \frac{1}{\log\{\eta_t(\boldsymbol{\theta})^{-1}\}} \left[\frac{\lambda^2_{\eta_t(\boldsymbol{\theta})}\eta_t(\boldsymbol{\theta})}{\log\{\lambda_{\eta_t(\boldsymbol{\theta})}\}^2} + \sqrt{\frac{2}{\pi}} \frac{\eta_t(\boldsymbol{\theta})\lambda_{\eta_t(\boldsymbol{\theta})}}{\lambda_{\eta_t(\boldsymbol{\theta})} - \log\{\lambda_{\eta_t(\boldsymbol{\theta})}\}/\lambda_{\eta_t(\boldsymbol{\theta})}}\right] = O\{\eta_t(\boldsymbol{\theta})\}. \tag{S7}$$

It remains to bound $J_1(\boldsymbol{\theta})$. The initial steps are similar to those for bounding $J_2(\boldsymbol{\theta})$, but now we must account for the fact that the thresholding level $\hat{\lambda}$ is random:

$$\begin{aligned}
J_1(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i \in A_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}}\left[\{s_{\hat{\lambda}}(x_i) - s_{\lambda_{\kappa_t(\boldsymbol{\theta})}}(x_i)\}^2\right] + \frac{1}{n}\sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}}\left[\{s_{\lambda_{\hat{\lambda}}}(x_i) - s_{\lambda_{\kappa_t(\boldsymbol{\theta})}}(x_i)\}^2\right] \\
&\leq \frac{a_{\boldsymbol{\theta}}(\rho)}{n} E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2\right] + \frac{1}{n}\sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})}\right] \\
&\leq \frac{\eta_t(\boldsymbol{\theta})}{\rho^2 \wedge 1} E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2\right] + \frac{1}{n}\sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})}\right].
\end{aligned}$$

By standard large deviations results,

$$E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2\right] \leq \frac{2}{\log\{\kappa_t(\boldsymbol{\theta})^{-1}\}} E_{\boldsymbol{\theta}}\left[\log\{(\hat{\kappa}_t \vee n^{-1})/\kappa_t(\boldsymbol{\theta})\}^2\right] = O\left[\frac{1}{n\kappa_t(\boldsymbol{\theta})^2 \log\{\kappa_t(\boldsymbol{\theta})^{-1}\}}\right].$$

Thus, it follows that

$$J_1(\boldsymbol{\theta}) \leq \frac{1}{n}\sum_{i \in A^c_{\boldsymbol{\theta}}(\rho)} E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})}\right] + O\left[\frac{1}{(\rho^2 \wedge 1)n\kappa_t(\boldsymbol{\theta})\log\{\kappa_t(\boldsymbol{\theta})^{-1}\}}\right] \tag{S8}$$

Now fix $c_1, c_2 > 0$. Then

$$\begin{aligned}
E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})}\right] &\leq E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})},\ \hat{\kappa}_t \leq \kappa_t(\boldsymbol{\theta}) - c_1\right] \\
&\quad + E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})},\ \kappa_t(\boldsymbol{\theta}) - c_1 < \hat{\kappa}_t \leq \kappa_t(\boldsymbol{\theta})\right] \\
&\quad + E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})},\ \kappa_t(\boldsymbol{\theta}) < \hat{\kappa}_t \leq \kappa_t(\boldsymbol{\theta}) + c_2\right] \\
&\quad + E_{\boldsymbol{\theta}}\left[\{\hat{\lambda} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i| \geq \hat{\lambda} \wedge \lambda_{\kappa_t(\boldsymbol{\theta})},\ \kappa_t(\boldsymbol{\theta}) + c_2 < \hat{\kappa}_t\right] \\
&\leq \lambda^2_{\text{univ}} P_{\boldsymbol{\theta}}\{\hat{\kappa}_t \leq \kappa_t(\boldsymbol{\theta}) - c_1\} \\
&\quad + \{\lambda_{\kappa_t(\boldsymbol{\theta})-c_1} - \lambda_{\kappa_t(\boldsymbol{\theta})}\}^2 P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\kappa_t(\boldsymbol{\theta})}\} \\
&\quad + \{\lambda_{\kappa_t(\boldsymbol{\theta})} - \lambda_{\kappa_t(\boldsymbol{\theta})+c_2}\}^2 P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\kappa_t(\boldsymbol{\theta})+c_2}\} \\
&\quad + \lambda^2_{\kappa_t(\boldsymbol{\theta})} P_{\boldsymbol{\theta}}\{\kappa_t(\boldsymbol{\theta}) + c_2 < \hat{\kappa}_t\} \\
&\leq \lambda^2_{\text{univ}} P_{\boldsymbol{\theta}}\{\hat{\kappa}_t \leq \kappa_t(\boldsymbol{\theta}) - c_1\} \\
&\quad + \frac{2c_1^2}{\{\kappa_t(\boldsymbol{\theta}) - c_1\}^2 \log\{\kappa_t(\boldsymbol{\theta})^{-1}\}} P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\kappa_t(\boldsymbol{\theta})}\} \\
&\quad + \frac{2c_2^2}{\kappa_t(\boldsymbol{\theta})^2 \log\{\kappa_t(\boldsymbol{\theta})^{-1}\}} P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\kappa_t(\boldsymbol{\theta})+c_2}\} \\
&\quad + \lambda^2_{\kappa_t(\boldsymbol{\theta})} P_{\boldsymbol{\theta}}\{\kappa_t(\boldsymbol{\theta}) + c_2 < \hat{\kappa}_t\}.
\end{aligned}$$

Now take $c_1 = c_2 = \{\log(n)/n\}^{1/2}$ and assume that $i \in A^c_{\boldsymbol{\theta}}(\rho)$, with $\rho = \log\{\lambda_{\kappa_t(\boldsymbol{\theta})+c_2}\}/\lambda_{\kappa_t(\boldsymbol{\theta})+c_2}$. Then, by (S6),

$$P_{\boldsymbol{\theta}}\{|x_i| \geq \lambda_{\kappa_t(\boldsymbol{\theta})+c_2}\} = O\{\eta_t(\boldsymbol{\theta})\}$$

and we conclude that
$$E_{\boldsymbol{\theta}}\left[\{\hat{\lambda}-\lambda_{\kappa_t(\boldsymbol{\theta})}\}^2;\ |x_i|\geq \hat{\lambda}\wedge\lambda_{\kappa_t(\boldsymbol{\theta})}\right]\leq O\{\eta_t(\boldsymbol{\theta})\}$$

Hence, combining this with (S8) implies
$$J_1(\boldsymbol{\theta}) = O\{\eta_t(\boldsymbol{\theta})\}. \tag{S9}$$

Finally, (S5), (S7), and (S9) together imply (S4), which proves the theorem.

S1.3 Proof of Theorem 3

Let
$$h(x)=\int_{-\infty}^{\infty}\phi(x-\theta)\ d\pi(\theta),$$

where $\phi(x)$ is the standard normal density. By Brown's identity (Brown, 1971; Bickel, 1981), the Bayes risk $r(\pi)$ satisfies
$$r(\pi)=1-\int_{-\infty}^{\infty}\frac{\{h'(x)\}^2}{h(x)}\ dx.$$

Since
$$h'(x)=-\int_{-\infty}^{\infty}(x-\theta)\phi(x-\theta)\ d\pi(\theta)$$

and
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\theta)^2\phi(x-\theta)\ d\pi(\boldsymbol{\theta})\ dx=1,$$

it follows that
$$r(\pi)=\int_{-\infty}^{\infty}\frac{1}{h(x)}\left[\int_{-\infty}^{\infty}\phi(x-\theta)\ d\pi(\theta)\int_{-\infty}^{\infty}(x-\theta)^2\phi(x-\theta)\ d\pi(\theta)-\left\{\int_{-\infty}^{\infty}(x-\theta)\phi(x-\theta)\ d\pi(\theta)\right\}^2\right]\ dx.$$

By Jensen's inequality,
$$\int_{-\infty}^{\infty}\phi(x-\theta)\ d\pi(\theta)\int_{-\infty}^{\infty}(x-\theta)^2\phi(x-\theta)\ d\pi(\theta)\geq\left\{\int_{-\infty}^{\infty}(x-\theta)\phi(x-\theta)\ d\pi(\theta)\right\}^2,\quad x\in\mathbb{R}.$$

Thus,
$$\begin{aligned}r(\pi)&\geq\sqrt{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\phi(x-\theta)\ d\pi(\theta)\int_{-\infty}^{\infty}(x-\theta)^2\phi(x-\theta)\ d\pi(\theta)-\left\{\int_{-\infty}^{\infty}(x-\theta)\phi(x-\theta)\ d\pi(\theta)\right\}^2\right]\ dx\\&=\sqrt{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{(x-\theta_1)^2-(x-\theta_1)(x-\theta_2)\}\phi(x-\theta_1)\phi(x-\theta_2)\ dx\ d\pi(\theta_1)\ d\pi(\theta_2).\end{aligned}$$

The inner integral above is
$$\int_{-\infty}^{\infty}\{(x-\theta_1)^2-(x-\theta_1)(x-\theta_2)\}\phi(x-\theta_1)\phi(x-\theta_2)\ dx=\frac{1}{4\sqrt{\pi}}(\theta_1-\theta_2)^2e^{-(\theta_1-\theta_2)^2/4},$$

which implies that
$$r(\pi)\geq\frac{1}{2\sqrt{2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\theta_1-\theta_2)^2e^{-(\theta_1-\theta_2)^2/4}\ d\pi(\theta_1)\ d\pi(\theta_2)=\frac{1}{2\sqrt{2}}E\left(Se^{-S/4}\right), \tag{S10}$$

where the random variable S has the same distribution as $(\theta_1-\theta_2)^2$, under the assumption that $\theta_1,\theta_2\sim\pi$ are independent.

Now we work to lower bound $E(Se^{-S/4})$. Since the function $t\mapsto te^{-t/4}$ is convex on $[8,\infty)$, Jensen's inequality implies that
$$E\left(Se^{-S/4}\right)=E\left(Se^{-S/4}\middle|\,S\leq 8\right)P(S\leq 8)+E\left(Se^{-S/4}\middle|\,S>8\right)P(S>8)$$

$$\geq e^{-2}E(S;S \leq 8) + E(S;S > 8)e^{-E(S|S>8)/4}.$$

Dividing the analysis into two case, first suppose that $E(S;S \leq 8) \geq 8P(S > 8)$. Then

$$E\left(Se^{-S/4}\right) \geq e^{-2}E(S;S \leq 8) \geq \frac{1}{2e^2}\left\{E(S;S \leq 8) + 8P(S > 8)\right\} = \frac{1}{2e^2}E(S \wedge 8).$$

Furthermore, since π is symmetric,

$$\begin{aligned}
E(S \wedge 8) &= \int_{\mathbb{R}^2} (\theta_1 - \theta_2)^2 \wedge 8 \; d\pi(\theta_1) \; d\pi(\theta_2) \\
&\geq \int_{\theta_1 \geq 0, \theta_2 \leq 0} (\theta_1 - \theta_2)^2 \wedge 8 \; d\pi(\theta_1) \; d\pi(\theta_2 + \int_{\theta_1 \leq 0, \theta_2 \geq 0} (\theta_1 - \theta_2)^2 \wedge 8 \; d\pi(\theta_1) \; d\pi(\theta_2 \\
&\geq 2\int_{\theta_1 \geq 0, \theta_2 \leq 0} (\theta_1^2 + \theta_2^2) \wedge 8 \; d\pi(\theta_1) \; d\pi(\theta_2) \\
&\geq 4\int_{\theta_1 \geq 0, \theta_2 \leq 0} \theta_1^2 \wedge 4 \; d\pi(\theta_1) \; d\pi(\theta_2) \\
&\geq \int_{\mathbb{R}} \theta_1^2 \wedge 4 \; d\pi(\theta_1) \\
&\geq \int_{\mathbb{R}} \theta_1^2 \wedge 1 \; d\pi(\theta_1) \\
&= \eta_t(\pi).
\end{aligned}$$

Thus,

$$E\left(Se^{-S/4}\right) \geq \frac{1}{2e^2}\eta_t(\pi). \tag{S11}$$

Now suppose that $E(S;S \leq 8) < 8P(S > 8)$. Then

$$E(S|S > 8) = \frac{E(S;S > 8)}{P(S > 8)} \leq \frac{16E(S)}{E(S \wedge 8)} \leq \frac{32\eta_2(\pi)}{\eta_t(\pi)}.$$

We conclude that

$$E\left(Se^{-S/4}\right) \geq E(S)e^{-E(S|S>8)/4} \geq 2\eta_2(\pi)e^{-8\eta_2(\pi)/\eta_t(\pi)}. \tag{S12}$$

The theorem follows from (S10)–(S12).

S1.4 Proof of Proposition 1

For $p = 0$, the result follows immediately from (S3), with $\eta_t(\boldsymbol{\theta})$ in place of η, and the fact that $||\boldsymbol{\theta}||_t^2 \leq ||\boldsymbol{\theta}||_0$. Suppose that $0 < p < 2$ and $\boldsymbol{\theta} \in B_n^p(\eta)$. Let $\eta_t(\boldsymbol{\theta}) = n^{-1}||\boldsymbol{\theta}||_t^2$ and $\eta_p(\boldsymbol{\theta}) = n^{-1}||\boldsymbol{\theta}||_p^p$. By Lemma S1,

$$\begin{aligned}
R\{\hat{\boldsymbol{\theta}}_{\lambda_{\eta_t(\boldsymbol{\theta})}};\boldsymbol{\theta}\} &= \frac{1}{n}\sum_{i=1}^n r\{\lambda_{\eta_t(\boldsymbol{\theta})}, \theta_i\} \\
&\leq \eta_t(\boldsymbol{\theta}) + \left[1 + 2\log\{\eta_t(\boldsymbol{\theta})^{-1}\}\right]\eta_t(\boldsymbol{\theta})\gamma_p(\boldsymbol{\theta}) \\
&\leq \eta + \left[1 + 2\log\{\eta_t(\boldsymbol{\theta})^{-1}\}\right]\eta_t(\boldsymbol{\theta})\gamma_p(\boldsymbol{\theta}),
\end{aligned}$$

where

$$\gamma_p(\boldsymbol{\theta}) = \left(\frac{\eta_p(\boldsymbol{\theta})}{[1 + 2\log\{\eta_t(\boldsymbol{\theta})^{-1}\}]^{p/2}\,\eta_t(\boldsymbol{\theta})}\right) \wedge 1.$$

It is straightforward to check that if $\eta \to 0$, then

$$\sup_{\boldsymbol{\theta} \in B_n^p(\eta)} \left[1 + 2\log\{\eta_t(\boldsymbol{\theta})^{-1}\}\right]\eta_t(\boldsymbol{\theta})\gamma_p(\boldsymbol{\theta}) \lesssim \eta\{2\log(\eta^{-1})\}^{1-p/2}.$$

The proposition follows.

References

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Stat.* **9** 1301–1309.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903.

Johnstone, I. M. (2013). *Gaussian Estimation: Sequence and Wavelet Models*. Monograph. Available at http://www-stat.stanford.edu/∼imj/GE06-11-13.pdf.