SUPPLEMENT TO BAYESIAN ESTIMATION OF SPARSE SIGNALS WITH A CONTINUOUS SPIKE-AND-SLAB PRIOR

By Veronika Ročková[*,†],

University of Chicago [†]

1. Supplemental Proofs of Theorems.

1.1. *Proofs of Section 4.* Throughout this section, we will be using the notation Δ^L and Δ^U introduced in (3.15). We also denote by

$$\delta_c = \frac{1}{\lambda_0 - \lambda_1} \log\left[\frac{1-\theta}{\theta}\frac{\lambda_0}{\lambda_1}\frac{c}{1-c}\right] \tag{1.1}$$

and

(1.2)

$$c_+ = \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{(\lambda_0-\lambda_1)^2}}\right) \quad \text{and} \quad c_- = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{(\lambda_0-\lambda_1)^2}}\right).$$

With this notation, $\delta_{c-} < \delta_{c+}$ are the two roots of the equation $pen''(\beta) = 1$.

1.1.1. *Proof of Theorem 3.1.* First, we assume that $L(\beta, y)$ has two modes. Denote by $\widehat{\beta}$ the global mode and by $\widetilde{\beta}$ the local mode. First, assume $\widehat{\beta} = 0$ and $(|y| - \lambda_1)^2 > 2\log[1/p^\star(0)]$. By (3.10) we have

$$y = \widetilde{\beta} + \text{sign}(y)\lambda^\star(\widetilde{\beta}). \tag{1.3}$$

Because $\widetilde{\beta}$ is a local maximum, the second derivative test $pen''(\widetilde{\beta}) < 1$ yields

$$p^\star(\widetilde{\beta})[1 - p^\star(\widetilde{\beta})] < 1/(\lambda_0 - \lambda_1)^2. \tag{1.4}$$

Because $\widetilde{\beta} \neq 0$, (1.4) implies that $|\widetilde{\beta}| > \delta_{c+} > 0$, where δ_{c+} is defined in (3.14). Since $\widehat{\beta} = 0$, we can write

$$y^2 < (y - \widetilde{\beta})^2 + 2\log\left(\frac{\pi(0)}{\pi(\widetilde{\beta})}\right). \tag{1.5}$$

[*]Veronika.Rockova@Chicagobooth.edu.

By noting

$$\frac{\pi(0)}{\pi(\widetilde{\beta})} = \frac{p^\star(\widetilde{\beta})}{p^\star(0)} \exp(\lambda_1 |\widetilde{\beta}|) \tag{1.6}$$

and invoking (1.3), (1.5) is equivalent to

$$y^2 < \lambda^\star(\widetilde{\beta})^2 + 2\lambda_1|y| - 2\lambda_1\lambda^\star(\widetilde{\beta}) + 2\log\left(\frac{p^\star(\widetilde{\beta})}{p^\star(0)}\right).$$

Rearranging the terms above, we obtain

$$(|y| - \lambda_1)^2 < g(\widetilde{\beta}) + 2\log[1/p^\star(0)], \tag{1.7}$$

where $g(\cdot)$ is defined in (3.13). We will need the following lemma.

LEMMA 1.1. *Assume* $(\lambda_0 - \lambda_1) > 2$. *The function* $g(x)$ *is symmetric and* $g(|x|)$ *is negative and monotone-increasing on* $[\delta_{c+};\infty)$. *Moreover,* $\lim_{|x|\to\infty} g(x) = 0$ *and* $g(|x|)$ *has a minimum at* δ_{c+} *if* $g(0) > 0$.

PROOF. Writing $g(x) = [1 - p^\star(x)]^2(\lambda_0 - \lambda_1)^2 + 2\log p^\star(x)$, the condition $g(x) < 0$ will be satisfied when

$$\log p^\star(x) < -\frac{[1 - p^\star(x)]^2}{2}(\lambda_0 - \lambda_1)^2. \tag{1.8}$$

Using $\log(x) < x - 1$ for $x < 1$, (1.8) will be satisfied when $p^\star(x) > 1 - \frac{2}{(\lambda_0 - \lambda_1)^2}$. Equating the derivative of $g(|x|)$ to zero yields

$$2[1 - p^\star(|x|)](\lambda_0 - \lambda_1)\left\{1 - p^\star(|x|)[1 - p^\star(|x|)](\lambda_0 - \lambda_1)^2\right\} = 0$$

We obtain the two solutions $\delta_{c-} < \delta_{c+}$ defined by (1.1) and (1.2). Because $|\widetilde{\beta}| > \delta_{c+} > 0$ and $pen''(|\widetilde{\beta}|) < 1$, we have $g'(|x|) > 0$ on $[\delta_{c+}, \infty)$. Since $p^\star(x) \geq c_+ > 1 - \frac{2}{(\lambda_0 - \lambda_1)^2}$ on $[\delta_{c+}, \infty)$, we have $g(x) < 0$ on $[\delta_{c+}, \infty)$. Because $g(|x|)$ is monotone increasing on $[\delta_{c+}, \infty)$, $g(|x|)$ has a local minimum at δ_{c+}. The value $g(\delta_{c+})$ will be a global minimum if $g(0) > 0$. Because $\lim_{|x|\to\infty} p^\star(x) = 1$, we obtain $\lim_{|x|\to\infty} g(x) = 0$. □

According to Lemma 1.1 we have $g(x) < 0$ for $|x| \geq \delta_c$. Because $|\widetilde{\beta}| > \delta_c$, we have $g(\widetilde{\beta}) < 0$ and (1.7) contradicts the assumption $(|y| - \lambda_1)^2 > 2\log[1/p^\star(0)]$. Therefore $\widehat{\beta}$ has to be nonzero. Next, assume $(|y| - \lambda_1)^2 < 2\log[1/p^\star(0)] - d$. Since $-d = g(\delta_{c+})$ is the minimum of the function $g(x)$, we obtain (1.7), which is equivalent to $\widehat{\beta} = 0$. The selection threshold can be written as $\Delta = \sqrt{2\log[1/p^\star(0)] + g(\widetilde{\beta})} + \lambda_1$. Note that $|\widetilde{\beta}| > \delta_{c+}$. By Lemma 1.1, $g(x)$ is negative and monotone increasing on $[\delta_{c+};\infty)$ with a minimum at δ_{c+}, we have $\Delta^L < \Delta \le \Delta^U$.

Now, assume that the posterior is unimodal. First, let $\widehat{\beta} \neq 0$ and $|y| < \Delta$. Then $|y| > \lambda^\star(0)$ and because $g(0) > 0$ we have $\lambda^\star(0) > \Delta$, which yields a contradiction. Next, assume $\widehat{\beta} = 0$ and $|y| > \Delta$. Denote by $x_0 \in (0, \delta_{c+})$ the root of $g(x) = 0$. Then, it is necessary that $|y| < \lambda^\star(x_0) + x_0$. We show that $\lambda^\star(x_0) + x_0 < \Delta$, yielding a contradiction. Because $g(x_0) = 0$, we have $\lambda^\star(x_0) + x_0 = \lambda_1 + \sqrt{2\log[1/p^\star(x_0)]} + x_0$. Define $h(x) = x + \sqrt{2\log[1/p^\star(x)]}$, then $h'(x_0) = 0$ and $h'(x) < 0$ for $x \in (0, x_0)$. Therefore $h(x_0) < h(0)$ and $\lambda^\star(x_0) + x_0 < \Delta$. □

LEMMA 1.2. *Let* $d = -g(\delta_{c+})$*. Then* $0 < d < 2 - \left(\frac{1}{\lambda_0-\lambda_1} - \sqrt{2}\right)^2$.

PROOF. Because $p^\star(\delta_{c+}) = c_+ = (1 - c_-)$. We can invoke the inequality $1-\sqrt{1-x^2} > x^2/2$ for $0 < x < 1$ to obtain $(\lambda_0-\lambda_1)^2(1-c_+)^2 > 1/(\lambda_0-\lambda_1)^2$. Because $c_+ > 1 - \frac{2}{(\lambda_0-\lambda_1)^2} > 0.5$ and $\log(1 - x^2) > -x$ for $0 < x < 0.5$, we have $\log c_+ > -\frac{\sqrt{2}}{\lambda_0-\lambda_1}$. Altogether, we have

$$g(\delta_{c+}) > \frac{1}{(\lambda_0-\lambda_1)^2} - \frac{\sqrt{2}}{\lambda_0-\lambda_1} = \left(\frac{1}{\lambda_0-\lambda_1} - \sqrt{2}\right)^2 - 2.$$

Lemma 1.1 yields $f(\delta_{c+}) < 0$, which completes the proof. □

1.2. *Proofs of Section 4.*

1.2.1. *Proof of Theorem 4.1.* The risk of the Spike-and-Slab LASSO estimator $\widehat{\boldsymbol{\beta}}$ (3.7) can be written as

$$R(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{p_n} \mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - \beta_{0i}]^2 + \sum_{i=p_n+1}^{n} \mathsf{E}_0[\widehat{\beta}_i^2], \tag{1.9}$$

where $p_n = ||\boldsymbol{\beta}_0||_0$. We begin by finding an upper bound to the first summand in (1.9). We have

$$\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - \beta_{0i}]^2 \le 2 + 2\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - Y_i]^2. \tag{1.10}$$

By Theorem 3.1 and because $|\widehat{\beta}_i| < |Y_i|$, we can write

$$\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - Y_i]^2 \le \mathsf{E}_{\beta_{0i}}[Y_i^2\mathbb{I}(|Y_i| \le \Delta)] + \mathsf{E}_{\beta_{0i}}[\lambda^\star(\widehat{\beta}_i)^2\mathbb{I}(|Y_i| > \Delta)].$$

Because $|\widehat{\beta}_i| > \delta_{c+}$, we have $p^\star(\widehat{\beta}_i) > c_+$ when $|Y_i| > \Delta$, where c_+ is defined in the proof of Lemma 3.1. Then, we have $\lambda^\star(\widehat{\beta}_i) < c_+(\lambda_1 - \lambda_0) + \lambda_0$ when $|Y_i| > \Delta$. Next, we obtain

$$\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - Y_i]^2 \leq \Delta^2 + [\lambda_1 + (1-c_+)(\lambda_0 - \lambda_1)]^2.$$

Because $c_+(1-c_+) = 1/(\lambda_0-\lambda_1)^2$ and $c_+ > 0.5$, we have $(1-c_+)(\lambda_0-\lambda_1) < 2/(\lambda_0 - \lambda_1)$. Since $\Delta^2 < 4\log[1/p^\star(0)] + 2\lambda_1^2$ and because $\lambda_1 < \mathrm{e}^{-2} < 1$ and $\lambda_0 - \lambda_1 > 2$, then

$$\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - Y_i]^2 \leq 4\log[1/p^\star(0)] + 2.$$

Together with (1.10), this yields

$$\mathsf{E}_{\beta_{0i}}[\widehat{\beta}_i - \beta_{0i}]^2 \leq 8[\log[1/p^\star(0)] + 1].$$

Now we continue with the second summand in (1.9). Denote by $\phi(x)$ the standard normal density. Because $|\widehat{\beta}_i| < |Y_i|$ and $\Delta^L < \Delta \leq \Delta^U$, we can write

$$\mathsf{E}_0[\,\widehat{\beta}_i^2\,] < \mathsf{E}_0[Y_i^2\,\mathbb{I}(|Y_i| > \Delta)] < \mathsf{E}_0[Y_i^2\,\mathbb{I}(|Y_i| > \Delta^L)] < 4\,\Delta^L\phi(\Delta^L), \tag{1.11}$$

where the last inequality uses Mills's ratio (as in [**?**]). Because $\Delta^L > \sqrt{2\log[1/p^\star(0)] - d}$, we have $\phi(\Delta^L) < 1/\sqrt{2\pi}\mathrm{e}^{d/2}p^\star(0)$. Because $\Delta^L < \sqrt{2\log[1/p^\star(0)]} + \lambda_1$ and $\lambda_1 < \mathrm{e}^{-2} < \sqrt{2}$, (1.11) can be further simplified as

$$\mathsf{E}_0[\,\widehat{\beta}_i^2\,] < \frac{4\,\mathrm{e}^{d/2}}{\sqrt{\pi}}p^\star(0)[\sqrt{\log[1/p^\star(0)]} + 1].$$

Altogether, we obtain

$$\mathsf{E}_{\boldsymbol{\beta}_0}||\boldsymbol{\beta} - \boldsymbol{\beta}_0||^2 < 8\,p_n\,(1 + \log\,[1/p^\star(0)]) + 4/\sqrt{\pi}(n-p_n)\mathrm{e}^d p^\star(0)(1 + \sqrt{\log\,[1/p^\star(0)]}). \tag{1.12}$$

1.2.2. *Proof of Theorem 4.2.*

PROOF. Without loss of generality, we will assume $\frac{1-\theta}{\theta}\frac{\lambda_0}{\lambda_1} = n^{\alpha+\nu}/\lambda_1$. Denote by $Q(\boldsymbol{\beta}) = -\frac{1}{2}||\boldsymbol{y}^{(n)} - \boldsymbol{\beta}||^2 + \log\pi(\boldsymbol{\beta}\,|\,\theta)$, where $\pi(\boldsymbol{\beta}\,|\,\theta)$ is the SSL prior arising from (2.1) and (2.2). Using the global optimality $0 \geq Q(\boldsymbol{\beta}_0) - Q(\widehat{\boldsymbol{\beta}})$, we can write

$$0 \geq ||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||^2 - 2\boldsymbol{\varepsilon}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + 2\,\log\left(\frac{\pi(\boldsymbol{\beta}_0\,|\,\theta)}{\pi(\widehat{\boldsymbol{\beta}}\,|\,\theta)}\right). \tag{1.13}$$

Now, we focus on a set $\tau_0 = \left\{\boldsymbol{y}^{(n)} : ||\boldsymbol{y}^{(n)} - \boldsymbol{\beta}_0||_\infty \leq \bar{\lambda}\right\}$, where $\bar{\lambda} = 2\sqrt{\log n}$. This set has a large probability under the generative model, i.e. $\mathsf{P}(\tau_0) \geq 1 - \frac{2}{n}$ [**?**]. Conditioning on the set τ_0, we use the Hölder inequality $|\boldsymbol{\alpha}'\boldsymbol{\beta}| \leq |\boldsymbol{\alpha}|_\infty ||\boldsymbol{\beta}||_1$ to find that

$$0 \geq ||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||^2 - 2\,\bar{\lambda}||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_1 + 2\,\log\left(\frac{\pi(\boldsymbol{\beta}_0 \mid \theta)}{\pi(\widehat{\boldsymbol{\beta}} \mid \theta)}\right). \tag{1.14}$$

Denote by $\boldsymbol{\Delta} \equiv \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. Using the fact $||\boldsymbol{\Delta}||_1 \leq ||\boldsymbol{\Delta}||\,||\boldsymbol{\Delta}||_0^{1/2}$, we have

$$0 \geq ||\boldsymbol{\Delta}||^2 - 2\,\bar{\lambda}||\boldsymbol{\Delta}||\,||\boldsymbol{\Delta}||_0^{1/2} + 2\,\log\left(\frac{\pi(\boldsymbol{\beta}_0 \mid \theta)}{\pi(\widehat{\boldsymbol{\beta}} \mid \theta)}\right). \tag{1.15}$$

Because $p^\star(\widehat{\beta}_j) > c_+$ whenever $\widehat{\beta}_j \neq 0$, we can write

$$\begin{aligned}\log\left[\frac{\pi(\boldsymbol{\beta}_0 \mid \theta)}{\pi(\widehat{\boldsymbol{\beta}} \mid \theta)}\right] &\geq -\lambda_1||\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}||_1 + \sum_{j=1}^n \log\left[\frac{p^\star(\widehat{\beta}_j)}{p^\star(\beta_{0j})}\right]\\ &\geq -\lambda_1||\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}||_1 + \widehat{p}_n\,b + (\widehat{p}_n - p_n)\log[1/p^\star(0)],\end{aligned}$$

where $0 > b \equiv \log c_+ > \log 0.5$ is a constant very close to 0 and where $\lim_{\lambda_0\to\infty} b = 0$. To continue with (1.15), we can write

$$\left[||\boldsymbol{\Delta}|| - (\bar{\lambda} + \lambda_1)||\boldsymbol{\Delta}||_0^{1/2}\right]^2 + 2\,(\widehat{p}_n - p_n)\log[1/p^\star(0)] \leq (\bar{\lambda} + \lambda_1)^2||\boldsymbol{\Delta}||_0 - 2\,\widehat{p}_n\,b. \tag{1.16}$$

With (1.16) and using the fact $||\boldsymbol{\Delta}||_0 \leq \widehat{p}_n + p_n$, we obtain

$$2\,(\widehat{p}_n - p_n)\log[1/p^\star(0)] \leq (\bar{\lambda} + \lambda_1)^2(\widehat{p}_n + p_n) - 2\,\widehat{p}_n\,b$$

which is equivalent to writing

$$\widehat{p}_n \leq p_n\,(1 + C)\,,$$

where

$$C = \frac{2(\bar{\lambda} + \lambda_1)^2}{2\log[1/p^\star(0)] - (\bar{\lambda} + \lambda_1)^2 + 2\,b}$$

and where $C = \mathcal{O}(1)$ under the given choice of hyper-parameters. Moreover, with $\lambda_1 < \mathrm{e}^{-1}$ and $\frac{1-\theta}{\theta}\frac{\lambda_0}{\lambda_1} = n^{\alpha+\nu}/\lambda_1$, we have $2\log[1/p^\star(0)] - 2(\bar{\lambda} + \lambda_1)^2 + 2\,b > 2(\alpha + \nu - 4 - 4\lambda_1)\log n + 2\log(1/\lambda_1) - 2\lambda_1^2 + 2\,b > 0$ for some suitable $c > 0$. This inequality yields $0 < C < 2$.

What remains to be shown is that $p_n \leq \widehat{p}$ under the suitable beta-min condition. We prove this statement by contradiction. Assume $\widehat{p}_n < p_n$ and denote by $q_n = p_n - \widehat{p}_n > 0$. We continue with (1.15) and write

$$0 \geq ||\mathbf{\Delta}|| \left[||\mathbf{\Delta}|| - 2\,(\bar{\lambda} + \lambda_1)||\mathbf{\Delta}||_0^{1/2} \right] - 2\,q_n \log[1/p^{\star}(0)] + 2\,\widehat{p}_n\, b.$$

Because $||\mathbf{\Delta}||_0 < (1 + C)\,p_n$, we can write

$$0 \geq ||\mathbf{\Delta}|| \left[||\mathbf{\Delta}|| - 2\,(\bar{\lambda} + \lambda_1)\sqrt{(1 + C)\,p_n} \right] - 2\,q_n \log[1/p^{\star}(0)] + 2\,\widehat{p}_n\, \log 0.5. \tag{1.17}$$

Assuming the minimal-strength condition $|\beta_{0i}| > b_0 \geq D\sqrt{p_n \log n}$ when $\beta_{0i} \neq 0$, we can write

$$\frac{1}{2}||\mathbf{\Delta}|| > \frac{\sqrt{q_n}}{2}\,b_0 \geq \frac{D}{2}\,\sqrt{q_n\,p_n \log n} > 2\,(\bar{\lambda} + \lambda_1)\sqrt{p_n(1 + C)}$$

for a suitably large $D > 0$. Assuming $\frac{1-\theta}{\theta}\frac{\lambda_0}{\lambda_1} = n^{\alpha+\nu}/\lambda_1$, (1.17) yields a contradiction

$$\begin{aligned}
0 &\geq \frac{1}{2}||\mathbf{\Delta}||^2 - 2\,q_n \log\left(1 + n^{\alpha+\nu}/\lambda_1\right) + 2\,\widehat{p}_n\, \log 0.5 \\
&> \frac{D^2}{2} q_n\,p_n\, \log n - 2\,q_n \log\left(1 + n^{\alpha+\nu}/\lambda_1\right) + 2\,\widehat{p}_n\, \log 0.5 > 0
\end{aligned}$$

for D sufficiently large. □

Veronika Ročková
Booth School of Business
University of Chicago
369 Carles M. Harper Center
5807 S. Woodlawn Avenue
Chicago, IL 60637