9 Appendix

9.1 Discussion of Main Theorems and Conditions in Section 3

By lagrangian duality, the problem (3.1) can be reformulated as a non-constraint problem, which is known as a regular penalized optimization problem. But, deciding \( r_k \) in (3.1) is more straightforward by using an initial estimate, compared to deciding the weights of the quantile loss function for each \( \tau_k \)th quantile in the dual problem.

In (3.1), \( r_k \) controls the possible degrees of goodness of fit for \( \hat{\beta}^{(k)} \) with respect to quantile loss function at \( \tau_k \)th quantile. Controlling goodness of fit was also studied in Belloni & Chernozhukov (2013). They proposed the certain degrees of goodness of fit for their post fit estimator, in i.i.d linear regression model setting.

The RE condition have been used in high dimensional regression analysis (Bickel et al. (2009) and Belloni & Chernozhukov (2011)), and RNI condition is first introduced by Belloni & Chernozhukov (2011). Condition 2 imposes restriction on the weights and it requires \( \lambda < 0.5 \). For the constant weights case, where \( w_j^{(k)} = v_j^{(k)} = 1 \) \((k = 1, \ldots, K; j = 1, \ldots, p)\), Condition 2 reduces to \( \lambda < 0.5 \). But Condition 2 is not guaranteed for the regular adaptive lasso weight case, where \( w_j^{(k)} = 1/|\tilde{\beta}_j^{(k)}| \) and \( v_j^{(k)} = 1/|\tilde{\beta}_j^{(k)} - \tilde{\beta}_j^{(k-1)}| \) with an initial estimate \( \tilde{\beta}^{(k)} \) at quantile level \( \tau_k \). This motivates to use different type of weights, and in Section 4 the derivative of the SCAD penalty function is used for calculating the weights \( w_j^{(k)} \) and \( v_j^{(k)} \), which satisfies Condition 2 with high probability, as can be seen in the proof of Theorem 3.

The main idea of the proof of Theorem 2 is directly comparing the objective functions at \( \beta^o \) and \( \tilde{\beta} \). Since both \( \beta^o \) and \( \tilde{\beta} \) are feasible and \( \tilde{\beta} \) is the optimal solution of (3.1), \( \tilde{\beta} \) does not have greater objective function value than that of \( \beta^o \). When we apply this procedure to the regular lasso estimator whose consistency rate is \( O\{(s_0 \log p/n)^{0.5}\} \) uniformly for all \( k \), it provides the rate \( O\{s_0K(\log p/n)^{0.5}\} \) for the bound of (3.4), which is greater than our rate (3.5) provided that \( \eta = O(s_0 \log p/n) \) and \( W_1/W = o(1) \). Furthermore, applying the same procedure to the adaptive lasso estimator whose consistency rate is \( O\{(s \log n/n)^{0.5}\} \) uniformly for all \( k \) (Fan et al. (2014)), the bound of (3.4) is

\[
O \left\{ \frac{W_1}{W} s_0 K \left( \frac{\log n}{n} \right)^{0.5} \vee \frac{K(s_0 \log p \log n)^{0.5}}{n \lambda_a} \right\},
\]

(9.1)
where $\lambda_a$ is the regularization parameter used for adaptive lasso. The rate (9.1) is also greater than our rate (3.5) provided that $\eta = O(s_0 \log p/n)$ and $W_1/W = o(K(\log n)^{0.5}/\{\lambda_a(ns_0)^{0.5}\})$. As can be seen in Section 4, $\eta = O(s_0 \log p/n)$ is satisfied by using a consistent initial estimate, and sufficiently small $W_1/W$ could be also attained by a strong condition imposed on the magnitude of the true coefficients.

9.2 Discussion of Conditions in Section 4

In Subsection 4.1, we propose two-step procedure by using initial estimates $\tilde{\beta}^{(k)}$ ($k = 1, \ldots, K$). In Subsection 4.3, we exploit the theoretical results of Belloni & Chernozhukov (2011) as stated in Lemma 1. If one is uncomfortable with using the sparsity property in Lemma 1 for an initial estimate, then we can instead use the threshold version of $\tilde{\beta}^{(k)}$ as an initial estimate, which satisfies the sparsity property. Beta-min condition in Theorem 3 imposes the lower bound for nonzero coefficients by the rate $(\log p/n)^{0.5}$, which is not a significant level. This condition is necessary for an initial estimate $\tilde{\beta}^{(k)}$ to include at least $s_k/2$ nonzero components. Indeed, we only need that $\|\tilde{\beta}^{(k)}\|_0/s_k$ is greater than some constant. The lower bound of $\lambda_n$ stated in Theorem 3 is required to guarantee that the weights of the locations where the true quantile coefficients are zero and true quantile coefficients are same at adjacent quantile levels are well separated from zero.

While (4.3) in Theorem 4 has been considered in high dimensional analysis to establish the exact model selection property, the lower bound of nonzero interquantile differences (4.4) has not been studied. The bound in (4.4) can be demonstrated by using several models. For simplicity, we consider equal width quantile levels $\tau_k$ ($k = 1, \ldots, K$) with $\tau_k - \tau_{k-1} \approx 1/K$. First, consider a location scale model: $y_k = x_i^T \beta + x_i^T r_\epsilon$, where the design $x_i$ and the vector $r \in \mathbb{R}^p$ have nonnegative components. Then the quantile coefficients at $\tau_k$th quantile is $\beta(\tau_k) = \beta + rq_{\tau_k}$, where $q_{\tau_k}$ is the $\tau_k$th quantile of $\epsilon$. For any adjacent quantiles $\tau_k$ and $\tau_{k-1}$, the difference of their true quantile coefficients is $\beta(\tau_k) - \beta(\tau_{k-1}) = r(q_{\tau_k} - q_{\tau_{k-1}})$, thus, the difference of each components is

$$\beta_j(\tau_k) - \beta_j(\tau_{k-1}) = (q_{\tau_k} - q_{\tau_{k-1}})r_j = (\tau_k - \tau_{k-1})r_j/f\{F^{-1}(\tau)\}$$

for some $\tau \in (\tau_{k-1}, \tau_k)$. Since $f(x)$ is upper bounded by $\hat{U}$, the difference of each components are at least the rate $O\{(\tau_k - \tau_{k-1})r_j\} = O(r_j/K)$, which implies that the lower bound of nonzero interquantile differences in (4.4) holds as long as $r_j \asymp K(s_0 \log p/n)^{0.5}$ ($j = 1, \ldots, p$). Consider a second model, a random
coefficient model, where \( y_i = \sum_{j=1}^{p} x_{ij} \beta_j(u_i) \) with \( x_{ij} \geq 0, u_i \) are independent and identically distributed from uniform(0, 1) and independent of design \( x_{ij} \)s, and \( b_j(u) \) \((j = 1, \ldots, p)\) are strictly increasing and differentiable functions. Then the true quantile coefficients at the \( \tau \)th quantile is \( \beta(\tau) = \{b_1(\tau), \ldots, b_p(\tau)\}^T \), because \( b_j \)s are strictly increasing functions and \( \tau \) is the \( \tau \)th quantile of uniform(0, 1). Thus, the difference of quantile coefficients at adjacent quantile levels \( \tau_k \) and \( \tau_{k-1} \) is

\[
|\beta_j(\tau_k) - \beta_j(\tau_{k-1})| = |b'_j(\tau_k) - b'_j(\tau_{k-1})| \approx |b'_j(\tau)|/K
\]

for some \( \tau \in (\tau_{k-1}, \tau_k) \). So (4.4) holds as long as \( |b'_j(\tau)| \approx K(\log p/n)^{0.5} \) \((j = 1, \ldots, p)\).

### 9.3 Computation

Our Dantzig-type joint quantile regression is equivalent to linear programming problem with the aid of slack variables, and can be solved by existing optimization packages. We use the linear programming package “lpSolve” in the program R. For the chosen parameters \( \beta^{(k)} \) \((k = 1, \ldots, K)\), \( w_j^{(k)} \)s, \( v_j^{(k)} \)s, \( \lambda \) and \( \Lambda_k \) \((k = 1, \ldots, K)\), the problem (3.1) can be rewritten as following problem.

\[
\begin{align*}
\min_{\beta^{(k)} \in \mathbb{R}^p, k = 1, \ldots, K} & \sum_k \sum_j w_j^{(k)} \{ \beta_j(k) + \beta_j(k)^- \} + \lambda \sum_{k=2}^K \sum_j v_j^{(k)} \{ \beta_j(k) + \beta_j(k)^- \} \\
\text{subject to} & \quad \beta_j(k) = \beta_j(k)^+ - \beta_j(k)^- \quad (k = 1, \ldots, K; j = 1, \ldots, p) \\
& \quad d_j^{(k)} - d_j^{(k)^-} = \beta_j(k)^+ - \beta_j(k)^- - \beta_j(k-1)^+ + \beta_j(k-1)^- \quad (k = 2, \ldots, K; j = 1, \ldots, p) \\
& \quad a_{ki}^+, a_{ki}^- \leq \beta_j(k)^+ - \beta_j(k)^- \quad (i = 1, \ldots, n; j = 1, \ldots, p; k = 1, \ldots, K) \\
& \quad y_i - x_i^T \beta^{(k)} = a_{ki}^+- a_{ki}^- \quad (i = 1, \ldots, n; k = 1, \ldots, K) \\
& \quad \frac{1}{n} \sum_k a_{ki}^+ + \frac{1}{n} (1 - \tau_k) \sum_k a_{ki}^- \leq \frac{1}{n} \sum_k \rho_{nk} \{ y_i - x_i^T \beta^{(k)} \} + \Lambda_k \frac{\bar{s}_k \log p}{n} \quad (k = 1, \ldots, K),
\end{align*}
\]

where \( \beta^{(k)} \) is the initial estimate at \( \tau_k \)th quantile.

### 9.4 Lemmas and Proofs

Throughout the proofs, let \( F_i \) denote the conditional distribution of \( y_i \) given \( x_i \) for all \( i = 1, \ldots, n \). Define the diagonal matrices

\[
H_k = \text{diag}\{f_1\{x_1^T \beta(\tau_k)\}, \ldots, f_n\{x_n^T \beta(\tau_k)\}\} \quad (k = 1, \ldots, K).
\]
Then for any vector $\delta \in \mathbb{R}^p$, we define an intrinsic norm such that
\[
\|\delta\|_{k,2} = \left(\frac{\delta^T X^T H_k X \delta}{n}\right)^{0.5} \quad (k = 1, \ldots, K).
\]
For any positive constant $c$ and the sets $T^{(k)} (k = 1, \ldots, K)$ as defined in (2.2), let
\[
A^{(k)}(c) = \{ \delta : \delta \neq 0, \delta \in \mathbb{R}^p, \|\delta_{(T^{(k)})^c}\|_1 \leq c\|\delta_{T^{(k)}}\|_1 \}.
\]
Define two functions such that
\[
Q_k(\beta) = E\left\{\frac{1}{n} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - x_i^T \beta)\right\}, \quad \hat{Q}_k(\beta) = \frac{1}{n} \sum_{i} \rho_{\tau_k}(y_i - x_i^T \beta) \quad (k = 1, \ldots, K).
\]
For simplicity, for any $\beta = [\beta^{(1)}, \ldots, \beta^{(K)}]^T \in \mathbb{R}^{K \times p}$, let
\[
F(\beta) = \sum_{k} \sum_{j} w^{(k)}_j |\beta^{(k)}_j| + \lambda \sum_{k \geq 2} \sum_{j} v^{(k)}_j |\beta^{(k)}_j - \beta^{(k-1)}_j|,
\]
where $w^{(k)} (k = 1, \ldots, K)$ and $v^{(k)} (k = 2, \ldots, K)$ are $p$-dimensional weight vectors. We begin by providing several lemmas with proofs that will be used for the proof of our main theorems.

**Lemma 2.** Suppose Conditions 1 and RNI($X$, $2s_0$, $c_0$) hold. Then for any $\delta \in A^{(k)}(c_0) (k = 1, \ldots, K),$
\[
Q_k(\{\beta_{\tau_k} + \delta\}) - Q_k(\{\beta_{\tau_k}\}) \geq \frac{3L_0^{3/2} q(2s_0, c_0)}{8U'} \|\delta\|_{k,2} \wedge \frac{1}{4} \|\delta\|_{k,2}^2 \quad (k = 1, \ldots, K).
\]

Lemma 2 is a fixed design version of the lemma 4 in Belloni & Chernozhukov (2011). Lemma 2 provides the lower bound of the difference of the expected values of quantile loss function under defined cone $A^{(k)}(c_0)$.

**Proof.** The proof is similar to the lemma 4 in Belloni & Chernozhukov (2011). But we rewrite it by using
our settings. By using Knight’s identity, we have

\[
Q_k\{\beta(\tau_k) + \delta\} - Q_k\{\beta(\tau_k)\} = E \left[ \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - x_i^T \beta(\tau_k) - x_i^T \delta) - \rho_{\tau_k}(y_i - x_i^T \beta(\tau_k)) \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} E \left( -x_i^T \delta[\tau_k - I\{y_i - x_i^T \beta(\tau_k) \leq 0\}] \right)
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} E \left( \int_{0}^{x_i^T \delta} [I\{y_i - x_i^T \beta(\tau_k) \leq z\} - I\{y_i - x_i^T \beta(\tau_k) \leq 0\}] dz \right)
\]

\[
= \frac{1}{n} \sum_{i} \int_{0}^{x_i^T \delta} F_i\{x_i^T \beta(\tau_k) + z\} - F_i\{x_i^T \beta(\tau_k)\} dz
\]

\[
= \frac{1}{n} \sum_{i} \int_{0}^{x_i^T \delta} z f_i\{x_i^T \beta(\tau_k)\} + \frac{z^2}{2} f_i'\{x_i^T \beta(\tau_k)\} + \bar{z}_{k,i} \} dz
\]

\[
\geq \frac{1}{2} \|\delta\|_{k,2}^2 - \frac{\bar{U}'}{6} \frac{1}{n} \sum_{i} |x_i^T \delta|^3,
\]

where \(\bar{z}_{k,i} \in [0, \bar{z}]\) depends on \(k\) and \(i\) in the fourth equality, and we use Condition 1 for the last inequality.

Now if \(\delta \in A^{(k)}(c_0)\) and \(\|\delta\|_{k,2} \leq 1.5q(2s_0, c_0)L_{0}^{3/2}/\bar{U}'\), then

\[
\|\delta\|_{k,2} \leq \frac{1}{2} \frac{\bar{U}'}{6} \frac{1}{n} \sum_{i} |x_i^T \delta|^3 \leq \frac{3}{2} \frac{1}{n} \sum_{i} |x_i^T \delta|^3
\]

\[
\leq \frac{3}{2} \frac{1}{n} \sum_{i} \|\delta\|_{k,2}^{3/2} \leq \frac{3}{2} \frac{1}{n} \sum_{i} |x_i^T \delta|^3.
\]

(9.4)

where the second inequality uses the definition of \(L_0\) as stated in Condition 1. So we have \(\|\delta\|_{k,2}^{2}/4 \geq (\bar{U}'/6)\sum_{i} |x_i^T \delta|^3/n\), which yields \(Q_k\{\beta(\tau_k) + \delta\} - Q_k\{\beta(\tau_k)\} \geq \|\delta\|_{k,2}^{2}/4\) in (9.3).

Suppose \(\delta \in A^{(k)}(c_0)\) and \(\|\delta\|_{k,2} > 1.5q(2s_0, c_0)L_{0}^{3/2}/\bar{U}'\). Let \(\alpha(\delta) = 1.5q(2s_0, c_0)L_{0}^{3/2}/(\bar{U}'\|\delta\|_{k,2})\).

Since \(0 \leq \alpha(\delta) \leq 1\) and \(Q_k(\beta(\tau_k) + x) - Q_k(\beta(\tau_k))\) is a convex function of \(x\) and has zero when \(x = 0\), we have

\[
\alpha(\delta)(Q_k\{\beta(\tau_k) + \delta\} - Q_k\{\beta(\tau_k)\}) \geq Q_k\{\beta(\tau_k) + \alpha(\delta)\delta\} - Q_k\{\beta(\tau_k)\}.
\]
Lemma 3 controls the empirical error over the fixed design matrix case. For each \( k \),

\[
Q_k \{ \beta(\tau_k) + \delta \} - Q_k \{ \beta(\tau_k) \} \geq \frac{1}{\alpha(\delta)} \| Q_k \{ \beta(\tau_k) + \alpha(\delta)\delta \} - Q_k \{ \beta(\tau_k) \} \|_F^2
\]

\[
\geq \frac{1}{\alpha(\delta)} \frac{\| \delta \|_{k,2}^2}{4}
\]

\[
= \frac{3L_0^{3/2} q(2s_0, c_0) \| \delta \|_{k,2}}{8U'}
\]

where the second inequality holds due to \( \alpha(\delta) \delta \in A^{(k)}(c_0) \) and \( \| \alpha(\delta) \delta \|_{k,2} \leq 1.5q(2s_0, c_0)L_0^{3/2}/U' \) so that we can apply (9.4). This completes the proof. \( \Box \)

**Lemma 3.** Suppose Conditions 1, RE(X, 2s_0, c_0) and RNI(X, 2s_0, c_0) hold. Then for any \( (t_1, \ldots, t_K) \in \mathbb{R}_+^K \), there exists an universal constant \( C_7 > 0 \) with probability at least \( 1 - 1/n \), satisfying following inequality:

\[
\sup_{\delta \in A^{(k)}(c_0), \| \delta \|_{k,2} \leq t_k} \left| Q_k \{ \beta(\tau_k) + \delta \} - Q_k \{ \beta(\tau_k) \} - \tilde{Q}_k \{ \beta(\tau_k) + \delta \} + \tilde{Q}_k \{ \beta(\tau_k) \} \right| \leq C_7 \frac{1 + c_0}{k(s_0, c_0)} t_k \left( \frac{s_0 \log p}{n} \right)^{0.5} \quad (k = 1, \ldots, K).
\]

Lemma 3 controls the empirical error over \( A^{(k)}(c_0) \) and also analogous to the lemma 5 in the Belloni & Chernozhukov (2011).

**Proof.** The proof is analogous to the proof of Lemma 5 in the Belloni & Chernozhukov (2011), but this is the fixed design matrix case. For each \( k = 1, \ldots, K \), define a set of vectors \( \mathcal{F}^{(k)} \) by

\[
\mathcal{F}^{(k)} = \{ \delta \in \mathbb{R}^p : \delta \in A^{(k)}(c_0), \| \delta \|_{k,2} \leq t_k \}.
\]

For each \( k = 1, \ldots, K \), define independent stochastic processes \( \{ Z_i^{(k)}(\delta) : \delta \in \mathcal{F}^{(k)} \}_{i=1}^n \) such that

\[
n^{0.5} Z_i^{(k)}(\delta) = \rho_{rk} \{ y_i - x_i^T \beta(\tau_k) - x_i^T \delta \} - \rho_{rk} \{ y_i - x_i^T \beta(\tau_k) \} - E[\rho_{rk} \{ y_i - x_i^T \beta(\tau_k) - x_i^T \delta \}] + E[\rho_{rk} \{ y_i - x_i^T \beta(\tau_k) \}] \quad (i = 1, \ldots, n).
\]

Then

\[
\sup_{\delta \in A^{(k)}(c_0), \| \delta \|_{k,2} \leq t_k} \left| Q_k \{ \beta(\tau_k) + \delta \} - Q_k \{ \beta(\tau_k) \} - \tilde{Q}_k \{ \beta(\tau_k) + \delta \} + \tilde{Q}_k \{ \beta(\tau_k) \} \right| = \frac{1}{n^{0.5}} \| \sum_i Z_i^{(k)} \|_{\mathcal{F}^{(k)}}
\]

where \( \| \sum_i Z_i^{(k)} \|_{\mathcal{F}^{(k)}} = \sup_{\delta \in \mathcal{F}^{(k)}} | \sum_i Z_i^{(k)}(\delta) | \). Then we have

\[
\sup_{\delta \in A^{(k)}(c_0)} \text{var} \{ \sum_i Z_i^{(k)}(\delta) \} \leq \delta_k^2 \left( \frac{1}{n} \sum_i x_i x_i^T \right) \delta \leq \frac{t_k^2}{L_0},
\]

24
where the last inequality holds due to Condition 1. Since $Z_1^{(k)}, \ldots, Z_n^{(k)}$ are independent mean zero processes, applying the symmetrization lemma (Lemma 2.3.7 in Van & Wellner (1996)) yields

$$
P\left\{ \left\| \sum_i Z_i^{(k)} \right\|_{F^{(k)}} \geq z \right\} \leq 2P\left\{ \left\| \sum_i W_i Z_i^{(k)} \right\|_{F^{(k)}} \geq \frac{z}{4} \right\}
$$

where $W_1, \ldots, W_n$ be a rademacher sequence, independent of $x_i$s and the errors $\epsilon_i$s, and

$$
n^{0.5} Z_i^{(k)}(\delta) = \rho_{\tau_k} \left\{ y_i - x_i^T \beta(\tau_k) - x_i^T \delta \right\} - \rho_{\tau_k} \left\{ y_i - x_i^T \beta(\tau_k) \right\}.
$$

Now define the function $\rho_k : \mathbb{R}^2 \to \mathbb{R}$ such that $\rho_k(s, y) = \rho_{\tau_k}(y - s) - \rho_{\tau_k}(y)$, and then $Z_i^{(k)}(\delta)$ can be rewritten as $Z_i^{(k)}(\delta) = n^{-0.5} \rho_k \left\{ x_i^T \delta, y_i - x_i^T \beta(\tau_k) \right\}$. Then the following Lipschitz condition holds for $\rho_k \left\{ \cdot, y_i - x_i^T \beta(\tau_k) \right\}$:

$$
\left| z_k \{ s_1, y_i - x_i^T \beta(\tau_k) \} - z_k \{ s_2, y_i - x_i^T \beta(\tau_k) \} \right| \leq \left\{ \tau_k \vee (1 - \tau_k) \right\} |s_1 - s_2| \leq |s_1 - s_2|.
$$

Applying exponential moment inequality for contractions in Ledoux & Talagrand (1991), for fixed $\lambda > 0$,

$$
E \left\{ e^{\lambda \left\| \sum_i W_i Z_i^{(k)} \right\|_{F^{(k)}}} \right\} \leq E \left( e^{2\lambda \sup_{\| \delta \|_1 \leq tk \Gamma} \left\{ \sum_i \frac{1}{\sqrt{n}} W_i x_i^T \delta \right\}} \right)
\leq E \left( e^{2\lambda \sup_{\| \delta \|_1 \leq tk \Gamma} \left\{ \sum_i \frac{1}{\sqrt{n}} \sum_j \delta_j \sum_i W_i x_{ij} \right\}} \right)
\leq E \left( e^{2\lambda \max_j \left\{ \sup_{\| \delta \|_1 \leq tk \Gamma} \left\| \delta_1 \right\| \sum_i W_i x_{ij} \right\}} \right)
\leq E \left( e^{2\max_j \left\{ \sup_{\| \delta \|_1 \leq tk \Gamma} \left\| \delta \right\| \sum_i W_i x_{ij} \right\}} \right)
\leq p \max_j E \left( e^{2\lambda k \Gamma \max_j \left\{ \sup_{\| \delta \|_1 \leq tk \Gamma} \left\| \delta \right\| \sum_i W_i x_{ij} \right\}} \right),
$$

(9.5)

where $\Gamma = s_0^{0.5}(1 + c_0)/\{L_0^{0.5} k(s_0, c_0)\}$. To show the first inequality, we use that for $\delta \in A^{(k)}(c_0)$,

$$
\| \delta \|_1 \leq (1 + c_0) \| \delta_{T^{(k)}} \|_1 \leq (1 + c_0) s_0^{0.5} \| \delta_{T^{(k)}} \|_2 \leq (1 + c_0) s_0^{0.5} \| \delta \|_{k,2} / \{L_0^{0.5} k(s_0, c_0)\},
$$

which implies that

$$
\{ \delta \in A^{(k)}(c_0) : \| \delta \|_{k,2} \leq t_k \} \subset \{ \delta \in \mathbb{R}^p : \| \delta \|_1 \leq t_k \Gamma \}.
$$
And we also have
\[
\begin{align*}
\max_j E \left( e^{2\lambda_k \Gamma_k \frac{1}{n} \sum_i W_{ix_{ij}}} \right) & \leq \max_j \left\{ E \left( e^{2\lambda_k \Gamma_k \frac{1}{n} \sum_i W_{ix_{ij}}} \right) + E \left( e^{-2\lambda_k \Gamma_k \frac{1}{n} \sum_i W_{ix_{ij}}} \right) \right\} \\
& = 2 \max_j E \left( e^{2\lambda_k \Gamma_k \frac{1}{n} \sum_i W_{ix_{ij}}} \right) \\
& = 2 \max_j \prod_i E \left( e^{2\lambda_k \Gamma_k \frac{1}{n} x_{ij}} \right) \\
& = 2 \max_j \prod_i \left( \frac{1}{2} e^{2\lambda_k \Gamma_k \frac{1}{n} x_{ij}} + \frac{1}{2} e^{-2\lambda_k \Gamma_k \frac{1}{n} x_{ij}} \right) \\
& \leq 2 \max_j \prod_i \left( e^{-\frac{4\lambda_k^2 \|Z_{ij}\|^2}{n}} \right) \\
& \leq 2 e^{2\lambda_k^2 \|Z_{ij}\|^2}, \\
\end{align*}
\]  
(9.6)
where the second inequality uses the fact that $e^{x^2/2} \geq (e^x + e^{-x})/2$ for any $x$. By (9.5) and (9.6)
\[
E \left\{ e^{\lambda \| \sum_i W_i Z_i^{(k)} \|_{\|F\|}} \right\} \leq 2pe^{2\lambda_k^2 \|Z_{ij}\|^2}.
\]
Now by applying Markov inequality with $\lambda = z/(16\lambda_k^2 \|Z_{ij}\|^2)$,
\[
pr \left\{ \left\| \sum_i W_i Z_i^{(k)} \right\|_{\|F\|} > z/4 \right\} \leq e^{-\lambda z/4} E \left\{ e^{\lambda \| \sum_i W_i Z_i^{(k)} \|_{\|F\|}} \right\} \\
\quad \leq 2pe^{2\lambda_k^2 \|Z_{ij}\|^2 - \lambda z/4} = 2pe^{-\frac{z^2}{128\lambda_k^4 \|Z_{ij}\|^2}}.
\]
So
\[
pr \left\{ \left\| \sum_i Z_i^{(k)} \right\|_{\|F\|} \geq z \right\} \leq \frac{4pe^{-\frac{z^2}{128\lambda_k^4 \|Z_{ij}\|^2}}}{1 - \frac{t_k^2}{4L_0z^2}}.
\]
Plugging $z = (128)^{0.5}t_k \Gamma \{ \log(8p) + \log K + \log n \}^{0.5}$ yields $1 - t_k^2/(4L_0z^2) \geq 1/2$, because it is equivalent to $256s_0(1 + c_0)^2 \log(8pnK) \geq k^2(s_0, c_0)$. So we have
\[
pr \left[ \left\| \sum_i Z_i^{(k)} \right\|_{\|F\|} \geq (128)^{0.5}t_k \Gamma \{ \log(8p) + \log K + \log n \}^{0.5} \right] \leq \frac{1}{Kn}.
\] (9.7)
Since $p > n \lor K$, (9.7) implies that there exists an universal constant $C_7$, which only depends on $L_0$ and satisfies $\left\| \sum_i Z_i^{(k)} \right\|_{\|F\|} \leq C_7(1 + c_0)t_k s_0^{0.5}(\log p)^{0.5}/k(s_0, c_0)$ with probability at least $1 - 1/(Kn)$ for all $k = 1, \ldots, K$. So $\left\| \sum_i Z_i^{(k)} \right\|_{\|F\|} \leq C_7(1 + c_0)t_k s_0^{0.5}(\log p)^{0.5}/k(s_0, c_0)$ ($k = 1, \ldots, K$) holds with probability at least $1 - 1/n$. This completes the proof.

\[\square\]

\textbf{Lemma 4.} With probability at least $1 - 1/n$,
\[
\max_k \left\| \sum_{i=1}^n x_i [\tau_k] - I \{ y_i \leq x_i^T \beta(\tau_k) \} / n \right\|_\infty \leq 3(\log p/n)^{0.5}.
\]
Proof. Lemma 1.5 in Ledoux(1991) implies that for fixed $j, k$,

$$
pr \left( \left| \sum_i x_{ij} \tau_k - I \{ y_i \leq x_i^T \beta(\tau_k) \} \right| / n \geq t \right) \leq 2 \exp\left(-nt^2/2\right)
$$

Therefore,

$$
pr \left( \max_k \max_j \left| \sum_i x_{ij} \tau_k - I \{ y_i \leq x_i^T \beta(\tau_k) \} \right| / n \geq t \right) \leq 2kp \exp\left(-nt^2/2\right)
$$

Putting $t = 3(\log p/n)^{0.5}$ and using $p > n \vee K$ yields

$$
pr \left( \max_k \max_j \left| \sum_i x_{ij} \tau_k - I \{ y_i \leq x_i^T \beta(\tau_k) \} \right| / n \geq 3(\log p/n)^{0.5} \right) \leq \frac{1}{n}.
$$

This completes the proof. $\square$

**Lemma 5.** Consider the estimators $\tilde{\beta}^{(k)}$ ($k = 1, \ldots, K$), which are defined in (4.1) with the corresponding regularization parameters $\tilde{\lambda}_k$ ($k = 1, \ldots, K$). If $C_2 \sqrt{\log p/n} \geq \tilde{\lambda}_k \geq C_3 \sqrt{\log p/n}$ holds, where $C_2$ and $C_3$ are the constants used in Lemma 1, then with probability at least $1 - q_n$, there exists an universal constant $C_8$ satisfying

$$
\hat{Q}_k\{\tilde{\beta}^{(k)}\} - \hat{Q}_k\{\beta(\tau_k)\} \leq C_2 C_4 (C_5 + 1)^{0.5} \frac{s_0 \log p}{n} (k = 1, \ldots, K),
$$

where $C_5$ is the constant as stated in Lemma 1.

Lemma 5 provides the upper bound of $\hat{Q}_k\{\tilde{\beta}^{(k)}\}$.

Proof. Since the initial estimate $\tilde{\beta}^{(k)}$ is a minimizer of $\hat{Q}_k(\beta) + \tilde{\lambda}_k \| \beta \|_1$, where $\tilde{\lambda}_k \leq C_2(\log p/n)^{0.5}$, the following holds with probability at least $1 - q_n$:

$$
\hat{Q}_k\{\tilde{\beta}^{(k)}\} - \hat{Q}_k\{\beta(\tau_k)\} \leq \tilde{\lambda}_k \{ \| \beta(\tau_k) \|_1 - \| \tilde{\beta}^{(k)} \|_1 \}
\leq C_2 \left( \frac{\log p}{n} \right)^{0.5} \| \beta(\tau_k) - \tilde{\beta}^{(k)} \|_1
\leq C_2 \left( \frac{\log p}{n} \right)^{0.5} (C_5 + 1)^{0.5} s_0^{0.5} C_4 \left( \frac{s_0 \log p}{n} \right)^{0.5}
= C_2 C_4 (C_5 + 1)^{0.5} s_0^{0.5} C_4 \left( \frac{s_0 \log p}{n} \right)^{0.5}
$$

where the second inequality comes from the upper bound of $\tilde{\lambda}_k$ as stated in Lemma 1. The third inequality holds with probability at least $1 - q_n$ due to the $\ell_2$ estimation error bound and the sparsity bound of $\tilde{\beta}^{(k)}$ as stated in Lemma 1. This completes the proof. $\square$
9.5 Proofs of Theorems

Proof of Theorem 1. We first fix \( \eta \) and \( c_0 \) such that \( 0 \leq \eta < \frac{9L_0^3q^2(2s_0,c_0)}{32U^2} \) and \( c_0 \geq (W_1 + 2\lambda W)/(W - 2\lambda W) \). Now we will show that \( \hat{\beta}^{(k)} - \beta(\tau_k) \) are included in the cone \( A^{(k)}(c_0) \) for all \( k = 1, \ldots, K \). By using the fact that \( \hat{\beta} \) and \( \beta' = [\hat{\beta}(1), \ldots, \hat{\beta}(k-1), \beta(\tau_k), \hat{\beta}(k+1), \ldots, \hat{\beta}(K)]^T \) are feasible and \( \hat{\beta} \) is a global minimizer of (3.1), we have

\[
\sum_j w_j^{(k)} |\hat{\beta}_j^{(k)}| \leq \sum_j w_j^{(k)} |\beta_j(\tau_k)| + \lambda \sum_j v_j^{(k+1)} |\beta_j^{(k)} - \hat{\beta}_j^{(k+1)}|.
\]

By applying triangle inequality, it reduces to

\[
\sum_j w_j^{(k)} |\beta_j(\tau_k)| + \lambda \sum_j v_j^{(k)} |\beta_j^{(k)} - \beta_j^{(k-1)}| + \lambda \sum_j v_j^{(k+1)} |\hat{\beta}_j^{(k+1)} - \beta_j^{(k+1)}|.
\]

Rearranging yields

\[
\sum_j [w_j^{(k)} - \lambda(v_j^{(k)} + v_j^{(k+1)})]|\hat{\beta}_j^{(k)} - \beta_j(\tau_k)| \leq \sum_j [w_j^{(k)} + \lambda(v_j^{(k)} + v_j^{(k+1)})]|\hat{\beta}_j^{(k)} - \beta_j(\tau_k)|.
\]

Using \( W - 2\lambda W \geq 0 \) as stated in Condition 2, we have

\[
\sum_j |\hat{\beta}_j^{(k)} - \beta_j(\tau_k)| \leq \frac{W_1 + 2\lambda W}{W - 2\lambda W} \sum_j |\hat{\beta}_j^{(k)} - \beta_j(\tau_k)| \quad (k = 1, \ldots, K).
\]

This implies that \( \hat{\beta}^{(k)} - \beta(\tau_k) \in A^{(k)}(c_0) \) for all \( k = 1, \ldots, K \), with probability at least \( 1 - 1/n - P(E''_n) \),

\[
\frac{\|\hat{\beta}^{(k)} - \beta(\tau_k)\|_{k,2}^2}{4} \geq \left\{ \frac{3\epsilon_0^3/2q(2s_0,c_0)}{8U^2} \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_{k,2} \right\} \leq Q_k\{\hat{\beta}^{(k)}\} - Q_k\{\beta(\tau_k)\}
\]

where the first inequality comes from Lemma 2, the second inequality holds for any \( \eta > 0 \) with probability at least \( 1 - P(E''_n) \) and the last inequality uses lemma 3 with \( t_k = \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_{k,2} \) for all \( k = 1, \ldots, K \).
To satisfy (9.8), in the left hand side of (9.8), the first term should be less than the second term. Because if the second term is less than the first term, then \( \|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2} \geq 3L_0^{3/2}q(2s_0, c_0)/(2\tilde{U}')\), so (9.8) implies

\[
\eta \geq \left\{ \frac{3L_0^{3/2}q(2s_0, c_0)}{8\tilde{U}'} - C_7 \frac{1 + c_0}{k(s_0, c_0)} \left( \frac{s_0 \log p}{n} \right)^{0.5} \right\} \|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2}
\]

\[
\geq \frac{3L_0^{3/2}q(2s_0, c_0) 3L_0^{3/2}q(2s_0, c_0)}{16\tilde{U}'}
\]

where the last inequality comes from \(3L_0^{3/2}q(2s_0, c_0)/(16\tilde{U}') \geq C_7(1 + c_0)(s_0 \log p/n)^{0.5}/k(s_0, c_0)\). This contradicts to \(\eta < 9L_0^{3/2}q(2s_0, c_0)/(32\tilde{U}')\). Thus, (9.8) implies that, with probability at least \(1 - 1/n - P(E_\eta^c)\),

\[
\frac{\|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2}^2}{4} \leq \eta + C_7 \frac{1 + c_0}{k(s_0, c_0)} \left( \frac{s_0 \log p}{n} \right)^{0.5} \|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2},
\]

which yields

(9.9) \[
\|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2} \leq 4C_7 \frac{1 + c_0}{k(s_0, c_0)} \left( \frac{s_0 \log p}{n} \right)^{0.5} + 2\eta^{0.5} (k = 1, \ldots, K).
\]

We can derive the upper bound of \(\|\hat{\beta}(k) - \beta(\tau_k)\|_2\) by \(\|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2}\) as follow: For any \(\delta \in A(\kappa)(c_0)\), we know that \(m\)th largest components of \(\delta\) outside of the set \(T(k)\) is less than \(\|\delta(T(k))^c\|_1/m\). Therefore

\[
\|\hat{\delta}_{\{T(k) \cup T(k)(\delta,s_0)^c\}}\|_2^2 \leq \sum_{m = s_0 + 1} \frac{\|\hat{\delta}_{\{T(k)^c\}}\|_1^2}{m^2} \leq \frac{\|\hat{\delta}_{\{T(k)^c\}}\|_1^2}{s_0} \leq c_0^2 \frac{\|\hat{\delta}_{T(k)}\|_2^2}{s_0}
\]

\[
\leq c_0^2 \|\delta_{\{T(k) \cup T(k)(\delta,s_0)^c\}}\|_2^2,
\]

where \(T(k)(\delta,s_0) \subseteq \{1, \ldots, p\} \cap \{T(k)^c\}\) is the support of the \(s_0\) largest in absolute value components of the vector \(\delta\) outside of the support set \(T(k)\). Hence \(\|\delta\|_2^2 \leq (1 + c_0^2)\|\delta_{\{T(k) \cup T(k)(\delta,s_0)^c\}}\|_2^2\) for any \(\delta \in A(\kappa)(c_0)\).

Since \(\hat{\beta}(k) - \beta(\tau_k) \in A(\kappa)(c_0)\) and \(\text{RE}(X, 2s_0, c_0)\), we have

(9.10) \[
\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq \frac{1 + c_0}{L_0^{0.5}k(2s_0, c_0)}\|\hat{\beta}(k) - \beta(\tau_k)\|_{k,2} (k = 1, \ldots, K).
\]

Combining (9.9) and (9.10) yields that, with probability at least \(1 - 1/n - P(E_\eta^c)\),

\[
\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq 4C_7 \frac{(1 + c_0)^2}{k(2s_0, c_0)k(s_0, c_0)L_0^{0.5}} \left( \frac{s_0 \log p}{n} \right)^{0.5} + 2 \frac{1 + c_0}{k(2s_0, c_0)L_0^{0.5}} \eta^{0.5} (k = 1, \ldots, K).
\]

This implies that with probability at least \(1 - 1/n - P(E_\eta^c)\), there exists an universal constant \(C_1\), which depends on \(L_0\) and \(C_7\), satisfying

\[
\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq C_1 \left( \frac{(1 + c_0)^2}{k(2s_0, c_0)} \frac{1}{k(s_0, c_0)} \right) \left( \frac{s_0 \log p}{n} + \eta \right)^{0.5} (k = 1, \ldots, K).
\]
This holds for any $\eta$ and $c_0$ such that $0 \leq \eta < \frac{9L_0^2g^2(2s_0,c_0)}{32L^2}$ and $c_0 \geq (W_1 + 2\lambda W)/(W - 2\lambda W)$, which completes the proof.

\[
\frac{9L_0^2g^2(2s_0,c_0)}{32L^2} < \eta < \frac{9L_0^2g^2(2s_0,c_0)}{32L^2},
\]

\[
c_0 \geq (W_1 + 2\lambda W)/(W - 2\lambda W).
\]

Proof of Theorem 2. We first fix $\eta$ and $c_0$ such that $0 \leq \eta < \frac{9L_0^2g^2(2s_0,c_0)}{32L^2}$ and $c_0 \geq (W_1 + 2\lambda W)/(W - 2\lambda W)$. The main idea is to compare the objective functions at two vectors $\hat{\beta}$ and $\beta^o$. Since $\beta^o$ is feasible, $F(\hat{\beta})$ should not be greater than $F(\beta^o)$, where the function $F(\cdot)$ is defined in (9.2). So

\[
0 \leq F(\beta^o) - F(\beta) = \left\{ \sum_k \sum_{j \in T^{(k)}} w_j^{(k)} |\beta_j(\tau)| - \beta_j(\tau - 1) + \lambda \sum_k \sum_{j \in B^{(k)}} v_j^{(k)} |\beta_j(\tau) - \beta_j(\tau - 1)| \right\}
\]

\[
- \left\{ \sum_k \sum_{j \in T^{(k)}} w_j^{(k)} |\hat{\beta}_j^{(k)} - \beta_j(\tau)| + \lambda \sum_k \sum_{j \in B^{(k)}} v_j^{(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| + \sum_k \sum_{j \in (T^{(k)})} w_j^{(k)} |\hat{\beta}_j^{(k)}| + \lambda \sum_k \sum_{j \in B^{(k)}} v_j^{(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| \right\}
\]

By triangle inequality, above yields

\[
0 \leq \sum_k \sum_{j \in (T^{(k)})} w_j^{(k)} |\hat{\beta}_j^{(k)}| + \lambda \sum_k \sum_{j \in B^{(k)}} v_j^{(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| \leq \sum_k \sum_{j \in T^{(k)}} w_j^{(k)} |\hat{\beta}_j^{(k)} - \beta_j(\tau)| + \lambda \sum_k \sum_{j \in B^{(k)}} v_j^{(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)} - \beta_j(\tau) + \beta_j(\tau - 1)|
\]

\[
\leq W_1 \sum_k \sum_{j \in T^{(k)}} |\hat{\beta}_j^{(k)} - \beta_j(\tau)| + W_1 \lambda \sum_k \sum_{j \in B^{(k)}} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)} - \beta_j(\tau) + \beta_j(\tau - 1)|
\]

\[
\leq W_1 \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}_{T^{(k)}}\|_1 + W_1 \lambda \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}_{B^{(k)}}\|_1
\]

\[
+ W_1 \lambda \sum_{k \geq 2} \|\{\hat{\beta}^{(k-1)} - \beta(\tau - 1)\}_{B^{(k)}}\|_1
\]

\[
\leq W_1 K_{0.5} \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}_{T^{(k)}}\|_1^{0.5} + W_1 \lambda K_{0.5} \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}_{B^{(k)}}\|_1^{0.5}
\]

\[
+ W_1 \lambda \sum_{k \geq 2} \|\{\hat{\beta}^{(k-1)} - \beta(\tau - 1)\}_{B^{(k)}}\|_1^{0.5}
\]

\[
\leq W_1 K_{0.5} \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}_{T^{(k)}}\|_2^{0.5} + 2W_1 \lambda K_{0.5} (2s_0)^{0.5} \sum_k \|\{\hat{\beta}^{(k)} - \beta(\tau)\}\|_2^{0.5}
\]

\[
\leq C_1(W_1 + 3\lambda W_1) s_0^{0.5} K \left(1 + \frac{c_0}{k(2s_0, c_0)}\right) \left\{ 1 + \frac{1}{k(s_0, c_0)} \right\} \left( s_0 \log \frac{p}{n} + \eta \right)^{0.5},
\]

where the third inequality uses triangle inequality, the fourth and the fifth inequalities comes from the cauchy schwarz inequality with $|T^{(k)}| \leq s_0$ and $|B^{(k)}| \leq 2s_0$, and the last inequality follows from the estimation
To prove (9.16), let where we use triangle inequality and the event \( \tilde{E} \) of (9.15) \( \parallel \hat{\beta}^{(k)} \parallel - \beta^{(k-1)} \)

First, \( \text{pr} \) Lemma 1 implies that \( C \parallel \hat{\beta}^{(k)} \parallel \leq \tilde{E} \parallel (\hat{\beta}^{(k)}) \parallel \leq \parallel \hat{\beta}^{(k)} \parallel \leq \parallel \beta^{(k-1)} \parallel \). By using (9.12) \( \parallel C \parallel \leq \parallel \hat{\beta}^{(k)} \parallel \leq \parallel \beta^{(k-1)} \parallel \) \( \text{pr} \) Lemma 1. And \( \tilde{E} \leq \parallel \beta^{(k-1)} \parallel \). Note that \( \parallel \hat{\beta}^{(k)} \parallel \) is defined in (4.1) with the regularization parameter \( \tilde{\lambda} \) satisfying \( C_2 \sqrt{\log p / n} \leq \tilde{\lambda} \leq C_3 \sqrt{\log p / n} \), where \( C_2 \) and \( C_3 \) are the constants used in Lemma 1. And \( \Lambda_k \) satisfy

\[
6(C_5 + 1)^0.5 \leq \Lambda_k \quad (k = 1, \ldots, K).
\]

So satisfying the event \( E_\eta \), which is defined in (3.2), is equivalent to

\[
\tilde{Q}_k \{ \beta(\tau_k) \} \leq \tilde{Q}_k \{ \hat{\beta}^{(k)} \} + \Lambda_k \frac{\tilde{s}_k \log p}{n} \leq \tilde{Q}_k \{ \beta(\tau_k) \} + \eta \quad (k = 1, \ldots, K).
\]

We define two events as follows:

\[
E_1 := \{ \parallel \hat{\beta}^{(k)} - \beta(\tau_k) \parallel \leq C_4(s_k \log p / n)^0.5 \quad (k = 1, \ldots, K) \},
\]

\[
E_2 := \{ \parallel \hat{\beta}^{(k)} \parallel \leq \tilde{s}_k \leq C_5s_k \quad (k = 1, \ldots, K) \},
\]

where \( C_4 \) and \( C_5 \) are constants defined in Lemma 1. From now on, we assume \( E_1 \cap E_2 \) holds. Indeed, Lemma 1 implies that \( \text{pr}(E_1 \cap E_2) \geq 1 - q_n \). Given the event \( E_1 \cap E_2 \), the following two corollaries hold.

First,

\[
\parallel \hat{\beta}^{(k)} - \beta(\tau_k) \parallel \leq \parallel \hat{\beta}^{(k)} \parallel \leq \parallel \beta(\tau_k) \parallel \leq (C_5 + 1)s_k \quad (k = 1, \ldots, K),
\]

where we use triangle inequality and the event \( E_2 \). Second, we have,

\[
\tilde{s}_k \geq s_k / 2.
\]

To prove (9.16), let \( n_k \) be the number of false negatives of \( \hat{\beta}^{(k)} \), then

\[
n_k^{0.5}C_4 \left( \frac{2\log p}{n} \right)^{0.5} \leq n_k^{0.5} \min_{j \in T^{(k)}} \left| \beta_j(\tau_k) \right| \leq \parallel \hat{\beta}^{(k)} - \beta(\tau_k) \parallel \leq C_4 \left( \frac{s_k \log p}{n} \right)^{0.5} \quad (k = 1, \ldots, K),
\]
where the first inequality comes from the beta-min condition as stated in Theorem 3, and the last inequality holds from \( E_2 \). So \( n_k \leq s_k/2 \), which implies that \( \bar{s}_k \geq s_k - s_k/2 = s_k/2 \). Now we are ready to find \( \eta \) with \( \text{pr}(E_\eta) \) being close to 1. We can first show that the left hand side of (9.14) holds for all \( k \), with probability at least \( 1 - 1/n \) as follows:

\[
\hat{Q}_k \{ \tilde{\beta}^{(k)} \} - \hat{Q}_k \{ \beta(\tau_k) \} \geq \frac{1}{n} \sum_i (x_i[\tau_k - I\{y_i \leq x_i^T \beta(\tau_k)\}])^T \{ \tilde{\beta}^{(k)} - \beta(\tau_k) \} \\
\geq - \frac{1}{n} \sum_i (x_i[\tau_k - I\{y_i \leq x_i^T \beta(\tau_k)\}])\| \tilde{\beta}^{(k)} - \beta(\tau_k) \|_1 \\
\geq - 3 \left( \frac{\log p}{n} \right)^{0.5} (C_5 + 1)^{0.5}s_k^{0.5}C_4 \left( \frac{s_k \log p}{n} \right)^{0.5} \\
\geq - 3(C_5 + 1)^{0.5}C_4 \frac{s_k \log p}{n} \\
(9.17)
\]

For the first inequality, we use the convexity of the function \( \hat{Q}_k \) and that \( \sum_i x_i[\tau_k - I\{y_i \leq x_i^T \beta(\tau_k)\}] \) is a subgradient of \( \hat{Q}_k \) at \( \beta(\tau_k) \). The third inequality holds with probability at least \( 1 - 1/n \); we use Lemma 4 to bound the \( \| \sum_i (x_i[\tau_k - I\{y_i \leq x_i^T \beta(\tau_k)\}] \|_\infty \), and use (9.15). The last inequality uses (9.13) and \( \bar{s}_k \geq s_k/2 \).

To find \( \eta \) which satisfies the right hand side of (9.14) with high probability, we have for all \( k \) that

\[
\hat{Q}_k \{ \tilde{\beta}^{(k)} \} + \Lambda_k(\bar{s}_k \log p)/n \leq \hat{Q}_k \{ \beta(\tau_k) \} + C_4 C_2 (C_5 + 1)^{0.5} s_0 \log p \frac{s_k \log p}{n} + \Lambda_k \bar{s}_k \log p \frac{s_k \log p}{n} \\
\leq \hat{Q}_k \{ \beta(\tau_k) \} + \{ C_4 C_2 (C_5 + 1)^{0.5} + C_5 \Lambda_k \} s_0 \log p \frac{s_k \log p}{n} \\
(9.18)
\]

where the first inequality uses Lemma 5, which holds given \( E_1 \cap E_2 \), and the second inequality comes from \( E_2 \). Combining (9.17) and (9.18) implies that (9.14) holds with probability at least \( 1 - 1/n \), with

\[
\eta = \{ C_4 C_2 (C_5 + 1)^{0.5} + C_5 \Lambda_k \} s_0 \log p \frac{s_k \log p}{n} < \frac{9L_0^2 q^2(2s_0, c_0)}{32U^2}. \\
(9.19)
\]

That is, for \( \eta = \{ C_4 C_2 (C_5 + 1)^{0.5} + C_5 \Lambda_k \} s_0 \log p \frac{s_k \log p}{n} \), we have \( \text{pr}(E_\eta) \geq 1 - 1/n \) given the event \( E_1 \cap E_2 \).

We can also find \( c_0 \) which satisfies Condition 2: given the event \( E_1 \cap E_2 \), \( \overline{W} = 1 \) and \( \overline{\overline{W}} \leq 1 \) because the maximum absolute value of \( p_{\lambda_n}(\cdot) \) is at most 1 and \( p_{\lambda_n}(\tilde{\beta}^{(k)}_j) = 1 \) for any \( j \in \{T^{(k)}\} \), which follows from

\[
|\tilde{\beta}^{(k)}_j| \leq \| \tilde{\beta}^{(k)} - \beta(\tau_k) \|_2 \leq C_4 (s_k \log p/n)^{0.5} < \lambda_n. \\
\]

Therefore, Condition 2 holds with \( c_0 = (1 + 2\lambda)/(1 - 2\lambda) \) given the event \( E_1 \cap E_2 \).
Hence, given the event $E_1 \cap E_2$, we can exploit the results of Theorem 1 with

$$
\eta = \{C_4C_2(C_5 + 1)^{0.5} + C_5\lambda_k\} \frac{s_0 \log p}{n} \quad (k = 1, \ldots, K), \quad c_0 = (1 + 2\lambda)/(1 - 2\lambda).
$$

It provides that given the event $E_1 \cap E_2$, the following convergence bound for all $k$ with probability at least $1 - \text{pr}(E_{\eta}^c) - 1/n$:

$$
\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq \frac{4C_1}{(1 - 2\lambda)^2 k(2s_0, \psi_\lambda)} \left\{ 1 + \frac{1}{k(s_0, \psi_\lambda)} \right\} \left[ \frac{s_0 \log p}{n} + \{C_4C_2(C_5 + 1)^{0.5} + C_5\lambda_k\} \frac{s_0 \log p}{n} \right]^{0.5}
$$

(9.20)$\leq$ \frac{C_6}{(1 - 2\lambda)^2 k(2s_0, \psi_\lambda)} \left\{ 1 + \frac{1}{k(s_0, \psi_\lambda)} \right\} (1 + \Lambda_k) \left( \frac{s_0 \log p}{n} \right)^{0.5},$

where $\psi_\lambda = \frac{1 + 2\lambda}{2 - 2\lambda}$ and $C_6$ is an universal constants that only depend on $C_4, C_5$, and $C_2$. Since $\text{pr}(E_{\eta}) \geq 1 - 1/n$ and $\text{pr}(E_1 \cap E_2) \geq 1 - q_n$, (9.20) holds with probability at least $1 - q_n - 2/n$. This completes the proof of theorem.

**Proof of Theorem 4.** Define the event $E_3$ such that

$$
E_3 := \{\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq \frac{C_6}{(1 - 2\lambda)^2 k(2s_0, \psi_\lambda)} \left\{ 1 + \frac{1}{k(s_0, \psi_\lambda)} \right\} (1 + \Lambda_k) \sqrt{\frac{s_0 \log p}{n}} \},
$$

where $C_6$ is the universal constant used in Theorem 3. Recall the event $E_1$ and $E_2$, which are defined in the proof of Theorem 3.

$$
E_1 := \{\|\hat{\beta}(k) - \beta(\tau_k)\|_2 \leq C_4(s_k \log p/n)^{0.5} \quad (k = 1, \ldots, K),
$$

$$
E_2 := \{\|\hat{\beta}(k)\|_0 = \tilde{s}_k \leq C_5s_k \quad (k = 1, \ldots, K)\},
$$

where $C_4$ and $C_5$ are the constants as defined in Lemma 1. Note that Lemma 1 implies that $\text{pr}(E_1 \cap E_2) \geq 1 - q_n$. From the proof of Theorem 3, we also have $\text{pr}(E_3|E_1 \cap E_2) \geq 1 - 2/n$.

First, we will show that $W_1 = 0$ holds given the event $E_1$. We have,

$$
\min_k \min_{j \in T(k)} |\hat{\beta}_j^{(k)}| \geq \min_k \min_{j \in T(k)} |\beta_j(\tau_k)| - \max_k \|\hat{\beta}(k) - \beta(\tau_k)\|_2
$$

$$
\geq (a\alpha + C_4) \left( \frac{s_0 \log p}{n} \right)^{0.5} - C_4 \left( \frac{s_0 \log p}{n} \right)^{0.5}
$$

(9.21)$= a\alpha \left( \frac{s_0 \log p}{n} \right)^{0.5} = a\lambda_n,$

where the second inequality comes from the event $E_1$ and the beta-min condition as stated in Theorem 4.
We know that given the event \(E\)

objective function of our optimization problem as defined in (9.2). Thus

\[
\hat{E}
\]

where the second inequality holds from the event \(E\). By combining (9.21) and (9.22), we have

(9.22)

Similarly,

\[
\min_{k \geq 2} \min_{j \in B(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| \geq \min_{k \geq 2} \min_{j \in B(k)} |\beta_j^{(k)}(\tau_k) - \beta_j^{(k-1)}(\tau_{k-1})| - \max_k \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2
\]

\[
\geq (a\alpha + 2C_4) \left( \frac{s_0 \log p}{n} \right)^{0.5} - 2C_4 \left( \frac{s_0 \log p}{n} \right)^{0.5}
\]

\[
= a\alpha \left( \frac{s_0 \log p}{n} \right)^{0.5}
\]

\[
\geq a\lambda_n.
\]

(9.22)

By combining (9.21) and (9.22), we have \(\overline{W}_1 = 0\), which implies that \(F(\beta^o) = 0\), where \(F(\cdot)\) is the objective function of our optimization problem as defined in (9.2). Thus \(\beta^o\) becomes one of optimal solution given the event \(E_1\), which implies that \(\beta^o\) is one of optimal solution with probability at least \(1 - q_n\).

We know that given the event \(E_3\), the inequality of false positives that is stated in Theorem 2, holds with

\[
\eta = \{C_4C_2(C_5 + 1)^{0.5} + C_5\Lambda_k\} \frac{s_0 \log p}{n} (k = 1, \ldots, K), \quad c_0 = (1 + 2\lambda)/(1 - 2\lambda).
\]

Given the event \(E_1 \cap E_3\), the inequality in Theorem 2 implies with \(\overline{W}_1 = 0\) that \(\hat{\beta}\) satisfies the following:

(9.23) \[\hat{\beta}_j^{(k)} \in \{0\} \quad (k = 1, \ldots, K), \quad \{\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}\} \in \{0\} \quad (k = 2, \ldots, K).\]

Given the event \(E_3\), we also have

(9.24) \[
\min_k \min_{j \in T^{(k)}} |\hat{\beta}_j^{(k)}| \geq \min_k \min_{j \in T^{(k)}} |\beta_j^{(k)}(\tau_k) - \max_k \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2
\]

\[
\geq \xi \left( \frac{s_0 \log p}{n} \right)^{0.5} - \xi \left( \frac{s_0 \log p}{n} \right)^{0.5}
\]

\[
= 0,
\]

where the second inequality holds from the event \(E_3\) and the beta-min condition as stated in Theorem 4.

Similarly,

(9.25) \[
\min_{k \geq 2} \min_{j \in B(k)} |\hat{\beta}_j^{(k)} - \hat{\beta}_j^{(k-1)}| \geq \min_{k \geq 2} \min_{j \in B(k)} |\beta_j^{(k)}(\tau_k) - \beta_j^{(k-1)}(\tau_{k-1})| - \max_k \|\hat{\beta}^{(k)} - \beta(\tau_k)\|_2
\]

\[
\geq 2\xi \left( \frac{s_0 \log p}{n} \right)^{0.5} - 2\xi \left( \frac{s_0 \log p}{n} \right)^{0.5}
\]

\[
= 0
\]

By combining (9.23), (9.24) and (9.25), we have that \(\hat{\beta}\) provides the exact model structure given the event \(E_1 \cap E_3\). Since \(pr(E_3|E_1 \cap E_2) \geq 1 - 2/n\) and \(pr(E_1 \cap E_2) \geq 1 - q_n\), we have that

\[
pr(E_1 \cap E_3) \geq pr(E_1 \cap E_2 \cap E_3) = pr(E_1 \cap E_2)pr(E_3|E_1 \cap E_2) \geq (1 - q_n)(1 - 2/n) \geq 1 - q_n - 2/n.
\]

34
This completes the proof.

References


