Biometrics 000, 000–000 DOI: 000
000 0000

Regularized Estimation for the Accelerated Failure Time Model"

T. Cai[1,*], J. Huang[2,], and L. Tian[2,***]**

[1]Department of Biostatistics, Harvard University, Boston, MA 02115, USA

[2]Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA.

**email:* tcai@hsph.harvard.edu

***email:* jjhuang@northwestern.edu

****email:* lutian@northwestern.edu

Figure 1. The estimated loss (as a function of the penalty parameter λ used in the LASSO regularization) based on five independent replicates of 5-fold cross validations in the breast cancer example

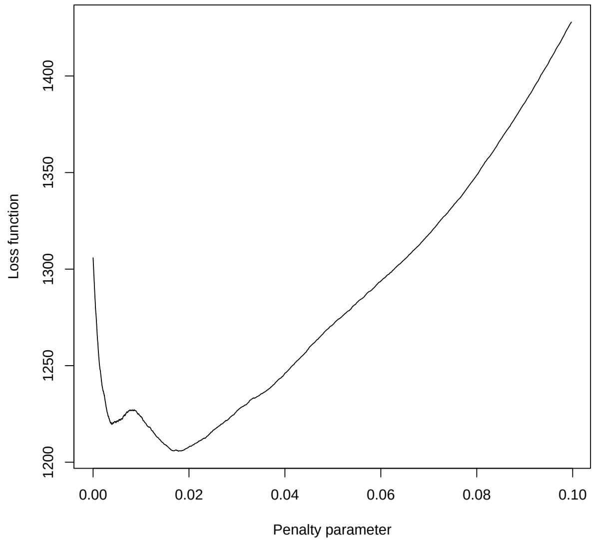
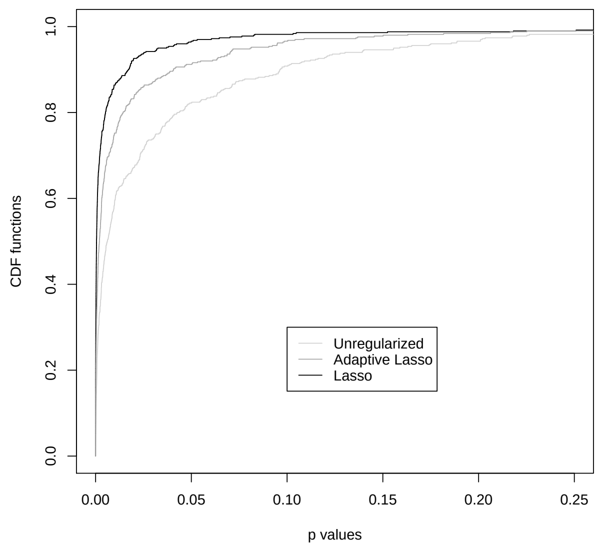


Figure 2. Empirical cumulative distribution functions of p values from the logrank test in comparing high and low risk groups in the validation sample.



Appendix

For the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{LG}$ and $\widehat{\boldsymbol{\beta}}_{AG}$, we require the same assumptions that are required in Tsiatis (1990). Briefly, the censoring C is assumed to be independent of T conditional on $\mathbf{Z}$. $E(\mathbf{Z}'\mathbf{Z}) < \infty, \mathbb{A}_0 > 0$ and the true parameter $\boldsymbol{\beta}_0$ is an interior point of a compact parameter space Ω. The error distribution has a finite Fisher information and the conditional distribution of $\mathbf{Z}|\delta = 1$ does not concentrate on a low dimensional hyperplane in R^p. Throughout, the dimension p is assumed to be a fixed constant.

Appendix A: Asymptotic Properties of $\widehat{\boldsymbol{\beta}}_{LG}$

Throughout, we assume that $n^{\frac{1}{2}}\lambda_n \to \lambda_0 \geqslant 0$ and $\boldsymbol{\beta}_0$ is an inner point of to a compact parameter space Ω. Let $\ell(\boldsymbol{\beta}) = E\{L(\boldsymbol{\beta})\}$ be the limit of $L(\boldsymbol{\beta})$ and $L_{\lambda_n}(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) + \lambda_n \sum_{k=1}^{p} |\beta_k|$. It follows from the fact that $\lambda_n \to 0$ and a uniform law of large numbers for U-processes (Honore and Powell, 1994) that $L_{\lambda_n}(\boldsymbol{\beta}) \to \ell(\boldsymbol{\beta})$ almost surely, uniformly in $\boldsymbol{\beta} \in \Omega$. This, together with the uniqueness of $\operatorname{argmin}_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$ and Corollary 1 of (Honore and Powell, 1994), implies that $\widehat{\boldsymbol{\beta}}_{LG} \to \boldsymbol{\beta}_0$ almost surely.

To derive the limiting distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{LG} - \boldsymbol{\beta}_0)$, we follow similar arguments as given in Knight and Fu (2000) and define

$$
\begin{aligned}
V_n(\mathbf{u}) &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \left| \epsilon_{ij}^{\star} - n^{-\frac{1}{2}} \mathbf{u}'\mathbf{Z}_{ij} \right|_{+} - \left| \epsilon_{ij}^{\star} \right|_{+} \right\} + \lambda_n \sum_{k=1}^{p} \left\{ \left| \beta_{0k} + n^{-\frac{1}{2}} u_k \right| - |\beta_{0k}| \right\} \\
&= n \left\{ L(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - L(\boldsymbol{\beta}_0) \right\} + n\lambda_n \sum_{k=1}^{p} \left\{ \left| \beta_{0k} + n^{-\frac{1}{2}} u_k \right| - |\beta_{0k}| \right\}
\end{aligned}
$$

where $|a|_+ = aI(a \geqslant 0)$, $\epsilon_i^{\star} = \epsilon_i \wedge (C_i - \boldsymbol{\beta}_0'\mathbf{Z}_i)$, $\epsilon_{ij}^{\star} = \epsilon_i^{\star} - \epsilon_j^{\star}$ and $\mathbf{Z}_{ij} = \mathbf{Z}_i - \mathbf{Z}_j$. Then $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{LG} - \boldsymbol{\beta}_0)$ is the minimizer of $V_n(\mathbf{u})$. To derive the limiting distribution of $n^{\frac{1}{2}} \left\{ L(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - L(\boldsymbol{\beta}_0) \right\}$, we note that by a functional central limit theorem for U-processes (Nolan and Pollard, 1988), $n^{\frac{1}{2}} \{\mathbf{S}_n(\mathbf{b}) - \mathbf{s}(\mathbf{b})\}$ converges weakly to a zero mean Gaussian process in $\mathbf{b}$, where $\mathbf{s}(\mathbf{b}) = E\{\mathbf{S}_n(\mathbf{b})\}$. Therefore, $n^{\frac{1}{2}}\mathbf{S}_n(\boldsymbol{\beta}_0)$ converges in distribution to $\mathbf{W} \sim N(0, \mathbb{B}_0)$. Furthermore, for any given finite vector $\mathbf{u}$,

$$
n^{\frac{1}{2}} \left\{ \mathbf{S}_n(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - \mathbf{s}(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) \right\} - n^{\frac{1}{2}}\mathbf{S}_n(\boldsymbol{\beta}_0) = o_p(1) \tag{A.1}
$$

and thus $n^{\frac{1}{2}}\{\mathbf{S}_n(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - n^{\frac{1}{2}}\mathbf{S}_n(\boldsymbol{\beta}_0)\} \approx \mathbf{u}'\mathbb{A}_0$. It follows that

$$
n \left\{ L(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - L(\boldsymbol{\beta}_0) \right\} = \mathbf{u}' \left\{ n^{\frac{1}{2}}\mathbf{S}_n(\boldsymbol{\beta}_0) \right\} + \frac{1}{2}\mathbf{u}'\mathbb{A}_0\mathbf{u} + o_p(1) \tag{A.2}
$$

which converges in distribution to $\mathbf{u}'\mathbf{W} + \frac{1}{2}\mathbf{u}'\mathbb{A}_0\mathbf{u}$. This, together with

$$n\lambda_n \sum_{k=1}^{p} \left\{ \left|\beta_{0k} + n^{-\frac{1}{2}} u_k\right| - |\beta_{0k}| \right\} \to \lambda_0 \sum_{k=1}^{p} \{u_j \mathrm{sgn}(\beta_{0k}) I(\beta_{0k} \neq 0) + |u_j| I(\beta_{0k} = 0)\}$$

implies that $V_n(\mathbf{u}) \to V(\mathbf{u})$ in distribution for any fixed $\mathbf{u}$. Since V_n is convex and V has a unique minimum, it follows from the same arguments as given in Knight and Fu (2000) by invoking the convergence properties of random convex functions that

$$\operatorname*{argmin}_{\mathbf{u}} V_n(\mathbf{u}) = n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{LG} - \boldsymbol{\beta}_0) \to \operatorname*{argmin}_{\mathbf{u}} V(\mathbf{u}) \quad \text{in distribution.}$$

Appendix B: Asymptotic Properties of $\widehat{\boldsymbol{\beta}}_{AG}$

Under the assumption that $|\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0| = O_p(n^{-\frac{1}{2}})$, $\widetilde{\lambda}_n \to 0$ and $n^{\frac{1}{2}}\widetilde{\lambda}_n \to \infty$ as $n \to \infty$, it is straightforward to show that $L(\boldsymbol{\beta}) + \widetilde{\lambda}_n \sum_{k=1}^{p} |\beta_k|/|\widehat{\beta}_{Gk}| \to \ell(\boldsymbol{\beta})$ almost surely, uniformly in $\boldsymbol{\beta}$. It then follows from the same arguments as given for the consistency of $\widehat{\boldsymbol{\beta}}_{LG}$ that $\widehat{\boldsymbol{\beta}}_{AG} \to \boldsymbol{\beta}_0$ almost surely.

To establish the oracle property for $\widehat{\boldsymbol{\beta}}_{AG}$, we follow steps similar to the proofs given in Zou (2006) and Zhang and Lu (2007). We first prove the asymptotic normality part. Let

$$\begin{aligned} D_n(\mathbf{u}) &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \left| \epsilon_{ij}^{\star} - n^{-\frac{1}{2}} \mathbf{u}'(\mathbf{Z}_{ij}) \right|_+ - \left| \epsilon_{ij}^{\star} \right|_+ \right\} + n^{\frac{1}{2}} \widetilde{\lambda}_n \sum_{k=1}^{p} \frac{\left|\beta_{0k} + n^{-\frac{1}{2}} u_k\right| - |\beta_{0k}|}{|\widehat{\beta}_{Gk}|} \\ &= n \left\{ L(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\mathbf{u}) - L(\boldsymbol{\beta}_0) \right\} + n^{\frac{1}{2}} \widetilde{\lambda}_n \sum_{k=1}^{p} \frac{\left|\beta_{0k} + n^{-\frac{1}{2}} u_k\right| - |\beta_{0k}|}{|\widehat{\beta}_{Gk}|} \end{aligned}$$

Then $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{AG} - \boldsymbol{\beta}_0)$ is the unique minimizer of $D_n(\mathbf{u})$. It follows from (A.2) that

$$D_n(\mathbf{u}) = \mathbf{u}' \left\{ n^{\frac{1}{2}} \mathbf{S}_n(\boldsymbol{\beta}_0) \right\} + \frac{1}{2}\mathbf{u}'\mathbb{A}_0\mathbf{u} + n^{\frac{1}{2}} \widetilde{\lambda}_n \sum_{k=1}^{p} \frac{\left|\beta_{0k} + n^{-\frac{1}{2}} u_k\right| - |\beta_{0k}|}{|\widehat{\beta}_{Gk}|} + o_p(1)$$

For the third term in the right handside of the above expression, we note that if $k \in \mathcal{A}$, $\widehat{\beta}_{Gk}^{-1} \to \beta_{0k}^{-1}$ in probability and $n^{\frac{1}{2}}(|\beta_{0k} + n^{-\frac{1}{2}}u_k| - |\beta_{0k}|) \to u_k \mathrm{sgn}(\beta_{0k})$. By Slutsky's Theoremm, $\widetilde{\lambda}_n n^{\frac{1}{2}}(|\beta_{0k} + n^{-\frac{1}{2}}u_k| - |\beta_{0k}|)/|\widehat{\beta}_{Gk}| \to 0$ as $n \to \infty$. If $k \in \mathcal{A}_c = \{k' : \beta_{0k'} = 0\}$, then $n^{\frac{1}{2}}(|\beta_{0k} + n^{-\frac{1}{2}}u_k| - |\beta_{0k}|) = |u_j|$, $\widetilde{\lambda}_n/|\widehat{\beta}_{Gk}| = n^{\frac{1}{2}}\widetilde{\lambda}_n/|n^{\frac{1}{2}}\widehat{\beta}_{Gk}| \to \infty$. Therefore, for every $\mathbf{u}$,

$$D_n(\mathbf{u}) = \begin{cases} \left(\mathbf{u}^{\mathcal{A}}\right)' \left\{ n^{\frac{1}{2}} \mathbf{S}_n(\boldsymbol{\beta}_0)^{\mathcal{A}} \right\} + \frac{1}{2} \left(\mathbf{u}^{\mathcal{A}}\right)' \mathbb{A}_1 \mathbf{u}^{\mathcal{A}} + o_p(1) & \text{if } \mathbf{u}^{\mathcal{A}_c} = 0 \\ O_p\left(\sum_{k \in \mathcal{A}_c} \widetilde{\lambda}_n |\widehat{\beta}_{Gk}|^{-1} \right) & \text{otherwise} \end{cases}$$

which converges weakly to

$$D(\mathbf{u}) = \begin{cases} \left(\mathbf{u}^{\mathcal{A}}\right)' \mathbf{W}^{\mathcal{A}} + \frac{1}{2} \left(\mathbf{u}^{\mathcal{A}}\right)' \mathbb{A}_1 \mathbf{u}^{\mathcal{A}} & \text{if } \mathbf{u}^{\mathcal{A}_c} = 0 \\ \infty & \text{otherwise} \end{cases}$$

It follows from the the convexity of $D_n(\mathbf{u})$ and the fact that $D(\mathbf{u})$ attains its unique minimum at $\mathbf{u} = \binom{\mathbb{A}_1^{-1}\mathbf{W}^{\mathcal{A}}}{\mathbf{0}}$. Following the epi-convergence results in **?**) and Knight and Fu (2000), we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{AG}^{\mathcal{A}} - \boldsymbol{\beta}_0^{\mathcal{A}}) \to N(0, \mathbb{A}_1^{-1}\mathbb{B}_1\mathbb{A}_1^{-1}) \quad \text{and} \quad n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{AG}^{\mathcal{A}_c} - \boldsymbol{\beta}_0^{\mathcal{A}_c}) \to \mathbf{0} \quad \text{in distribution.}$$

We next show $P(\widehat{\mathcal{A}} = \mathcal{A}) \to 1$ as $n \to \infty$. For any $k \in \mathcal{A}$, the almost sure convergence of $\widehat{\beta}_{AGk} \to \beta_{0k}$ implies that $P(k \in \widehat{\mathcal{A}}) \to 1$. Thus, it suffices to show that for any $k_c \in \mathcal{A}_c$, $P(k_c \in \widehat{\mathcal{A}}) \to 0$. Consider the event $k_c \in \widehat{\mathcal{A}}$. The Karush-Kuhn-Tucker optimality conditions (Osborne et al, 2000; Efron et al., 2004) implies that

$$n^{\frac{1}{2}} S_{nk_c}(\widehat{\boldsymbol{\beta}}_{AG}) = -\widetilde{\lambda}_n |\widehat{\beta}_{Gk_c}|^{-1} \operatorname{sgn}(\widehat{\beta}_{AGk_c}) + o_p(1).$$

By (A.1) and weak convergence of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{AG} - \boldsymbol{\beta}_0)$, we have

$$n^{\frac{1}{2}} S_{nk_c}(\widehat{\boldsymbol{\beta}}_{AG}) = n^{\frac{1}{2}} S_{nk_c}(\boldsymbol{\beta}_0) + \mathbb{A}_{k_c \cdot} n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{AG} - \boldsymbol{\beta}_0) + o_p(1) = O_p(1)$$

where $\mathbb{A}_{k_c \cdot}$ is the k_cth row of $\mathbb{A}$. On the other hand, $\widetilde{\lambda}_n |\widehat{\beta}_{Gk_c}|^{-1} \to \infty$ and therefore

$$P(k_c \in \widehat{\mathcal{A}}) \leqslant P\left\{ n^{\frac{1}{2}} S_{nk_c}(\widehat{\boldsymbol{\beta}}_{AG}) = -\widetilde{\lambda}_n |\widehat{\beta}_{Gk_c}|^{-1} \operatorname{sgn}(\widehat{\beta}_{AGk_c}) + o_p(1) \right\} \to 0.$$

This implies $P(\widehat{\mathcal{A}} = \mathcal{A}) \to 1$ as $n \to \infty$.

Appendix C: Algorithm for Computing the Exact LASSO Path

In the initialization stage, we let $\boldsymbol{\beta}^{[1]} = 0$ and compute the derivative of $L(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ at $\boldsymbol{\beta} = \boldsymbol{\beta}^{[1]}$ as

$$\dot{\mathbf{L}}^{[1]} = n^{-2} \sum_{n \geqslant i > j > 1} \{(\mathbf{Z}_j - \mathbf{Z}_i) I(Y_i > Y_j)\delta_j + (\mathbf{Z}_i - \mathbf{Z}_j) I(Y_j > Y_i)\delta_i\}.$$

Then, the direction of the next move is set to be $\dot{\boldsymbol{\beta}}^{[1]} = -\operatorname{sgn}(\dot{L}_{k_m}^{[1]})\boldsymbol{\alpha}(k_m)$, where $k_m = \operatorname{argmax}_k |\dot{L}_k^{[1]}|$ and $\boldsymbol{\alpha}(k)$ is a p-dimensional vector with the kth element being 1 and all other elements being 0. Furthermore, we set $\lambda^{[1]} = (\dot{\boldsymbol{\beta}}^{[1]})'\dot{\mathbf{L}}^{[1]}$. In summary, $\boldsymbol{\beta}^{[1]}$, $\dot{\mathbf{L}}^{[1]}$, $\dot{\boldsymbol{\beta}}^{[1]}$ and $\lambda^{[1]}$ are the initial joint, the derivative of the objective function at the initial joint, the direction for the next move, and the penalty parameter corresponding to the current constraint, respectively.

In the iteration stage, for the current values of $\{\boldsymbol{\beta}^{[k]}, \dot{\mathbf{L}}^{[k]}, \dot{\boldsymbol{\beta}}^{[k]}, \lambda^{[k]}\}$, find the smallest step size ϵ such that either one of the following two events happens:

(a) exists a pair (i_k, j_k) such that $e_{i_k}(\boldsymbol{\beta}^{[k]}) \neq e_{j_k}(\boldsymbol{\beta}^{[k]})$ and $e_{i_k}(\boldsymbol{\beta}^{[k]} + \epsilon\dot{\boldsymbol{\beta}}^{[k]}) = e_{j_k}(\boldsymbol{\beta}^{[k]} + \epsilon\dot{\boldsymbol{\beta}}^{[k]})$.

(b) exists l_k such that $\beta_{l_k}^{[k]} \neq 0$ and $\beta_{l_k}^{[k]} + \epsilon\dot{\beta}_{l_k}^{[k]} = 0$.

We then update $\boldsymbol{\beta}^{[k]}$ as $\boldsymbol{\beta}^{[k+1]} = \boldsymbol{\beta}^{[k]} + \epsilon\dot{\boldsymbol{\beta}}^{[k]}$, and let $\mathcal{L}^{[k+1]} = \{(i,j) \mid e_i(\boldsymbol{\beta}^{[k+1]}) = e_j(\boldsymbol{\beta}^{[k+1]}), i > j\}$, $\mathcal{S}^{[k+1]} = \{l \mid \beta_l^{[k+1]} = 0\}$ and $\mathcal{S}^{[k+1]c}$ be the complement of $\mathcal{S}$. Note that typically the total size of $\mathcal{S}$ and $\mathcal{L}$ is p. To determine the direction of the next move, we first search for all possible candidate directions. For any given direction $\dot{\mathbf{b}}$, we determine the descending rate of $L(\boldsymbol{\beta})$ along the direction of $\dot{\mathbf{b}}$ relative to the change in $\sum_{k=1}^{p} |\beta_k|$. The final direction for the next move is the direction among all candidate directions with the highest relative descending rate. Specifically,

- for any $(i_0, j_0) \in \mathcal{L}$, we solve the equation for $\dot{\boldsymbol{\beta}}$:

$$\dot{\beta}_l = 0, \text{ for all } l \in \mathcal{S}^{[k+1]}, \qquad \sum_{l \in \mathcal{S}^{[k+1]c}} \text{sgn}(\beta_l^{[k+1]})\dot{\beta}_l = 1,$$
$$\dot{\boldsymbol{\beta}}'(\mathbf{Z}_i - \mathbf{Z}_j) = 0, \text{ for all } (i,j) \in \mathcal{L}^{[k+1]} - \{(i_0, j_0)\}$$

  and obtain the solution $\dot{\boldsymbol{\beta}}_{(i_0,j_0)}^{[k+1]}$ as a candidate direction for the next move. The relative descending rate of $L(\boldsymbol{\beta})$ along the direction of $\dot{\boldsymbol{\beta}}_{(i_0,j_0)}^{[k+1]}$ is $\lambda_{(i_0,j_0)}^{[k+1]} = (\dot{\boldsymbol{\beta}}^{[k+1]})'\dot{\mathbf{L}}_{(i_0,j_0)}^{[k+1]}$. Here, the derivative $\dot{\mathbf{L}}_{(i_0,j_0)}^{[k+1]}$ is updated such that

  – if event (a) happens in the foregoing update, then

$$\dot{\mathbf{L}}_{(i_0,j_0)}^{[k+1]} = \dot{\mathbf{L}}^{[k]} + \{\boldsymbol{\tau}_{i_0,j_0}(\dot{\boldsymbol{\beta}}_{(i_0,j_0)}^{[k+1]}) - \boldsymbol{\eta}_{i_k,j_k}(\boldsymbol{\beta}^{[k]})\}/n^2,$$
$$\text{where} \quad \boldsymbol{\tau}_{i,j}(\mathbf{b}) = (\mathbf{Z}_j - \mathbf{Z}_i)I(\mathbf{Z}_i'\mathbf{b} < \mathbf{Z}_j'\mathbf{b})\delta_j + (\mathbf{Z}_i - \mathbf{Z}_j)I(\mathbf{Z}_i'\mathbf{b} > \mathbf{Z}_j'\mathbf{b})\delta_i$$
$$\text{and} \quad \boldsymbol{\eta}_{i,j}(\boldsymbol{\beta}) = (\mathbf{Z}_j - \mathbf{Z}_i)I\{e_i(\boldsymbol{\beta}) > e_j(\boldsymbol{\beta})\}\delta_j + (\mathbf{Z}_i - \mathbf{Z}_j)I\{e_i(\boldsymbol{\beta}) < e_j(\boldsymbol{\beta})\}\delta_i.$$

  – if event (b) happens in the foregoing update, then $\dot{\mathbf{L}}_{(i_0,j_0)}^{[k+1]} = \dot{\mathbf{L}}^{[k]} + \boldsymbol{\tau}_{i_0,j_0}(\dot{\boldsymbol{\beta}}_{(i_0,j_0)}^{[k+1]})/n^2$.

- for any $l_0 \in \mathcal{S}^{[k+1]}$, we obtain $\dot{\boldsymbol{\beta}}_{(l_0)}^{[k+1]}$, the solution to

$$\dot{\beta}_l = 0, \text{ for all } l \in \mathcal{S} - \{l_0\}, \qquad |\dot{\beta}_{l_0}| + \sum_{l \in \mathcal{S}^{[k+1]c}} \text{sgn}(\beta_l^{[k+1]})\dot{\beta}_l = 1,$$
$$\dot{\boldsymbol{\beta}}'(\mathbf{Z}_i - \mathbf{Z}_j) = 0, \text{ for all } (i,j) \in \mathcal{L}^{[k+1]}$$

  as a candidate direction for the next move. The relative descending rate of $L(\boldsymbol{\beta})$ along $\dot{\boldsymbol{\beta}}_{(l_0)}^{[k+1]}$ is $\lambda_{(l_0)}^{[k+1]} = (\dot{\boldsymbol{\beta}}_{(l_0)}^{[k+1]})'\dot{\mathbf{L}}_{(l_0)}^{[k+1]}$, where

  – if event (a) happens in the foregoing update, then

$$\dot{\mathbf{L}}_{(l_0)}^{[k+1]} = \dot{\mathbf{L}}^{[k]} - \boldsymbol{\eta}_{i_k,j_k}(\boldsymbol{\beta}^{[k]})/n^2.$$

– if event (b) happens in the foregoing update, then

$$\dot{\mathbf{L}}_{(l_0)}^{[k+1]} = \dot{\mathbf{L}}^{[k]}.$$

The iteration stops if the relative descending rates along all the directions considered above are positive. Otherwise, let $\dot{\boldsymbol{\beta}}^{[k+1]}$ denote the direction with the fastest relative descending rate among $\{\widehat{\boldsymbol{\beta}}_{(i_0,j_0)}^{[k+1]}, (i_0, j_0) \in \mathcal{S}^{[k+1]}; \widehat{\boldsymbol{\beta}}_{(l_0)}^{[k+1]}, l_0 \in \mathcal{L}^{[k+1]}\}$. Let $\dot{\mathbf{L}}^{[k+1]}$ and $\lambda^{[k+1]}$ be the derivative and descending rate corresponding to the selected direction. In the aforementioned algorithm, we did not consider the degenerated case that more than p hyperplanes from the set $\{Y_i - Y_j - \boldsymbol{\beta}'(\mathbf{Z}_i - \mathbf{Z}_j) = 0, \beta_i = 0 \mid 1 \leqslant i < j \leqslant n\}$ intersect at the same point in the parameter space., i.e., the total sizes of $\mathcal{S}^{k+1}$ and $\mathcal{L}^{[k+1]}$ is greater than p. In practice, this can be avoided by randomly perturbing the data by an arbitrarily small amount.

References

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R.t (2004). Least angle regression The *Annals of Statistics*, **32**, 407-99.

Honore, B. and Powell, J. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, **64**, 241-78.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics* **28**, 1356-1378.

Nolan, D. and Pollard, D. (1988). Functional liit theorems for U-processes *The Annals of Probability* **16**, 1291-1298.

Osborne, M., Presnell, B. and Turlach, B. (2000) On the LASSO and its dual. *Journal of Computational and Graphical Statistics*, **9**, 319-337

Tsiatis, A. (1996). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18**, 305-328.

Zhang, H. and Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94**, 691-703.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418-1429.