Multi-Stage Dantzig Selector: Supplemental Material

Theorem 1 is fundamental for the rest of the theorems. We first highlight a brief architecture for its proof. Theorem 1 estimates $\|\hat{\beta}-\beta^*\|_p$, which is bounded by the sum of two parts: $\|\hat{\beta}-\beta^*\|_p \leq \|\hat{\beta}-\bar{\beta}\|_p + \|\bar{\beta}-\beta^*\|_p$. We use the upper bounds of these two parts to estimate the bound of $\|\hat{\beta}-\beta^*\|_p$. The analysis in Section 3.2 shows that the first term $\|\hat{\beta}-\bar{\beta}\|_p$ may be much larger than the second term $\|\bar{\beta}-\beta^*\|_p$. In Lemma 1, we estimate the bound of $\|\bar{\beta}-\beta^*\|_p$ and its holding probability. The remaining part of the proof focuses on the estimation of the bound of $\|\hat{\beta}-\bar{\beta}\|_p$. For convenience, we use h to denote $\hat{\beta}-\bar{\beta}$. h can be divided into $h_{\bar{F}_1-T_1}$ and $h_{F_1+T_1}$, where $F_0 \subset F_1 \subset F$. Lemma 3 studies the relationship between $h_{\bar{F}_1-T_1}$ and $h_{F_1+T_1}$, if $\bar{\beta}$ is feasible (Lemma 2 computes its holding probability). Then, Lemma 5 shows that $\|h\|_p$ can be bounded in term of $\|h_{F_1+T_1}\|_p$. In Theorem 7, we estimate the bound of $\|h_{F_1+T_1}\|_p$. Finally, letting $F_1 = F$, we prove Theorem 1.

Lemma 1. *With probability larger than* $1-\eta(\pi\log(s/\eta))^{-1/2}$*, the following holds:*

$$\|\bar{\beta}-\beta^*\|_p \leq \frac{s^{1/p}\sigma\sqrt{2\log(s/\eta)}}{\mu^{(p)}_{(X_F^TX_F)^{1/2},s}} \tag{13}$$

Proof. According to the definition of $\bar{\beta}$, we have

$$\begin{aligned}\bar{\beta}_F &= (X_F^TX_F)^{-1}X_F^Ty = (X_F^TX_F)^{-1}X_F^T(X\beta^*+\epsilon) = (X_F^TX_F)^{-1}X_F^T(X_F\beta_F^*+\epsilon)\\ &= \beta_F^* + (X_F^TX_F)^{-1}X_F^T\epsilon.\end{aligned} \tag{14}$$

It follows that

$$\bar{\beta}_F - \beta_F^* = (X_F^TX_F)^{-1}X_F^T\epsilon \sim N(0,(X_F^TX_F)^{-1}\sigma^2).$$

Since $\|\bar{\beta}-\beta^*\|_p = \|\bar{\beta}_F-\beta_F^*\|_p$, we only need to consider the bound for $\|\bar{\beta}_F-\beta_F^*\|_p$. Let $Z = (X_F^TX_F)^{1/2}(\beta_F^*-\bar{\beta}_F)/\sigma \sim N(0,I)$. We have

$$\begin{aligned}P(\|Z\|_p\geq t) =& (2\pi)^{-s/2}\int_{\|Z\|_p\geq t} e^{-Z^TZ/2}dZ\\ &\leq (2\pi)^{-s/2}\int_{s^{1/p}\|Z\|_\infty\geq t} e^{-Z^TZ/2}dZ \quad (\text{due to } \|Z\|_p\leq s^{1/p}\|Z\|_\infty)\\ &= 1-(2\pi)^{-s/2}\int_{\|Z\|_\infty\leq s^{-1/p}t} e^{-Z^TZ/2}dZ\\ &= 1-\left[(2\pi)^{-1/2}\int_{|Z_i|\leq s^{-1/p}t} e^{-Z_i^2/2}dZ_i\right]^s\\ &= 1-\left[1-2(2\pi)^{-1/2}\int_{s^{-1/p}t}^{\infty} e^{-Z_i^2/2}dZ_i\right]^s\\ &\leq s\left[2(2\pi)^{-1/2}\int_{s^{-1/p}t}^{\infty} e^{-Z_i^2/2}dZ_i\right]\\ &\leq \frac{2s^{1+1/p}}{t(2\pi)^{1/2}}\exp\left[\frac{-t^2}{2s^{2/p}}\right]\end{aligned}$$

Thus the following bound holds with probability larger than $1-\frac{2s^{1+1/p}}{t(2\pi)^{1/2}}\exp\left[\frac{-t^2}{2s^{2/p}}\right]$:

$$\begin{aligned}P(\|Z\|_p\leq t) &= P(\|(X_F^TX_F)^{1/2}(\beta_F^*-\bar{\beta}_F)\|_p\leq t\sigma)\\ &\leq P(\mu^{(p)}_{(X_F^TX_F)^{1/2},s}\|\beta_F^*-\bar{\beta}_F\|_p\leq t\sigma) = P(\|\beta_F^*-\bar{\beta}_F\|_p\leq t\sigma/\mu^{(p)}_{(X_F^TX_F)^{1/2},s})\end{aligned}$$

Taking $t = \sqrt{2\log(s/\eta)}s^{1/p}$, we prove the claim. Note that the presented bound holds for any $p\geq 1$. $\square$

Lemma 2. *With probability larger than* $1 - \eta(\pi \log \frac{m-s}{\eta})^{-1/2}$*, the following bound holds:*

$$\|X_{\bar{F}}^T(X\bar{\beta} - y)\|_\infty \leq \lambda, \tag{15}$$

where $\lambda = \sigma\sqrt{2\log(m-s)/\eta}$*.*

Proof. Let us first consider the probability of $\|X_{\bar{F}}^T(X\bar{\beta} - y)\|_\infty \leq \lambda$. For any $j \in \bar{F}$, define v_j as

$$\begin{aligned}
v_j &= X_j^T(X\bar{\beta} - y) \\
&= X_j^T\left(X_F(X_F^TX_F)^{-1}X_F^T(X_F\beta_F^* + \epsilon) - X_F\beta_F^* - \epsilon\right) \\
&= X_j^T\left(X_F(X_F^TX_F)^{-1}X_F^T - I\right)\epsilon \\
&\sim\ N(0, X_j^T(I - X_F(X_F^TX_F)^{-1}X_F^T)X_j\sigma^2)
\end{aligned}$$

Since $(I - X_F(X_F^TX_F)^{-1}X_F^T)$ is a projection matrix, we have $X_j^T(I - X_F(X_F^TX_F)^{-1}X_F^T)X_j\sigma^2 \leq \sigma^2$. Thus,

$$P(\|X_{\bar{F}}^T(X\bar{\beta} - y)\|_\infty \geq \lambda) = P(\sup_{j \in \bar{F}} |v_j| \geq \lambda) \leq \frac{2(m-s)\sigma}{\lambda(2\pi)^{1/2}} \exp\{-\lambda^2/2\sigma^2\}.$$

Taking $\lambda = \sigma\sqrt{2\log(m-s)/\eta}$ in the inequality above, we prove the claim. $\square$

It follows from the definition of $\bar{\beta}$ that $\|X_F^T(X\bar{\beta} - y)\|_\infty = 0$ always holds. In the following discussion, we assume that the following assumption holds:

Assumption 2. $\bar{\beta}$ *is a feasible solution of the problem* (6)*, if* $F_0 \subset F$*.*

Under the assumption above, both $\|X_{\bar{F}}^T(X\bar{\beta} - y)\|_\infty \leq \lambda$ and $\|X_F^T(X\bar{\beta} - y)\|_\infty = 0$ hold.

Note that this assumption is just used to simplify the description for following proofs. Our proof for the final theorems will substitute this assumption by the probability it holds.

In the following, we introduce an additional set F_1 satisfying $F_0 \subset F_1$ as in [20].

Lemma 3. *Let* $F_0 \subset F$ *and* $h = \hat{\beta} - \bar{\beta}$*. Assume that Assumption 2 holds. Given any index set* F_1 *such that* $F_0 \subset F_1$*, we have the following conclusions:*

$$\begin{aligned}
\|h_{\bar{F}_0 - \bar{F}_1}\|_1 + 2\|\bar{\beta}_{\bar{F}_1}\|_1 \geq& \|h_{\bar{F}_1}\|_1 \\
\|X_{F_0}^TXh\|_\infty =& 0 \\
\|X_{\bar{F}}^TXh\|_\infty \leq& 2\lambda \\
\|X_{\bar{F}_0 - \bar{F}}^TXh\|_\infty \leq& \lambda.
\end{aligned} \tag{16}$$

Proof. Since $\bar{\beta}$ is a feasible solution, the following holds

$$\begin{aligned}
\|\hat{\beta}_{\bar{F}_0}\|_1 &\leq \|\bar{\beta}_{\bar{F}_0}\|_1 \\
\|\hat{\beta}_{\bar{F}_0 - \bar{F}_1}\|_1 + \|\hat{\beta}_{\bar{F}_1}\|_1 &\leq \|\bar{\beta}_{\bar{F}_0 - \bar{F}_1}\|_1 + \|\bar{\beta}_{\bar{F}_1}\|_1 \\
\|\hat{\beta}_{\bar{F}_1}\|_1 &\leq \|h_{\bar{F}_0 - \bar{F}_1}\|_1 + \|\bar{\beta}_{\bar{F}_1}\|_1 \\
\|h_{\bar{F}_1} + \bar{\beta}_{\bar{F}_1}\|_1 &\leq \|h_{\bar{F}_0 - \bar{F}_1}\|_1 + \|\bar{\beta}_{\bar{F}_1}\|_1 \\
\|h_{\bar{F}_1}\|_1 &\leq \|h_{\bar{F}_0 - \bar{F}_1}\|_1 + 2\|\bar{\beta}_{\bar{F}_1}\|_1
\end{aligned}$$

Thus, the first inequality holds. Since

$$X_{F_0}^TXh = X_{F_0}^TX(\hat{\beta} - \bar{\beta}) = X_{F_0}^T(X\hat{\beta} - y) - X_{F_0}^T(X\bar{\beta} - y),$$

the second inequality can be obtained as follows:

$$\|X_{F_0}^TXh\|_\infty \leq \|X_{F_0}^T(X\hat{\beta} - y)\|_\infty + \|X_{F_0}^T(X\bar{\beta} - y)\|_\infty = 0.$$

The third inequality holds since

$$\|X_{\bar{F}}^T X h\|_\infty \leq \|X_{\bar{F}}^T (X\hat{\beta} - y)\|_\infty + \|X_{\bar{F}}^T (X\bar{\beta} - y)\|_\infty \leq 2\lambda.$$

Similarly, the fourth inequality can be obtained as follows:

$$\|X_{\bar{F}_0 - \bar{F}}^T X h\|_\infty \leq \|X_{\bar{F}_0 - \bar{F}}^T (X\hat{\beta} - y)\|_\infty + \|X_{\bar{F}_0 - \bar{F}}^T (X\bar{\beta} - y)\|_\infty \leq \lambda.$$

□

Lemma 4. *Given any $v \in \mathbb{R}^m$, its index set T is divided into a group of subsets T_j's without intersection such that $\bigcup_j T_j = T$. If $\max_j |T_j| \leq l$ and $\max_{i \in T_{j+1}} v_{T_{j+1}}[i] \leq \|v_{T_j}\|_1 / l$ hold for all j's, then we have*

$$\|v_{\bar{T}_1}\|_p \leq \|v\|_1 l^{1/p-1}. \tag{17}$$

Proof. Since $|v_{T_{j+1}}[i]| \leq \|v_{T_j}\|_1 / l$, we have

$$\|v_{T_{j+1}}\|_p^p = \textstyle\sum_{i \in T_{j+1}} v_{T_{j+1}}^p[i] \leq \|v_{T_j}\|_1^p l^{1-p},$$
$$\|v_{T_{j+1}}\|_p \leq \|v_{T_j}\|_1 l^{1/p-1}.$$

Thus,

$$\|v_{\bar{T}_1}\|_p \leq \sum_{j \geq 1} \|v_{T_{j+1}}\|_p \leq \sum_{j \geq 1} \|v_{T_j}\|_1 l^{1/p-1} = \|v\|_1 l^{1/p-1},$$

which proves the claim. □

Note that similar techniques as those in Lemma 4 have been used in the literature [7, 20].

Lemma 5. *Assume that $F_0 \subset F$ and $F_0 \subset F_1$. We divide the index set $\bar{F}_1$ into a group of subsets T_j's such that they satisfy all conditions in Lemma 4. Then the following holds:*

$$\begin{aligned}
\|h_{\bar{F}_1 - T_1}\|_p \leq & l^{1/p-1} \left(|\bar{F}_0 - \bar{F}_1|^{1-1/p} \|h_{\bar{F}_0 - \bar{F}_1}\|_p + 2\|\bar{\beta}_{\bar{F}_1}\|_1 \right), \\
\|h\|_p \leq & \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{p-1} \right]^{1/p} \|h_{F_1 + T_1}\|_p + \left[2l^{1/p-1} + 2 \right] \|\bar{\beta}_{\bar{F}_1}\|_1, \\
\|h\|_p \leq & \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{1-1/p} \right] \|h_{F_1 + T_1}\|_p + 2l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1.
\end{aligned} \tag{18}$$

Proof. Using Lemma 4 with $T = \bar{F}_1$, the first inequality can be obtained as follows using the first inequality in lemma 3 :

$$\begin{aligned}
\|h_{\bar{F}_1 - T_1}\|_p \leq & l^{1/p-1} \|h_{\bar{F}_1}\|_1 \leq l^{1/p-1} \left(\|h_{\bar{F}_0 - \bar{F}_1}\|_1 + 2\|\bar{\beta}_{\bar{F}_1}\|_1 \right) \\
\leq & l^{1/p-1} \left(|\bar{F}_0 - \bar{F}_1|^{1-1/p} \|h_{\bar{F}_0 - \bar{F}_1}\|_p + 2\|\bar{\beta}_{\bar{F}_1}\|_1 \right).
\end{aligned}$$

For any $x \geq 0$, $y \geq 0$, $p \geq 1$, and $a > 0$, it can be easily verified that

$$(x^p + (ax + y)^p)^{1/p} \leq (1 + a^p)^{1/p} x + (1 + a^{-p})^{1/p} y. \tag{19}$$

First we consider the case $F_0 \neq F_1$. In this case, we have

$$\begin{aligned}
\|h\|_p = & \left[\|h_{F_1 + T_1}\|_p^p + \|h_{\bar{F}_1 - T_1}\|_p^p \right]^{1/p} \\
\leq & \left[\|h_{F_1 + T_1}\|_p^p + \left[\left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{1-1/p} \|h_{\bar{F}_0 - \bar{F}_1}\|_p + 2l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1 \right]^p \right]^{1/p} \\
\leq & \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{p-1} \right]^{1/p} \|h_{F_1 + T_1}\|_p + \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{1-p} \right]^{1/p} 2l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1 \\
\leq & \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l} \right)^{p-1} \right]^{1/p} \|h_{F_1 + T_1}\|_p + \left[2l^{1/p-1} + 2|\bar{F}_0 - \bar{F}_1|^{1/p-1} \right] \|\bar{\beta}_{\bar{F}_1}\|_1.
\end{aligned}$$

The first inequality is due to the first claim in this lemma; the second inequality is due to $\|h_{\bar{F}_0-\bar{F}_1}\|_p \leq \|h_{F_1+T_1}\|_p$ and (19); the third inequality holds since $p \geq 1$. Next, we consider the case $F_0 = F_1$. We have

$$\begin{aligned}\|h\|_p =& \left[\|h_{F_1+T_1}\|_p^p + \|h_{\bar{F}_1-T_1}\|_p^p\right]^{1/p} \leq \left[\|h_{F_1+T_1}\|_p^p + \left[2l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right]^p\right]^{1/p} \\ \leq& \|h_{F_1+T_1}\|_p + 2l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1.\end{aligned}$$

Considering two cases above simultaneously, we obtain the second claim as follows:

$$\|h\|_p \leq \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{p-1}\right]^{1/p} \|h_{F_1+T_1}\|_p + \left[2l^{1/p-1} + 2\right] \|\bar{\beta}_{\bar{F}_1}\|_1$$

The third claim can be obtained by using the first claim as follows:

$$\begin{aligned}\|h\|_p \leq& \|h_{F_1+T_1}\|_p + \|h_{\bar{F}_1-T_1}\|_p \leq \|h_{F_1+T_1}\|_p + l^{1/p-1}\left(|\bar{F}_0 - \bar{F}_1|^{1-1/p}\|h_{\bar{F}_0-\bar{F}_1}\|_p + 2\|\bar{\beta}_{\bar{F}_1}\|_1\right) \\ =& \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{1-1/p}\right] \|h_{F_1+T_1}\|_p + 2l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1.\end{aligned}$$

□

Theorem 7. *Under the assumption 1, taking $F_0 \subset F$ and $\lambda = \sigma\sqrt{2\log\left(\frac{m-s}{\eta_1}\right)}$ into the optimization problem* (6)*, for any given index set F_1 satisfying $F_0 \subset F_1 \subset F$, if there exists some l such that $\mu^{(p)}_{A,s_1+l} - \theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p} > 0$ holds where $s_1 = |F_1|$, then with probability larger than $1 - \eta_1'$, the l_p-norm ($1 \leq p \leq \infty$) of the difference between the optimizer of the problem* (6) *and the oracle solution is bounded as*

$$\begin{aligned}\|\hat{\beta} - \bar{\beta}\|_p \leq& \frac{\left[1 + \left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{p-1}\right]^{1/p} \left((|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta^{(p)}_{A,s_1+l,l} l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu^{(p)}_{A,s_1+l} - \theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}} \\ &+ \left[2l^{1/p-1} + 2\right] \|\bar{\beta}_{\bar{F}_1}\|_1\end{aligned} \tag{20}$$

and

$$\begin{aligned}\|\hat{\beta} - \bar{\beta}\|_p \leq& \frac{\left[1 + \left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}\right] \left((|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta^{(p)}_{A,s_1+l,l} l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu^{(p)}_{A,s_1+l} - \theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}} \\ &+ 2l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\end{aligned} \tag{21}$$

and with probability larger than $1 - \eta_1' - \eta_2'$, the l_p-norm ($1 \leq p \leq \infty$) of the difference between the optimizer of the problem (6) *and the true solution is bounded as*

$$\begin{aligned}\|\hat{\beta} - \beta^*\|_p \leq& \frac{\left[1 + \left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{p-1}\right]^{1/p} \left((|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta^{(p)}_{A,s_1+l,l} l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu^{(p)}_{A,s_1+l} - \theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}} \\ &+ \left[2l^{1/p-1} + 2\right] \|\bar{\beta}_{\bar{F}_1}\|_1 + \frac{s^{1/p}}{\mu^{(p)}_{(X_F^T X_F)^{1/2},s}}\sigma\sqrt{2\log(s/\eta_2)}\end{aligned} \tag{22}$$

and

$$\begin{aligned}\|\hat{\beta} - \beta^*\|_p \leq& \frac{\left[1 + \left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}\right] \left((|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta^{(p)}_{A,s_1+l,l} l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu^{(p)}_{A,s_1+l} - \theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}} \\ &+ 2l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1 + \frac{s^{1/p}}{\mu^{(p)}_{(X_F^T X_F)^{1/2},s}}\sigma\sqrt{2\log(s/\eta_2)}\end{aligned} \tag{23}$$

Proof. First, we assume the Assumption 2 and the inequality (13) hold. Divide $\bar{F}_1$ into a group of subsets T_j's without intersection such that $\bigcup_j T_j = \bar{F}_1$, $\max_j |T_j| \leq l$ and $\max_{i\in T_{j+1}} h_{T_{j+1}}[i] \leq \|h_{T_j}\|_1/l$ hold. Note that such a division always exists. Simply, let T_1 be the index set of the largest l elements in h, T_2 be the index set of the largest l elements among the remaining elements, and so on (the size of the last set may be less than l). It is easy to verify that this group of sets satisfy all conditions above. For convenience of presentation, we denote $T_0 = \bar{F}_0 - \bar{F}_1$ and $T_{01} = T_0 + T_1$. Since

$$
\begin{aligned}
&\|X_{T_{01}+F_0}^T Xh\|_p \\
=&\|X_{T_{01}+F_0}^T X_{T_{01}+F_0} h_{T_{01}+F_0} + \sum_{j\geq 2} X_{T_{01}+F_0}^T X_{T_j} h_{T_j}\|_p \\
\geq&\mu_{A,s_1+l}^{(p)} \|h_{T_{01}+F_0}\|_p - \sum_{j\geq 2} \theta_{A,s_1+l,l}^{(p)} \|h_{T_j}\|_p \\
\geq&\mu_{A,s_1+l}^{(p)} \|h_{T_{01}+F_0}\|_p - \theta_{A,s_1+l,l}^{(p)} \sum_{j\geq 2} \|h_{T_j}\|_p \\
\geq&\mu_{A,s_1+l}^{(p)} \|h_{T_{01}+F_0}\|_p - \theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|h_{\bar{F}_1}\|_1 \quad \text{(due to lemma 4)} \\
\geq&\mu_{A,s_1+l}^{(p)} \|h_{T_{01}+F_0}\|_p - \theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \left(\|h_{T_0}\|_1 + 2\|\bar{\beta}_{\bar{F}_1}\|_1\right) \quad \text{(due to lemma 3)} \\
\geq&\mu_{A,s_1+l}^{(p)} \|h_{T_{01}+F_0}\|_p - \theta_{A,s_1+l,l}^{(p)} \left(\frac{l}{|T_0|}\right)^{1/p-1} \|h_{T_0}\|_p - 2\theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1 \\
\geq&\left(\mu_{A,s_1+l}^{(p)} - \theta_{A,s_1+l,l}^{(p)} \left(\frac{l}{|T_0|}\right)^{1/p-1}\right) \|h_{T_{01}+F_0}\|_p - 2\theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1
\end{aligned}
$$

and

$$
\begin{aligned}
&\|X_{T_{01}+F_0}^T Xh\|_p^p \\
=&\|X_{F_0}^T Xh\|_p^p + \|X_{T_{01}\cap F}^T Xh\|_p^p + \|X_{T_{01}\cap \bar{F}}^T Xh\|_p^p \\
\leq&|T_{01}\cap F|\lambda^p + |T_{01}\cap \bar{F}|(2\lambda)^p \quad \text{(due to lemma 3)} \\
\leq&|T_0\cap F|\lambda^p + |T_1\cap F|\lambda^p + |T_0\cap \bar{F}|(2\lambda)^p + |T_1\cap \bar{F}|(2\lambda)^p \quad \text{(due to } F_1 \subset F) \\
\leq&|T_0|\lambda^p + l(2\lambda)^p, \quad \text{(due to } T_0\cap \bar{F} = \emptyset)
\end{aligned}
$$

we have

$$
\begin{aligned}
\|h_{F_1+T_1}\|_p =&\|h_{T_{01}+F_0}\|_p \leq \frac{(|T_0| + 2^p l)^{1/p}\lambda + 2\theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1}{\mu_{A,s_1+l}^{(p)} - \theta_{A,s_1+l,l}^{(p)} \left(\frac{l}{|T_0|}\right)^{1/p-1}} \\
=&\frac{(|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1}{\mu_{A,s_1+l}^{(p)} - \theta_{A,s_1+l,l}^{(p)} \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{1-1/p}}
\end{aligned}
$$

Due to the second inequality in Lemma 5, we have

$$
\begin{aligned}
\|h\|_p \leq& \left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{p-1}\right]^{1/p} \|h_{F_1+T_1}\|_p + \left[2l^{1/p-1} + 2\right] \|\bar{\beta}_{\bar{F}_1}\|_1 \\
=&\frac{\left[1 + \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{p-1}\right]^{1/p} \left((|\bar{F}_0 - \bar{F}_1| + 2^p l)^{1/p}\lambda + 2\theta_{A,s_1+l,l}^{(p)} l^{1/p-1} \|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu_{A,s_1+l}^{(p)} - \theta_{A,s_1+l,l}^{(p)} \left(\frac{|\bar{F}_0 - \bar{F}_1|}{l}\right)^{1-1/p}} + \\
&\left[2l^{1/p-1} + 2\right] \|\bar{\beta}_{\bar{F}_1}\|_1
\end{aligned}
$$

Thus, we can bound $\|\hat{\beta}-\beta^*\|_p$ as

$$\begin{aligned}\|\hat{\beta}-\beta^*\|_p \leq & \|\hat{\beta}-\bar{\beta}\|_p + \|\bar{\beta}-\beta^*\|_p \\ \leq & \frac{\left[1+\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{p-1}\right]^{1/p}\left((|\bar{F}_0-\bar{F}_1|+2^p l)^{1/p}\lambda + 2\theta^{(p)}_{A,s_1+l,l} l^{1/p-1}\|\bar{\beta}_{\bar{F}_1}\|_1\right)}{\mu^{(p)}_{A,s_1+l}-\theta^{(p)}_{A,s_1+l,l}\left(\frac{|\bar{F}_0-\bar{F}_1|}{l}\right)^{1-1/p}} \\ & + \left[2l^{1/p-1}+2\right]\|\bar{\beta}_{\bar{F}_1}\|_1 + \frac{s^{1/p}}{\mu^{(p)}_{(X_F^TX_F)^{1/2},s}}\sigma\sqrt{2\log(s/\eta_2)}.\end{aligned}$$

Similarly, by the third inequality in Lemma 5, we can get the bounds in (21) and (23).

Finally, taking $\lambda=\sigma\sqrt{2\log\left(\frac{m-s}{\eta_1}\right)}$, Lemma 2 with $\eta=\eta_1$ implies that the assumption 2 holds with probability larger than $1-\eta_1'$ and Lemma 1 with $\eta=\eta_2$ implies that (13) holds with probability larger than $1-\eta_2'$. Thus, these two bounds above hold with probabilities larger than $1-\eta_1'$ and $1-\eta_1'-\eta_2'$, respectively. □

Remark 3. *[7] provided a more general upper bound for the Dantzig selector solution in the order of* $\mathcal{O}\left(k^{1/2}\sigma\sqrt{\log m}+r_k^{(2)}(\beta^*)\sqrt{\log m}\right)$*, where* $1\leq k\leq s$ *and* $r_k^{(p)}(\beta)=\left(\sum_{i\in L_k}|\beta_i|^p\right)^{1/p}$ *(*L_k *is the index set of the k largest entries in* β*). We argue that the result in Theorem 7 potentially implies a tighter bound for Dantzig selector. Setting* $F_0=\varnothing$ *(equivalent to the standard Dantzig selector) and* $l=k$ *with* $k=|\bar{F}_1|$ *in Theorem 7, it is easy to verify that the order of the bound for* $\|\hat{\beta}_D-\bar{\beta}\|_p$ *is determined by* $\mathcal{O}\left(k^{1/p}\sigma\sqrt{\log m}+k^{1/p-1}r_k^{(1)}(\bar{\beta})\right)$*, or* $\mathcal{O}\left(k^{1/p}\sigma\sqrt{\log m}+k^{1/p-1}r_k^{(1)}(\beta^*)\right)$ *due to Lemma 1. This bound achieves the same order as the bound of the LASSO solution given by [20], which is the sharpest bound for LASSO.*

We are now ready to prove Theorem 1.
Proof of Theorem 1: Taking $F_1=F$ in theorem 7, we conclude that

$$\|\hat{\beta}-\bar{\beta}\|_p \leq \frac{\left[1+\left(\frac{|\bar{F}_0-\bar{F}|}{l}\right)^{p-1}\right]^{1/p}(|\bar{F}_0-\bar{F}|+l2^p)^{1/p}}{\mu^{(p)}_{A,s+l}-\theta^{(p)}_{A,s+l,l}\left(\frac{|\bar{F}_0-\bar{F}|}{l}\right)^{1-1/p}}\sigma\sqrt{2\log\left(\frac{m-s}{\eta_1}\right)} \tag{24}$$

holds with the probability larger than $1-\eta_1'$ and

$$\begin{aligned}\|\hat{\beta}-\beta^*\|_p \leq \|\hat{\beta}-\bar{\beta}\|_p + \|\bar{\beta}-\beta^*\|_p \leq & \frac{\left[1+\left(\frac{|\bar{F}_0-\bar{F}|}{l}\right)^{p-1}\right]^{1/p}(|\bar{F}_0-\bar{F}|+l2^p)^{1/p}}{\mu^{(p)}_{A,s+l}-\theta^{(p)}_{A,s+l,l}\left(\frac{|\bar{F}_0-\bar{F}|}{l}\right)^{1-1/p}} \\ & \sigma\sqrt{2\log\left(\frac{m-s}{\eta_1}\right)}+\frac{s^{1/p}}{\mu^{(p)}_{(X_F^TX_F)^{1/2},s}}\sigma\sqrt{2\log(s/\eta_2)}.\end{aligned}$$

holds with the probability larger than $1-\eta_1'-\eta_2'$. □

Proof of Theorem 3: From the proof in Theorem 7, the bounds (7) and (8) in Theorem 1 hold with probability 1 if Assumption 2 and the inequality (13) hold. It is easy to verify by Theorem 1 that for any $j\in J$, the following holds: $|\beta_j^*|>\alpha_0\geq\|\hat{\beta}-\bar{\beta}\|_\infty+\|\hat{\beta}-\beta^*\|_\infty$. For any $j\in J$, we have

$$|\hat{\beta}_j|\geq|\beta_j^*|-|\hat{\beta}_j-\beta_j^*|>\|\hat{\beta}-\bar{\beta}\|_\infty+\|\hat{\beta}-\beta^*\|_\infty-|\hat{\beta}_j-\beta_j^*|\geq\|\hat{\beta}-\bar{\beta}\|_\infty\geq\|\hat{\beta}_{\bar{F}}\|_\infty.$$

Thus, there exist at least $|J|$ elements of $\hat{\beta}_F$ larger than $\|\hat{\beta}_{\bar{F}}\|_\infty$. If we pick up the largest $|J|$ elements in $\hat{\beta}$, then all of them correspond to the location of nonzero entries in the true solution β^*. Since Assumption 2 and the inequality (13) hold, the bounds (7) and (8) in Theorem 1 hold with the probability larger than $1 - \eta_1' - \eta_2'$. Thus the claim above holds with probability larger than $1 - \eta_1' - \eta_2'$. Note that the probability will not accumulate, as we only need the holding probability of Assumption 2 and the inequality (13). The proofs below follow the same principle. □

Proof of Theorem 4: From the proof in Theorem 7, the bounds (7) and (8) in Theorem 1 hold with probability 1 if assumption 2 and the inequality (13) hold. In the multi-stage algorithm, the problem in (6) is solved N times. It is easy to verify that the following holds:

$$\alpha_0 \geq \|\hat{\beta}^{(0)} - \bar{\beta}\|_\infty + \|\hat{\beta}^{(0)} - \beta^*\|_\infty.$$

Since $|supp_{\alpha_0}(\beta_J^*)| > 0$, there exists at least 1 element in $\hat{\beta}_J^{(0)}$ larger than $\|\hat{\beta}_{\bar{F}}^{(0)}\|_\infty$. Thus, $F_0^{(1)}$ must be a subset of F. Then, we can verify that

$$\alpha_1 \geq \|\hat{\beta}^{(1)} - \bar{\beta}\|_\infty + \|\hat{\beta}^{(1)} - \beta^*\|_\infty,$$

and $|supp_{\alpha_1}(\beta_J^*)| > 1$ guarantees that there exist at least 2 elements in $\hat{\beta}_J^{(1)}$ larger than $\|\hat{\beta}_{\bar{F}}^{(1)}\|_\infty$. Thus, $F_0^{(2)}$ must be a set of F. Similarly, we can show that $F_0^{(N)}$ is guaranteed to be a subset of F. Since the bounds (7) and (8) in Theorem 1 hold with probability larger than $1 - \eta_1' - \eta_2'$, the claim $F_0^{(N)} \subset F$ holds with probability larger than $1 - \eta_1' - \eta_2'$. □

Proof of Theorem 5: From Theorem 1, the first conclusion holds with probability larger than $1 - \eta_1' - \eta_2'$ by choosing $F_0 = \varnothing$ and $l = s$.

Assuming Assumption 2 and the inequality (13) hold, the bounds (7) and (8) in Theorem 1 hold with probability 1. Since the conditions in Theorem 4 are satisfied, the $|J|$ correct features can be selected from the feature set, i.e., $F_0^{(|J|)} \subset F$. Using the conclusion in (8) of Theorem 1, the bound of the multi-stage method can be estimated by taking $l = |\bar{F}_0 - \bar{F}|$ as follows:

$$\|\hat{\beta}_{mul} - \beta^*\|_p \leq \frac{(2^{p+1}+2)^{1/p}(s-N)^{1/p}}{\mu_{A,2s-N}^{(p)} - \theta_{A,2s-N,s-N}^{(p)}}\sigma\sqrt{2\log\left(\frac{m-s}{\eta_1}\right)} + \frac{s^{1/p}}{\mu_{(X_F^TX_F)^{1/2},s}^{(p)}}\sigma\sqrt{2\log(s/\eta_2)}.$$

Note that since

$$\mu_{A,2s-N}^{(p)} - \theta_{A,2s-N,s-N}^{(p)} \geq \mu_{A,2s}^{(p)} - \theta_{A,2s,s}^{(p)},$$

the following always holds: $\mu_{A,2s-N}^{(p)} - \theta_{A,2s-N,s-N}^{(p)} > 0$. Since Assumption 2 and the inequality (13) hold, the bounds (7) and (8) in Theorem 1 hold with probability larger than $1 - \eta_1' - \eta_2'$. Thus the claim above holds with probability larger than $1 - \eta_1' - \eta_2'$. □

Proof of Theorem 6: First, we assume that Assumption 2 and the inequality (13) hold. In this case, the claim in Theorem 4 holds with probability 1. Since all conditions in Theorem 4 are satisfied, after s iterations, s correct features will be selected (i.e., $F_0^{(N)} = F$) with probability 1. Since all correct features are obtained, the optimization problem in the last iteration can be formulated as:

$$\begin{aligned} \min : &\|\beta_{\bar{F}}\|_1 \\ s.t. : &\|X_{\bar{F}}^T(X\beta - y)\|_\infty \leq \lambda \\ &\|X_F^T(X\beta - y)\|_\infty = 0. \end{aligned} \tag{25}$$

The oracle solution minimizes the objective function to 0. Since Assumption 2 indeed implies that the oracle is a feasible solution, the oracle solution is one optimizer. We can also show that it is the unique optimizer. If there is another optimizer $\beta \neq \bar{\beta}$, then $\beta_{\bar{F}} = 0$ and $\beta_F = (X_F^TX_F)^{-1}X_F^Ty$, which is identical to the definition of the oracle solution. Thus, we conclude that the oracle is the unique optimizer for the optimization problem (25) with probability 1. Since the holding probability of Assumption 2 and the inequality (13) is larger than $1 - \eta_1' - \eta_2'$, the oracle solution can be achieved with the same probability. □