Appendix to "A Lasso-Type Approach for Estimation and Variable Selection in Single Index Models" published in the Journal of Computational and Graphical Statistics

Peng Zeng
Department of Mathematics and Statistics, Auburn University

Tianhong He
Department of Statistics, Purdue University

Yu Zhu
Department of Statistics, Purdue University

Proof of Proposition 1. Let $Y_j = (y_1 - \tilde{y}_j, y_2 - \tilde{y}_j, \ldots, y_n - \tilde{y}_j)^T$, $X_j = (x_1 - \tilde{x}_j, x_2 - \tilde{x}_j, \ldots, x_n - \tilde{x}_j)^T$, $W_j = \mathrm{diag}(w_{1j}, w_{2j}, \ldots, w_{nj})$. Then the target function

$$L_\lambda(b, \theta) = \sum_{j=1}^n \sum_{i=1}^n (y_i - \tilde{y}_j - b_j \theta^T (x_i - \tilde{x}_j))^2 w_{ij} + \lambda \sum_{j=1}^n |b_j| \sum_{k=1}^p |\theta_k|$$

can be re-written as

$$L(b, \theta) = \sum_{j=1}^n \Big[(Y_j - X_j b_j \theta)^T W_j (Y_j - X_j b_j \theta) + \lambda \sum_{k=1}^p |b_j \theta_k| \Big].$$

We have

$$\begin{aligned} L(b, \theta) &= \sum_{j=1}^n \Big[Y_j^T W_j Y_j - 2 b_j \theta^T X_j^T W_j Y_j + b_j \theta^T X_j^T W_j X_j b_j \theta + \lambda \sum_{k=1}^p |b_j \theta_k| \Big] \\ &\geq \sum_{j=1}^n \Big[Y_j^T W_j Y_j - 2 b_j \theta^T X_j^T W_j Y_j + \lambda \sum_{k=1}^p |b_j \theta_k| \Big] \end{aligned}$$

The above inequality follows from the fact that $b_j \theta^T X_j^T W_j X_j b_j \theta \geq 0$. Recall that $R_j = X_j^T W_j Y_j$. Therefore,

$$L(b, \theta) \geq \sum_{j=1}^n \Big[Y_j^T W_j Y_j + \sum_{k=1}^p (\lambda - 2 r_{jk} \mathrm{sgn}(b_j \theta_k)) |b_j \theta_k| \Big],$$

where $\mathrm{sgn}(\cdot)$ is the sign function. Further recall that $r_{11} = \max |r_{jk}|$. When $\lambda \geq 2|r_{11}|$, $\lambda - 2 r_{jk} \mathrm{sgn}(b_j \theta_k) \geq 0$ for all j and k, and thus $L(b, \theta) \geq \sum_{j=1}^n Y_j^T W_j Y_j$. The minimum is achieved at $b_j = 0$ and arbitrary θ. Therefore, $b_j(\lambda) = 0$ and θ is not identifiable when $\lambda \geq 2|r_{11}|$.

Assume that the second largest $|r_{jk}|$ is $|r_{j_2 k_2}|$. Consider $\lambda \in (2|r_{j_2 k_2}|, 2|r_{11}|)$. Following argument similar to the above, we conclude that $b_j(\lambda) = 0$ for $j \geq 2$. Therefore, we only need to minimize

$$L_1(b_1, \theta) = (Y_1 - X_1 b_1 \theta)^T W_1 (Y_1 - X_1 b_1 \theta) + \lambda \sum_{k=1}^p |b_1 \theta_k|,$$

which is the Lasso with weight least squares. Applying existing results of the regular Lasso or using the KKT conditions, it is easy to show that there exists $\delta > 0$ such that for $\lambda \in (2|r_{11}|-\delta, 2|r_{11}|)$, the minimizer of $L_1(b_1,\theta)$ is $b_1\theta_1 = [\sum_{i=1}^n (x_{i1}-\tilde{x}_{i1})^2 w_{i1}]^{-1}[r_{11}-\mathrm{sgn}(r_{11})\lambda/2]$ and $b_1\theta_k = 0$ for $2 \le k \le p$. Therefore,

$$\hat{b}_1(\lambda) = [\sum_{i=1}^n (x_{i1}-\tilde{x}_{i1})^2 w_{i1}]^{-1}[r_{11}-\mathrm{sgn}(r_{11})\lambda/2], \quad \text{and} \quad \theta(\lambda) = (1,0,0,\ldots,0)^T.$$

Proof of Proposition 2. Let b_j^+ and b_j^- denote the positive and negative parts of b_j ($1 \le j \le n$) and let θ_k^+ and θ_k^- denote the positive and negative parts of θ_k ($1 \le k \le p$). Let $\theta^+ = (\theta_1^+, \theta_2^+, \ldots, \theta_p^+)^T$, $\theta^- = (\theta_1^-, \theta_2^-, \ldots, \theta_p^-)^T$, $b^+ = (b_1^+, b_2^+, \ldots, b_n^+)^T$, and $b^- = (b_1^-, b_2^-, \ldots, b_n^-)^T$. Then, $\theta = \theta^+ - \theta^-$ and $b = b^+ - b^-$. The sim-lasso problem can be re-written as to minimize

$$L(b,\theta) = Q(b,\theta) + \lambda \sum_{j=1}^n (b_j^+ + b_j^-) \sum_{k=1}^p (\theta_k^+ + \theta_k^-) \tag{1}$$

subject to $\theta_k^+ \ge 0$, $\theta_k^- \ge 0$, $b_j^+ \ge 0$, $b_j^- \ge 0$, $\theta_k^+\theta_k^- = 0$ and $b_j^+ b_j^- = 0$ for $1 \le k \le p$ and $1 \le j \le n$. It is not difficult to verify that the last two constraints will be satisfied by the minimizer of (1) automatically and thus are redundant. Therefore, they are dropped in the discussion below. The KKT conditions given below characterize the minimizer of (1).

$$\frac{\partial L}{\partial b_j^+} - u_j^+ = \frac{\partial Q}{\partial b_j} + \lambda \sum_{k=1}^p |\theta_k| - u_j^+ = 0$$
$$\frac{\partial L}{\partial b_j^-} - u_j^- = -\frac{\partial Q}{\partial b_j} + \lambda \sum_{k=1}^p |\theta_k| - u_j^- = 0$$

$$\frac{\partial L}{\partial \theta^+} - v^+ = \frac{\partial Q}{\partial \theta} + \lambda \sum_{j=1}^n |b_j| - v^+ = 0$$
$$\frac{\partial L}{\partial \theta^-} - v^- = -\frac{\partial Q}{\partial \theta} + \lambda \sum_{j=1}^n |b_j| - v^- = 0$$

where u_j^+, u_j^-, $v^+ = (v_1^+, \ldots, v_k^+)$, and $v^- = (v_1^-, \ldots, v_k^-)$ are Lagrange (dual) variables such that $u_j^+ \ge 0$, $u_j^- \ge 0$,$v_k^+ \ge 0$, $v_k^- \ge 0$, $u_j^+ b_j^+ = u_j^- b_j^- = 0$, $v_k^+\theta_k^+ = v_k^-\theta_k^- = 0$. For any $j \in \mathcal{B}$, $b_j \ne 0$, which implies either $b_j^+ \ne 0$ or $b_j^- \ne 0$. If $b_j^+ \ne 0$, then $u_j^+ = 0$; if $b_j^- \ne 0$, then $u_j^- = 0$. In either case, the following equation holds

$$\frac{\partial Q}{\partial b_j} = -\mathrm{sgn}(b_j)\lambda \sum_{k=1}^p |\theta_k|.$$

When both $u_j^+ > 0$ and $u_j^- > 0$, then,

$$|\frac{\partial Q}{\partial b_j}| < \lambda \sum_{k=1}^p |\theta_k|.$$

This implies that $b_j = 0$ and $j \in \mathcal{B}^c$. Whenever b_j with $j \in \mathcal{B}$ becomes zero or u_j^+ or u_j^- with $j \in \mathcal{B}^c$ becomes zero, the corresponding data point need to be removed from $\mathcal{B}$ or added to $\mathcal{B}$ for reconsideration. Combining these facts together gives (iii) and (iv) of the proposition. Similarly we can obtain (i) and (ii) of the proposition.

Proof of Proposition 3. We have

$$\frac{\partial Q}{\partial \theta} = -2\sum_{j=1}^{n}\sum_{i=1}^{n}(y_i - \tilde{y}_j - b_j\theta^T(x_i - \tilde{x}_j))w_{ij}b_j(x_i - \tilde{x}_j) = -2\sum_{j=1}^{n} b_j R_j + 2\sum_{j=1}^{n} b_j^2 S_j \theta$$

and

$$\frac{\partial Q}{\partial b_j} = -2\sum_{i=1}^{n}(y_i - \tilde{y}_j - b_j\theta^T(x_i - \tilde{x}_j))w_{ij}\theta^T(x_i - \tilde{x}_j) = -2R_j^T\theta + 2b_j\theta^T S_j \theta$$

for $1 \le j \le n$. Suppose at a fixed λ which is not a transition value, the active sets are $\mathcal{A}$ and $\mathcal{B}$. From Proposition 2, we have

$$-2\sum_{j\in\mathcal{B}} b_j R_{j\mathcal{A}} + 2\sum_{j\in\mathcal{B}} b_j^2 S_{j\mathcal{A}}\theta_{\mathcal{A}} + \mathrm{sgn}(\theta_{\mathcal{A}})\lambda\sum_{j\in\mathcal{B}}|b_j| = 0; \tag{2}$$

and for $j \in \mathcal{B}$,

$$-2R_{j\mathcal{A}}^T\theta_{\mathcal{A}} + 2b_j\theta^T S_j\theta + \mathrm{sgn}(b_j)\lambda\sum_{k\in\mathcal{A}}|\theta_k| = 0. \tag{3}$$

When λ is between two consecutive transition values, $\theta_{\mathcal{A}}$ and b_j's do not change signs. Differentiating the two sides of (2) and (3) with respect to λ, we have

$$\sum_{j\in\mathcal{B}} b_j^2 S_{j\mathcal{A}}\frac{\partial\theta_{\mathcal{A}}}{\partial\lambda} + \sum_{j\in\mathcal{B}} U_{j\mathcal{A}}\frac{\partial b_j}{\partial\lambda} + \frac{1}{2}\mathrm{sgn}(\theta_{\mathcal{A}})\sum_{j\in\mathcal{B}}|b_j| = 0$$

and for $j \in \mathcal{B}$,

$$U_{j\mathcal{A}}^T\frac{\partial\theta_{\mathcal{A}}}{\partial\lambda} + \theta_{\mathcal{A}}^T S_{j\mathcal{A}}\theta_{\mathcal{A}}\frac{\partial b_j}{\partial\lambda} + \frac{1}{2}\mathrm{sgn}(b_j)\sum_{k\in\mathcal{A}}|\theta_k| = 0,$$

where

$$U_{j\mathcal{A}} = -R_{j\mathcal{A}} + 2b_j S_{j\mathcal{A}}\theta_{\mathcal{A}} + \lambda\mathrm{sgn}(b_j)\mathrm{sgn}(\theta_{\mathcal{A}})/2.$$

Solving the above two equations for the derivatives of $\theta_{\mathcal{A}}$ and $b_{\mathcal{B}}$ leads to the formulas given in the proposition.