Supplementary Appendix for “Sufficient Dimension Reduction for Censored Regression”

**Proof of Lemma 1:** Given \( \text{E}(X) = 0 \), one can show that

\[
\beta = \Sigma^{-1} \text{E}\{\text{E}(b(Y)X|Y)\} = \Sigma^{-1} \text{E}\{b(Y)X\} = \Sigma^{-1} \text{Cov}\{b(Y), X\}.
\]

We next show that \( \beta \in \mathcal{S}_{Y|X} \). The linearity condition implies that \( \text{E}(X|\Gamma^T X) = P_\Sigma X \), where \( P_\Sigma = \Gamma (\Gamma^T \Sigma \Gamma)^{-1} \Gamma^T \Sigma \) is the projection operator onto \( \mathcal{S}_{Y|X} = \text{Span}(\Gamma) \) with respect to the \( \Sigma \) inner product. Then,

\[
\beta = \Sigma^{-1} \text{E}\{b(Y)X\} = \Sigma^{-1} \text{E}\{\text{E}(b(Y)X|\Gamma^T X, Y)\} = \Sigma^{-1} \text{E}\{b(Y)\text{E}\{X|\Gamma^T X, Y\}\} = \Sigma^{-1} \text{E}\{b(Y)P_\Sigma X\} = P_\Sigma \beta,
\]

where the first equality on the second line holds because, by definition of the central subspace, \( Y \perp X|\Gamma^T X \), so that \( \text{E}\{X|\Gamma^T X, Y\} = \text{E}(X|\Gamma^T X) \). Consequently, the vector \( \beta \) belongs to the range of the projection operator \( P_\Sigma \), i.e., \( \mathcal{S}_{Y|X} \). \( \square \)

**Proof of Theorem 1:** As noted in Section 2.2, \( \hat{\beta}_k \) is the solution to the estimating equation \( U_k(\beta) = 0 \). Under the regularity conditions and the uniform consistency of \( \hat{G}_{X_i}(\cdot) \) assumed in (C.3), one can show that \( n^{-1}U_k(\beta) \) converges almost surely to a deterministic vector of function \( u_k(\beta) \) as \( n \to \infty \), with \( u_k(\beta_k) = 0 \). In addition, we have that \( -n^{-1}\partial U_k(\beta)/\partial \beta = n^{-1} \sum_{i=1}^n \delta_i X_i X_i^T / \hat{G}_{X_i}(\hat{Y}_i) \), which is positive definite and converges uniformly to a positive definite deterministic matrix \( A = -\partial u_k(\beta)/\partial \beta = \text{E}(XX^T) \). Thus \( \hat{\beta}_k \) is the unique solution of the above weighted least squares equation and converges to the true \( \beta_k \) almost surely as \( n \to \infty \). Then by the Taylor expansion
and the law of large numbers, we have \( \sqrt{n}(\hat{\beta}_k - \beta_k) = A^{-1}n^{-1/2}U_k(\beta_k) + o_p(1) \), where

\[
\begin{align*}
n^{-1/2}U_k(\beta_k) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}(Y_i)} X_i \left\{ b_k(\bar{Y}_i) - \beta_k^T X_i \right\} - \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}^2(Y_i)} X_i \left\{ b_k(\bar{Y}_i) - \beta_k^T X_i \right\} n^{1/2} \left\{ \hat{G}_{X_i}(\bar{Y}_i) - G_{X_i}(\bar{Y}_i) \right\} + o_p(1).
\end{align*}
\]

By the asymptotic representation assumed in (C.3), we have that the second term on the right-hand side of the above equation can be represented as

\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \int_0^\tau \phi_{1k}(s)\psi_2(s)dM_j(s) + \phi_{2k}\psi_{4j} \right\} + o_p(1),
\]

where \( \phi_{1k}(s) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}(Y_i)} X_i \left\{ b_k(\bar{Y}_i) - \beta_k^T X_i \right\} \psi_1(\bar{Y}_i, X_i)I(\bar{Y}_i \geq s) \) and \( \phi_{2k} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}^2(Y_i)} X_i \left\{ b_k(\bar{Y}_i) - \beta_k^T X_i \right\} \psi_3(\bar{Y}_i, X_i) \). Thus, we have \( U_k(\beta_k) = \sum_{i=1}^{n} u_{ik} \), where

\[
u_{ik} = \frac{\delta_i}{G_{X_i}(Y_i)} X_i \left\{ b_k(\bar{Y}_i) - \beta_k^T X_i \right\} - \left\{ \int_0^\tau \phi_{1k}(s)\psi_2(s)dM_i(s) + \phi_{2k}\psi_{4i} \right\}.
\]

Then it follows from the multivariate central limit theorem that \( \sqrt{n}\{\text{vec}(\hat{B}) - \text{vec}(B)\} \) converges in distribution to a normal random vector with mean zero and covariance matrix \( \Lambda = (\Lambda_{k,k'}) \), where \( \Lambda_{k,k'} = A^{-1}E(u_{1k}u_{1k'}^T)A^{-1} \), \( k, k' = 1, \ldots, h \). The \( \sqrt{n} \)-consistency of \( \text{Span}(\hat{\gamma}_1, \ldots, \hat{\gamma}_d) \) follows directly given the consistency of \( \hat{B} \).

**Proof of Theorem 2:** Following the similar techniques of Bondell & Li (2009), we first establish the equivalence between the regularization problem (5) and an adaptive-lasso type regularization with linear constrains. Specifically, define \( \tilde{\beta}_k \equiv \text{diag}(\alpha)\hat{\beta}_k \), \( k = 1, \ldots, h \). Then \( \alpha_j = \tilde{\beta}_{jk}/\tilde{\beta}_{jk}, \ j = 1, \ldots, p, \ k = 1, \ldots, h \), and \( \tilde{\beta}_{jk} \) is the \( j \)-th element of \( \hat{\beta}_k \). Without loss of generality, we assume that the first column of \( B_A \) is fully non-zero. It is easy to show that the minimization of (5) over \( \alpha \) is equivalent to
minimizing
\[
\frac{1}{nh} \sum_{k=1}^{h} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}(Y_i)} \left\{ b_k(\tilde{Y}_i) - \tilde{\beta}_i^* X_i \right\}^2 + \lambda \sum_{j=1}^{p} \frac{|\tilde{\beta}_{j1}|}{|\beta_{j1}|},
\]
(A.1)
over \(\tilde{\beta}_1 = (\tilde{\beta}_{11}, \ldots, \tilde{\beta}_{p1})^\top\) subject to \(\tilde{\beta}_{jk}/\hat{\beta}_{jk} = \tilde{\beta}_{j1}/\hat{\beta}_{j1}\), \(j = 1, \ldots, p\) and \(k = 2, \ldots, h\).

These linear constraints can be absorbed into (A.1), which leads to the following standard adaptive-lasso problem with the weighted least-squares loss
\[
\min_{(\hat{\beta}_{11}, \ldots, \hat{\beta}_{p1})} \frac{1}{nh} \sum_{k=1}^{h} \sum_{i=1}^{n} \frac{\delta_i}{G_{X_i}(Y_i)} \left\{ b_k(\tilde{Y}_i) - \hat{\beta}_i^* X_i^* \right\}^2 + \lambda \sum_{j=1}^{p} \frac{|\hat{\beta}_{j1}|}{|\beta_{j1}|},
\]
where \(X^*_i = (\tilde{\beta}_{i1} X_{i1}, \ldots, \tilde{\beta}_{ip} X_{ip})^\top\), \(k = 1, \ldots, h\) and \(i = 1, \ldots, n\). Based on the condition (C.3), \(\hat{G}_X(\cdot)\) is uniformly consistent and we have \(E \left\{ \frac{\delta_i}{G_{X_i}(Y_i)} | \tilde{Y}_i, X_i \right\} = 1\).

Hence the weighted least-squares loss in (A.2) has the same asymptotic limit as that of the ordinary least-squared loss. Moreover, from Theorem 1, we have that \(\hat{\beta}_k\), \(k = 1, \ldots, h\), are \(\sqrt{n}\)-consistent. Then following similar techniques of Zou (2006) and Zhang & Lu (2007) for establishing the oracle properties of the adaptive-lasso estimators in standard linear regression and Cox’s proportional hazards regression, we can show that, under the regularity conditions that \(n \lambda \to \infty\) and \(\sqrt{n} \lambda \to 0\),

(i) \(P(\tilde{\beta}_{A^c,1} = 0) \to 1\) (i.e. sparsity) and (ii) \(\sqrt{n}(\tilde{\beta}_{A,1} - \beta_{A,1}) \to N(0, \Omega_{11})\) (i.e. asymptotic normality), where \(\tilde{\beta}_1 = (\tilde{\beta}_{A,1}^\top, \tilde{\beta}_{A^c,1}^\top)^\top\) and \(\beta_1 = (\beta_{A,1}^\top, \beta_{A^c,1}^\top)^\top\). The detailed derivations are omitted here. Finally, note that \(\hat{\alpha}_j = \tilde{\beta}_{j1}/\hat{\beta}_{j1}\), \(j = 1, \ldots, p\). The sparsity of \(\tilde{\beta}_1\) is translated as \(\hat{\alpha}_j \neq 0 \Leftrightarrow j \in A\) with probability 1 as \(n \to \infty\). Therefore, \(P(A = A) \to 1\) as \(n \to \infty\). Furthermore, due to the linear constraints, we have \(\tilde{\beta}_{jk} = \frac{\hat{\beta}_{jk}}{\hat{\beta}_{j1}} \tilde{\beta}_{j1}\), \(j \in A\) and \(k = 2, \ldots, h\). Note that \(\mathbb{B}_A = (\beta_{A,1}, \ldots, \beta_{A,h})\) and \(\mathbb{B}_A = (\tilde{\beta}_{A,1}, \ldots, \tilde{\beta}_{A,h})\). Based on the asymptotic normality of \(\text{vec}(\mathbb{B})\) and \(\tilde{\beta}_{A,1}\), we can show that \(\sqrt{n} \left\{ \text{vec}(\mathbb{B}_A) - \text{vec}(\mathbb{B}_A) \right\} \to N(0, \Omega)\), for some \(\Omega > 0\). \(\square\)