Appendix 0: Proof of Lemma 1

Applying the Sherman-Morrison-Woodbury formula

\[(A + UDV)^{-1} = A^{-1} - A^{-1}U(D^{-1} + VA^{-1})^{-1}VA^{-1},\]

we have

\[r(I_p + X^T X)^{-1} = I_p - X^T (I_n + \frac{1}{r} XX^T)^{-1} X \frac{1}{r} = I_p - X^T (rI_n + XX^T)^{-1} X.\]

Multiplying \(X^T Y\) on both sides, we get

\[r(I_p + X^T X)^{-1} X^T Y = X^T Y - X^T (rI_n + XX^T)^{-1} XX^T Y.\]

The right hand side can be further simplified as

\[X^T Y - X^T (rI_n + XX^T)^{-1} XX^T Y = X^T Y - X^T (rI_n + XX^T)^{-1} Y = rX^T (rI_n + XX^T)^{-1} Y.\]

Therefore, we have

\[(rI_p + X^T X)^{-1} X^T Y = X^T (rI_n + XX^T)^{-1} Y.\]

Appendix A: Proof of Theorem 1

Recall the estimator \(\hat{\beta}^{(HD)} = X^T(XX^T)^{-1}Y = X^T(XX^T)^{-1}X\beta + X^T(XX^T)^{-1}\varepsilon = \xi + \eta\). The following two lemmas will be used to bound \(\xi\) and \(\eta\) respectively.

Lemma 2. Let \(\Phi = X^T(XX^T)^{-1}X\). Assume \(p > c_0 n\) for some \(c_0 > 1\), then for any \(C > 0\) there exists some \(0 < c_1 < 1 < c_2\) and \(c_3 > 0\) such that for any \(t > 0\) and any \(i \in Q, j \neq i\),

\[P\left(|\Phi_{ii} - c_1\kappa \frac{n}{p}\right) \leq 2e^{-Cn}, \quad P\left(|\Phi_{ii} | > c_2\kappa \frac{n}{p}\right) \leq 2e^{-Cn}\]

and

\[P\left(|\Phi_{ij} | > c_4 \kappa t \frac{\sqrt{n}}{p}\right) \leq 5e^{-Cn} + 2e^{-t^2/2},\]

where \(c_4 = \frac{\sqrt{c_2(c_0-c_1)}}{\sqrt{c_3(c_0-1)}}.\)
The proof can be found in the Lemma 4 and 5 in Wang and Leng (2015) for elliptical distributions. The special case of Gaussian is also proved in the Lemma 3 of Wang et al. (2015). Notice that the eigenvalue assumption in Wang and Leng (2015) is not used for proving Lemma 4 and 5.

**Lemma 3.** Assume \( x_i \) follows \( EN(l, \Sigma) \). If \( E[l^{-2}] < M_1 \) for some constant \( M_1 > 0 \), \( \text{var}(\epsilon) = \sigma^2 \) and \( \log p = o(n) \), then for any \( 0 < \alpha < 1 \) we have

\[
P \left( \| \eta \|_\infty \leq \frac{c_1 \kappa \tau n}{4} \right) \geq 1 - O \left( \frac{\sigma^2 \kappa^4 \log p}{\tau^2 \eta^{1-\alpha}} \right),
\]

where \( \tau = \min_{i \in S} |\beta_i| \) and \( \kappa = \text{cond}(\Sigma) \).

To prove Lemma 3 we need the following two propositions.

**Proposition 1.** (Lounici, 2008; Nemirovski, 2000; Akritas et al. (2014)) Let \( Y_i \in \mathbb{R}^p \) be random vectors with zero means and finite variances. Then we have for any \( k \) norm with \( k \in [2, \infty] \) and \( p \geq 3 \), we have

\[
E \left\| \sum_{i=1}^{n} Y_i \right\|_k^2 \leq \tilde{C} \min\{k, \log p\} \sum_{i=1}^{n} E \left\| Y_i \right\|_k^2,
\]

(3)

where \( \tilde{C} \) is some absolute constant.

As each row of \( X \) can be represented as \( X = LZ\Sigma^{1/2} \), where \( L = \text{diag}(l_1/\|z_1\|_2, \cdots, l_n/\|z_n\|_2) \) and \( Z \) is a matrix of independent Gaussian entries, i.e., \( Z \sim N(0, I_p) \). For \( Z \), we have the following result.

**Proposition 2.** Let \( Z \sim N(0, I_p) \), then we have the minimum eigenvalue of \( ZZ^T/p \) satisfies that

\[
P \left( \lambda_{\min}(ZZ^T/p) > \left( 1 - \frac{n}{p} - \frac{t}{p} \right)^2 \right) \geq 1 - 2 \exp(-t^2/2)
\]

for any \( t > 0 \). Assume \( p > c_0 n \) for \( c_0 > 1 \) and take \( t = \sqrt{n} \). When \( n > 4c_0^2/(c_0 - 1)^2 \), we have

\[
P \left( \lambda_{\min}(ZZ^T/p) > c \right) \geq 1 - 2 \exp(-n/2),
\]

(4)

where \( c = \frac{(c_0 - 1)^2}{4c_0^2} \).

The proof follows Corollary 5.35 in Vershynin (2010).

**Proof of Lemma 3.** Let \( A = pX^T(XX^T)^{-1}L \) and \( Z = L^{-1}X\Sigma^{-1/2} \). Then \( \eta = p^{-1}AL^{-1}\epsilon \).

**Part 1.** Bounding \( |A_{ij}| \). Consider the standard SVD on \( Z \) as \( Z = VDU^T \), where \( V \) and \( D \) are \( n \times n \) matrices and \( U \) is a \( p \times n \) matrix. Because \( Z \) is a matrix of iid Gaussian variables, its distribution is invariant under both left and right orthogonal transformation. In particular, for any \( T \in O(n) \), we have

\[
TVDU^T \overset{(d)}{=} VDU^T,
\]
i.e., $V$ is uniformly distributed on $O(n)$ conditional on $U$ and $D$ (they are in fact independent, but we don’t need such a strong condition). Therefore, we have

$$A = pX^T(XX^T)^{-1}L = p\Sigma^{1/2}Z^TL(Z\Sigma Z^T)^{-1}L = p\Sigma^{1/2}UDV^TL(LV,DU^T\Sigma UD^T)L^{-1}L$$
$$= p\Sigma^{1/2}U(U^T\Sigma)U^{-1}1-D^{-1}V^T = \sqrt{p}\Sigma^{1/2}U(U^T\Sigma)^{-1}1(D/\sqrt{p})^{-1}V^T.$$

Because $V$ is uniformly distributed conditional on $U$ and $D$, the distribution of $A$ is also invariant under right orthogonal transformation conditional on $U$ and $D$, i.e., for any $T \in O(n)$, we have

$$A \overset{(d)}{=} AT.$$  \hspace{1cm} (5)

Our first goal is to bound the magnitude of individual entries $A_{ij}$. Let $v_i = e_i^TAA^Te_i$, which is a function of $U$ and $D$ (see below). From (5), we know that $e_i^TA$ is uniformly distributed on the sphere $S^{n-1}(\sqrt{\kappa_i})$ if conditional on $v_i$ (i.e., conditional on $U$, $D$), which implies that

$$e_i^TA \overset{(d)}{=} \sqrt{v_i}\left(\frac{x_1}{\sqrt{\sum_{j=1}^{n}x_j^2}}, \frac{x_2}{\sqrt{\sum_{j=1}^{n}x_j^2}}, \ldots, \frac{x_n}{\sqrt{\sum_{j=1}^{n}x_j^2}}\right),$$  \hspace{1cm} (6)

where $x_j'$s are iid standard Gaussian variables. Thus, $A_{ij}$ can be bounded easily if we can bound $v_i$. Notice that for $v_i$ we have

$$v_i = e_i^TAATe_i = pe_i^T\Sigma^{1/2}U(U^T\Sigma)^{-1}1(D^2/p)^{-1}(U^T\Sigma)^{-1}U^T\Sigma^{1/2}e_i$$
$$= pe_i^TH(U^T\Sigma)^{-1/2}(D^2/p)^{-1}(U^T\Sigma)^{-1/2}HTe_i$$
$$\leq pe_i^T HHTe_i \cdot \lambda_{\min}^{-1}(U^T\Sigma) \cdot \lambda_{\min}^{-1}(D^2/p).$$

Here $H = \Sigma^{1/2}U(U^T\Sigma)^{-1/2}$ is defined the same as in Wang and Leng (2015) and can be bounded as $e_i^THHTe_i \leq c_2n\kappa/p$ with probability $1 - 2\exp(-Cn)$ (see the proof of Lemma 3 in Wang et al. (2015)). Therefore, we have

$$P\left(v_i \leq c_2\kappa^2\lambda_{\min}^{-1}(D^2/p)n\right) \geq 1 - 2\exp(-Cn)$$

Now applying the tail bound and the concentration inequality to (6) we have for any $t > 0$ and any $C > 0$

$$P(|x_j| > t) \leq 2\exp(-t^2/2) \quad P\left(\sum_{j=1}^{n}x_j^2/n \leq c_3\right) \leq \exp(-Cn).$$  \hspace{1cm} (7)

Putting the pieces all together, we have for any $t > 0$ and any $C > 0$ that

$$P\left(\max_{ij} |A_{ij}| \leq \kappa t\sqrt{c_2/c_3} \lambda_{\min}^{-1/2}(D^2/p)\right) \geq 1 - 2n\exp(-t^2/2) - 3p\exp(-Cn).$$

Now according to (4), we can further bound $\lambda_{\min}(D^2/p)$ and obtain that

$$P\left(\max_{ij} |A_{ij}| \leq \sqrt{c_2/c_3}\kappa t\right) \geq 1 - 2np\exp(-t^2/2) - 3p\exp(-Cn) - 2\exp(-n/2).$$  \hspace{1cm} (8)
Part 2. Bounding $\eta$. The second step is to use (8) and Proposition 1 to bound $\eta$. The procedure follows similarly as in Lounici’s paper. We first note that $\|z_i\|^2_p$ follows a chi-square distribution $\chi^2(p)$. We have for any $t$

$$P\left(\frac{\|z_i\|^2_p}{p} \geq 1 + 2\sqrt{\frac{t}{p} + \frac{2t}{p}}\right) \leq e^{-t},$$

from which we know

$$P\left(\max_i p^{-1}\|z_i\|^2_p < 5/2\right) \geq 1 - pe^{-p/4}. \quad (9)$$

Now define $W_j = (A_1p^{-1}\|z_j\|_2I^{-1}_j\epsilon_j, A_2p^{-1}\|z_j\|_2I^{-1}_j\epsilon_j, \ldots, A_hp^{-1}\|z_j\|_2I^{-1}_j\epsilon_j)$. It’s clear that $\eta = \sum_{j=1}^n W_j/p$. Applying Proposition 1 to $W_j$’s with the $l_\infty$ norm and noticing that $l_j$ is independent of $z_j$ we have

$$E\left[\sum_{j=1}^n W_j\right]^2_\infty \leq \log p \sum_{j=1}^n E\|W_j\|^2_\infty \leq \log p \frac{7c_2}{cc_3} \sigma^2 \kappa^2 t^2 \sum_{j=1}^n E[l_j^{-2}] \leq \frac{c_2}{cc_3} \sigma^2 \kappa^2 t^2 M_1^2 n \log p.$$

Using the Markov inequality on $\eta$, we have for any $r > 0$

$$P\left(\|\eta\|_\infty \geq \sqrt{nr} \right) = P\left(\frac{p}{\sqrt{n}} \|\eta\|_\infty \geq r \right) \leq \frac{p^2 E\|\eta\|^2_\infty}{nr^2} = \frac{E\|\sum_{j=1}^n W_j\|^2_\infty}{nr^2} \leq \frac{7c_2 \sigma^2 \kappa^2 M_1^2 t^2 \log p}{cc_3 r^2}.$$

To match our previous result, we take $r = c_1 \sqrt{n \tau} \kappa^{-1}/4$ and $t = n^{\alpha/2}$ for some small $\alpha$,

$$P\left(\|\eta\|_\infty \leq \frac{c_1 \kappa^{-1} \tau n}{4} \right) \geq 1 - \frac{7c_2 \sigma^2 \kappa^4 M_1 \log p}{cc_3 \tau^2 n^{1-\delta}} - 2np \exp(-n^{\delta}/2) - 3p \exp(-Cn) - 2 \exp(-n/2) \geq 1 - O\left(\frac{\sigma^2 \kappa^4 \log p}{\tau^2 n^{1-\delta}}\right).$$

Now we are ready to prove Theorem 1.

**Proof of Theorem 1.** Recall the definition of $\xi$ as $\xi = X^T(XX^T)^{-1}X\beta$. For any $i \in S$ we have

$$\xi_i = e_i^T X^T(XX^T)^{-1}X\beta = \sum_{j \in S} \Phi_{ii}\beta_i + \sum_{j \neq i, j \in S} \Phi_{ij}\beta_j,$$

and for $i \not\in S$,

$$\xi_i = e_i^T X^T(XX^T)^{-1}X\beta = \sum_{j \in S} \Phi_{ij}\beta_j.$$

According to our assumption we have $\min_{i \in S} ||\beta_i|\geq \tau$ and $\var(Y) = \var(X\beta) = \beta^T\Sigma\beta \leq M_0$ for some $M_0$. The latter one implies that

$$M_0 \geq \beta^T \Sigma \beta \geq \lambda_{\min}(\Sigma)\|\beta\|^2_2.$$
Therefore, we have for any $i \in S$

$$|\xi_i| \geq c_1 \kappa^{-1} \frac{t}{p} - \|\beta\|_2 \sqrt{\sum_{j \neq i, j \in S} \Phi_{ij}^2} \geq c_1 \kappa^{-1} \frac{t}{p} - \frac{c_4 \kappa \sqrt{sM_0 t \sqrt{n}}}{\lambda_{\min}(\Sigma)} = \frac{3c_1 \kappa^{-1} t}{4} n,$$

if $t$ is taken to be $t = \frac{c_1 \kappa^{-2} \frac{(\Sigma)_{i,i}}{\lambda_{\min}(\Sigma)}}{4c_4 \sqrt{sM_0 n}} \geq \frac{c_1 \kappa^{-2} \frac{(\Sigma)_{i,i}}{\lambda_{\min}(\Sigma)}}{4c_4 \sqrt{sM_0 n}}$. Hence, one can compute the probability to be greater than $1 - 7 \exp(-Cn) - 2 \exp \left( - \frac{c_1 \kappa^{-2} \frac{(\Sigma)_{i,i}}{\lambda_{\min}(\Sigma)}}{32c_4^2 sM_0 n} \right)$. Similarly, with the same $t$ we can show that for $i \notin S$

$$|\xi_i| \leq \|\beta\|_2 \sqrt{\sum_{j \neq i, j \in S} \Phi_{ij}^2} \leq \frac{c_1 \kappa^{-1} t}{4} n,$$

with probability greater than $1 - 7 \exp(-Cn) - 2 \exp \left( - \frac{c_1 \kappa^{-2} \frac{(\Sigma)_{i,i}}{\lambda_{\min}(\Sigma)}}{32c_4^2 sM_0 n} \right)$. Because we assume $s \log p = O(n)$, it is clear that $O\left( \frac{\sigma^2 \kappa^4 \log p}{\tau^2 n^{1-\alpha}} \right)$ will always be asymptotically greater than $2 \exp \left( - \frac{c_1 \kappa^{-2} \frac{(\Sigma)_{i,i}}{\lambda_{\min}(\Sigma)}}{32c_4^2 sM_0 n} \right)$.

Therefore, using the result from Lemma 3, we can obtain

$$P \left( \min_{i \in S} |\hat{\beta}_i| \geq \frac{c_1 \kappa^{-1} t}{2} \frac{n}{p} \right) \geq 1 - O \left( \frac{\sigma^2 \kappa^4 \log p}{\tau^2 n^{1-\alpha}} \right),$$

and

$$P \left( \max_{i \in S} |\hat{\beta}_i| \leq \frac{c_1 \kappa^{-1} t}{2} \frac{n}{p} \right) \geq 1 - O \left( \frac{\sigma^2 \kappa^4 \log p}{\tau^2 n^{1-\alpha}} \right).$$

Taking $\gamma = \frac{c_1 \kappa^{-1} t}{2} n p$, we have

$$P \left( \min_{i \in S} |\hat{\beta}_i| \geq \gamma \geq \max_{i \in S} |\hat{\beta}_i| \right) \geq 1 - O \left( \frac{\sigma^2 \kappa^4 \log p}{\tau^2 n^{1-\alpha}} \right).$$

\[ \square \]

**Proof of Theorem 2 and 3**

The proof follows similarly as Theorem 1. Recall $X$ possesses a representation as $X = LZ \Sigma^{1/2}$ and $X_{\hat{\mathcal{M}}_d} = LZ_{\hat{\mathcal{M}}_d} \Sigma_{\hat{\mathcal{M}}_d}^{1/2}$, we have

**Lemma 4.** Let $\hat{\mathcal{M}}_d$ be a submodel that contains the true model $\mathcal{M}^*$ and has a size of $d$. Define $A = n(X^T \hat{\mathcal{M}}_d X_{\hat{\mathcal{M}}_d})^{-1} X^T \hat{\mathcal{M}}_d L$ where $X_{\hat{\mathcal{M}}_d}$ is the submatrix formed by columns in $\hat{\mathcal{M}}_d$. Then for any $t > 0$ and $C > 0$, there exists some $c_3 > 0$ such that

$$P \left( \max_{[\mathcal{M}_d] = d, \mathcal{M}^{*} \subseteq \hat{\mathcal{M}}_d} \max_{ij} |A_{ij}| \leq \frac{t}{\sqrt{3} c_3 a_0} \right) \geq 1 - 2dn(p - s)^{d-s} \exp \left( - \frac{t^2}{2} \right) - d(p - s)^{d-s} \exp(-Cn),$$

where $\lambda_0 = \min_{[\mathcal{M}_d] = d, \mathcal{M}^{*} \subseteq \hat{\mathcal{M}}_d} \lambda_{\min}(\Sigma) \lambda_{\min}(Z^T \hat{\mathcal{M}}_d Z_{\hat{\mathcal{M}}_d} / n)$.

**Proof of Lemma 4.** The proof is similar to the argument in Lemma 3. For a given $\hat{\mathcal{M}}_d$, $X_{\hat{\mathcal{M}}_d}$ follows $EN(l, \Sigma_{\hat{\mathcal{M}}_d})$. Similarly, defining $Z_{\hat{\mathcal{M}}_d} = X_{\hat{\mathcal{M}}_d} \Sigma_{\hat{\mathcal{M}}_d}^{1/2}$, then $Z \sim N(0, I_d)$. Assuming the singular value decomposition of $Z$ is $Z_{\hat{\mathcal{M}}_d} = VDU^T$ where $V$ is a $n \times d$ matrix and $D, U$ are $d \times d$
matrices, and conditional on \( U, D, V \) is uniformly distributed on \( V_{n,d} \). Therefore, we have

\[
A = n(X^T_{\mathcal{M}_d} X_{\mathcal{M}_d})^{-1} X^T_{\mathcal{M}_d} = n\Sigma^{-1/2}_{\mathcal{M}_d}(Z^T Z)^{-1} Z^T = n\Sigma^{-1/2}_{\mathcal{M}_d} UD^{-1} V^T.
\]

We observe that

\[
\|e_1^T A\|^2 = n^2 e_1^T \Sigma^{-1/2}_{\mathcal{M}_d} UD^{-2} U^T \Sigma^{-1/2}_{\mathcal{M}_d} e_i \leq n^2 \lambda_{\min}^{-1}(\Sigma) \lambda_{\max}(Z^T_{\mathcal{M}_d} Z_{\mathcal{M}_d})^{-1} \leq \frac{n \lambda_{\min}^{-1}(\Sigma)}{\lambda_{\min}(Z^T_{\mathcal{M}_d} Z_{\mathcal{M}_d}/n)}.
\]

Next, following exactly the same argument in Lemma 3, we know that the distribution of \( A \) is invariant under the right orthogonal transformation and conditional on \( v_i = \|e_i^T A\|_2 \), \( e_i^T A \) is uniformly distributed on \( S^{n-1}(v_i) \). Using the same inequality in (7), we have

\[
P\left( \max_{ij} |A_{ij}| \leq \frac{t}{\sqrt{c_3 \lambda_{\min}(\Sigma) \lambda_{\min}(Z^T_{\mathcal{M}_d} Z_{\mathcal{M}_d}/n)}} \right) \geq 1 - 2dn \exp\left(-\frac{t^2}{2}\right) - d \exp(-Cn).
\]

Now the total number of possible \( \mathcal{M}_d \) is bounded by \((p-s) \times (p-s-1) \times \cdots \times (p-d+1) \leq (p-s)^{(d-s)} \). Therefore, we have

\[
P\left( \max_{|\mathcal{M}_d|=d, M^* \subseteq \mathcal{M}_d} \max_{ij} |A_{ij}| \leq \frac{t}{\sqrt{c_3 \lambda_0}} \right) \geq 1 - 2dn(p-s)^{d-s} \exp\left(-\frac{t^2}{2}\right) - d(p-s)^{d-s} \exp(-Cn),
\]

where \( \lambda_0 = \min_{|\mathcal{M}_d|=d, M^* \subseteq \mathcal{M}_d} \lambda_{\min}(\Sigma) \lambda_{\min}(Z^T_{\mathcal{M}_d} Z_{\mathcal{M}_d}/n) \).

\[\square\]

**Lemma 5** (Garvesh, Wainwright and Yu. (2010) Raskutti et al. (2010)). There exists some absolute constant \( c', c'' > 0 \) such that

\[
\frac{\|Xv\|_2}{\sqrt{n}} \geq \frac{1}{4} \|\Sigma^{1/2} v\|_2 - 9 \rho(\Sigma) \sqrt{\frac{\log p}{n}} \|v\|_1, \quad \forall v \in \mathbb{R}^p,
\]

with probability at least \( 1 - c'' \exp(-c'n) \), where \( \rho(\Sigma) = \max_{i=1,2,\ldots,p} \Sigma_{ii} \).

In our case, for any \( v \) with \( d \) nonzero coordinates, we have \( \|v\|_1 \leq \sqrt{d} \|v\|_2 \), \( \rho(\Sigma) = 1 \) and \( \|\Sigma^{1/2} v\|_2 \geq \kappa^{-1/2} \|v\|_2 \). Therefore,

\[
\frac{\|Xv\|_2}{\sqrt{n}} \geq \left( \frac{\kappa^{-1/2}}{4} - 9 \sqrt{\frac{d \log p}{n}} \right) \|v\|_2, \quad \|v\|_0 \leq d.
\]

**Proof of Theorem 2.** Lemma 5 essentially states that for any \( d \times d \) principal submatrix of \( X \), we can bound its smallest eigenvalue. Therefore, for any selected submodel \( \mathcal{M}_d \) from the first stage, we have with probability at least \( 1 - O(\exp(-c'n)) \)

\[
\min_{|\mathcal{M}_d|=d} \lambda_{\min}(\Sigma) \lambda_{\min}^{1/2}(Z^T_{\mathcal{M}_d} Z_{\mathcal{M}_d}/n) \geq \frac{\kappa^{-1/2}}{4} - 9 \sqrt{\frac{d \log p}{n}} \geq \frac{\kappa^{-1/2}}{8},
\]

as long as \( n \geq 6^d \kappa d \log p \), i.e., \( \lambda_0 \geq \frac{\kappa^{-1}}{64} \), where \( \lambda_0 \) is defined in Lemma 4.

A direct calculation shows that \( \beta^{(OLS)} = \beta + (X^T_{\mathcal{M}_d} X_{\mathcal{M}_d})^{-1} X^T_{\mathcal{M}_d} \varepsilon \). Therefore, we want to bound the error

\[
\hat{\eta} = (X^T_{\mathcal{M}_d} X_{\mathcal{M}_d})^{-1} X^T_{\mathcal{M}_d} \varepsilon = AL^{-1} \varepsilon/n.
\]

Following the same argument as Lemma 3, we define \( W_j = (A_{ij}l_j^{-1}p^{-1}l_j^{-1}) \|z_j\|_2 \varepsilon_j, \cdots, A_{ij}l_j^{-1}p^{-1}l_j^{-1} \|z_j\|_2 \varepsilon_j \) and \( \hat{\eta} = \sum_{j=1}^n W_j/n \). Using Proposition 1, Lemma 4 and (9), we have with probability at least
1 - 2d(p - s)^{d-s} \exp(-t^2/2) - d(p - s)^{d-s} \exp(-Cn)

\[ E\left\| \sum_{j=1}^{n} W_j \right\|_\infty^2 \leq \log d \sum_{j=1}^{n} E\left\| W_j \right\|_\infty^2 \leq \frac{7\sigma^2 n^2 M^2_1 \log d}{c_3 \lambda_0} \leq 448\sigma^{-1} M_1^2 t^2 n \log d. \tag{10} \]

Thus, for any \( r > 0 \)

\[ P\left( \left\| \eta \right\|_\infty \geq \frac{r}{n} \right) = P\left( \left\| \sum_{j=1}^{n} Z_j \right\|_\infty \geq r \right) \leq E\left\| \sum_{j=1}^{n} Z_j \right\|_\infty^2 \leq \frac{448\kappa M_1^2 n \sigma^2 t^2 \log d}{c_3 r^2}. \]

If we take \( t = \sqrt{2(\bar{c} + 3) \log p} \) for any \( \delta \in (0, 1) \), then it is ensured that

\[ 1 - 2dn(p - s)^{d-s} \exp\left(-\frac{t^2}{2}\right) - d(p - s)^{d-s} \exp(-Cn) \geq 1 - 2\exp\left((\bar{c} + 2) \log p - (\bar{c} + 3) \log p \right) - \exp\left((\bar{c} + 1) \log p - Cn \right) = 1 - O\left(\frac{1}{p}\right) \geq 1 - O\left(\frac{1}{n}\right). \]

Now taking \( r = \sigma n^{1-\alpha/2} \) for any \( \alpha \in (0, 1) \) we have

\[ P\left( \left\| \hat{\beta}^{(OLS)} - \beta_{\mathcal{X}_d} \right\|_\infty \leq \frac{\sigma}{n^{\delta/2}} \right) \geq 1 - O\left(\frac{\kappa \log p \log d}{n^{1-\delta}}\right). \tag{11} \]

Consequently, for any \( \alpha > 0 \) we have

\[ \left\| \hat{\beta}^{(OLS)} - \beta_{\mathcal{X}_d} \right\|_\infty \leq \frac{\sigma}{n^{\alpha/2}}. \tag{12} \]

with probability at least \( 1 - O\left(\frac{\kappa \log p \log d}{n^{1-\alpha}}\right) \). So if \( \tau \geq \frac{2\sigma}{n^{\gamma/2}} \), then by choosing \( \gamma' = \frac{\sigma}{n^{\alpha/2}} \) we have

\[ \min_{i \in S} \left| \beta_i^{(OLS)} \right| \geq \gamma' \geq \max_{i \in S} \left| \beta_i^{(OLS)} \right|. \]

\[ \square \]

**Proof of Theorem 3.** Denoting \( X_{\mathcal{X}_d} \) by \( X \), the definition of \( \hat{\beta}(r)^{(Ridge)} \) becomes

\[ \hat{\beta}(r)^{(Ridge)} = (X^T X + \tau I_d)^{-1} X^T X \beta + (X^T X + \tau I_d)^{-1} X^T \varepsilon \]

\[ = \beta - \tau (X^T X + \tau I_d)^{-1} \beta + (X^T X + \tau I_d)^{-1} X^T \varepsilon \]

\[ = \beta - \hat{\xi}(r) + \hat{\eta}(r). \]

For \( \hat{\xi}(r) \) we have

\[ \max |\hat{\xi}(r)| \leq r^2 \beta^T (X^T X + \tau I_d)^{-2} \beta \leq \frac{r^2 \|\beta\|_2^2}{n^2 \lambda_{\min}(X^T X / n + r / n)} \leq \frac{84r^2 \kappa^3 M_0}{n^2} \]

with probability \( 1 - c'' \exp(-c'n) \) if \( n \geq 64^4 \kappa d \log p \). This result is because of Lemma 5 and \( M_0 \geq \text{var}(Y) \geq \|\beta\|_2^2 \lambda_{\max}(\Sigma). \)

For \( \hat{\eta}(r) \), we follow the same technique in the proof of Theorem 2. Basically, one just needs to show a similar result as Lemma 4 exists. Let \( A = n(X^T X)^{-1} X^T \), which is the key quantity in Lemma 4, and \( A = n(X^T X + \tau I_d)^{-1} X^T \). If we can show that \( A \) does not differ too much
from $A$, then the proof is completed. Consider the singular value decomposition directly on $X$ as $X = VDU^T$ (not on $Z$), where $V$ is a $n \times d$ matrix and $D$ and $U$ are $d \times d$ matrices. We then have

$$A = n(UD^2U^T)^{-1}UDVT = nUD^{-1}VT,$$

and

$$\tilde{A} = n(UD^2U^T + rI_d)^{-1}UDVT = nUD^{-1}\left\{ I_d + \frac{r}{n} \left( \frac{D}{\sqrt{n}} \right)^{-2} \right\}^{-1}VT.$$

When $r \leq n\lambda_{\min}(X^TX/n)/2$, we can apply Taylor expansion on the inverse. Thus

$$\tilde{A} = nUD^{-1}\left\{ I_d + \sum_{k=1}^{\infty} \left( \frac{r}{n} \right)^k \left( \frac{D}{\sqrt{n}} \right)^{-2k} \right\}VT$$

$$= A + rUD^{-1}\left( \frac{D}{\sqrt{n}} \right)^{-2}VT + nUD^{-1}\left\{ \sum_{k=2}^{\infty} \left( \frac{r}{n} \right)^k \left( \frac{D}{\sqrt{n}} \right)^{-2k} \right\}VT$$

$$= A + \frac{rU(D/\sqrt{n})^{-3}VT}{n^{1/2}} + nUD^{-1}\left\{ \sum_{k=2}^{\infty} \left( \frac{r}{n} \right)^k \left( \frac{D}{\sqrt{n}} \right)^{-2k} \right\}VT.$$

Clearly, we have

$$\lambda_{\max}\left( \frac{rU(D/\sqrt{n})^{-3}VT}{n^{1/2}} \right) \leq \frac{8^3r\kappa^{3/2}}{\sqrt{n}},$$

and

$$\lambda_{\max}\left[ nUD^{-1}\left\{ \sum_{k=2}^{\infty} \left( \frac{r}{n} \right)^k \left( \frac{D}{\sqrt{n}} \right)^{-2k} \right\}VT \right] \leq \sqrt{n}\lambda_{\min}^{-1}\left( \frac{D}{\sqrt{n}} \right) \sum_{k=2}^{\infty} \frac{r^k}{n^k} \lambda_{\min}^{-1}\left( \frac{D^2}{n} \right)$$

$$\leq \sqrt{n}(8\kappa^2) \sum_{k=2}^{\infty} \left( \frac{8^2r\kappa}{n} \right)^k \leq \frac{\sqrt{n}(8\kappa^2)(\frac{8^2r\kappa}{n})^2}{1 - \frac{8^2r\kappa}{n}}$$

$$\leq \frac{2 \cdot 8^{5/2} \kappa^2 r^2}{n^{3/2}}.$$

The last inequality is because we assume $r \leq n\lambda_{\min}(X^TX/n)/2$. Together, we have

$$\|\tilde{A}\|_{\infty} \leq \|A\|_{\infty} + \frac{8^3r\kappa^{3/2}}{\sqrt{n}} + \frac{2 \cdot 8^{5/2} \kappa^2 r^2}{n^{3/2}},$$

with probability at least $1 - e^{\alpha} \exp(-c'n)$ if $n \geq 6^4\kappa d \log p$ and $r \leq \frac{n}{128\kappa}$. In the proof of Theorem 2, the value of $t$ in Lemma 4 is chosen to be $O(\log p)$. Thus, as long as $r \leq O(\kappa^{-1}\sqrt{n})$, (10) and (11) hold for $\tilde{\eta}(r)$ as well, i.e., for any $\delta \in (0, 1)$ we have

$$P\left( \|\tilde{\eta}(r)\|_{\infty} \leq \frac{\sigma}{n^{\delta/2}} \right) \geq 1 - O\left( \frac{\kappa \log p \log d}{n^{1-\delta}} \right).$$

On the other hand, if we require $r \leq 8^{-2}M_0^{-1/2}\kappa^{-3/2}\sigma^{1/2}n^{1-\delta/4}$, then we have

$$\max |\hat{\xi}(r)| \leq \frac{8^4r^2\kappa^3 M_0}{n^2} \leq \frac{\sigma}{n^{\delta/2}}.$$
Consequently, if the tuning parameter satisfies that

\[ r \leq O\left\{ \min \left( \frac{\sqrt{n}}{\kappa}, \frac{\sigma_1 n^{1-\delta/4}}{8^2 M_0^3 \kappa^3} \right) \right\}, \]

and \( n \geq 6^4 \kappa d \log p \), then we have

\[ P\left( \| \hat{\beta}^{(\text{Ridge})}(r) - \beta_{\tilde{M}_d} \|_\infty \leq \frac{\sigma}{n^{\delta/2}} \right) \geq 1 - O\left( \frac{\kappa \log p \log d}{n^{1-\delta}} \right). \]  \hspace{1cm} (13)

Therefore, if \( \tau \geq \frac{4\sigma}{n^{\delta/2}} \), then by choosing \( \gamma'(r) = \frac{2\sigma}{n^{\delta/2}} \) we have

\[ \min_{i \in S} |\hat{\beta}_i^{(\text{Ridge})}(r)| \geq \gamma' \geq \max_{i \in S} |\hat{\beta}_i^{(\text{Ridge})}(r)|. \]

\[
\]

**Proof of Corollary 1.** As mentioned before, we have \( \hat{\beta}^{(\text{OLS})} = \beta_{\tilde{M}_d} + (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} X_{\tilde{M}_d} \varepsilon \). Because \( \varepsilon \sim N(0, \sigma^2) \) for \( i = 1, 2, \ldots, n \), we have for any \( i \in \tilde{M}_d \),

\[ \hat{\eta}_i = e_i^T (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} X^T_{\tilde{M}_d} \varepsilon \sim N(0, \sigma^2 e_i^T (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} e_i) \equiv \sigma \sqrt{e_i^T (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} e_i} N(0, 1). \] \hspace{1cm} (14)

Likewise in the proof of Lemma 4, we know that as long as \( n \geq 64 \kappa d \log p \)

\[ \lambda_{\min}(X^T_{\tilde{M}_d} X_{\tilde{M}_d}/n) \geq \frac{1}{64 \kappa}. \]

Thus, we have

\[ \max_{i \in \tilde{M}_d} \frac{e_i^T (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} e_i}{n} \leq 64 \kappa/n. \]

Therefore, for any \( t > 0 \) and \( i \in \tilde{M}_d \), with probability at least \( 1 - c'' \exp(-c'n) - 2 \exp(-t^2/2) \) we have

\[ |\hat{\eta}_i| \leq \sigma t \sqrt{e_i^T (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} e_i} \leq \frac{8 \kappa^2 \sigma t}{\sqrt{n}}. \]

Then for any \( \delta > 0 \), if \( n > \log(2c'')/\delta + c' \), then with probability at least \( 1 - \delta \) we have

\[ \max_{i \in \tilde{M}_d} |\hat{\eta}_i| \leq 8 \sigma \sqrt{\frac{2 \kappa \log(4d/\delta)}{n}}. \] \hspace{1cm} (15)

Because \( \sigma \) needs to be estimated from the data, we need to obtain a bound as well. Notice that \( \hat{\sigma}^2 \) is an unbiased estimator for \( \sigma \), and

\[ \hat{\sigma}^2 = \sigma^2 e^T (I_n - X_{\tilde{M}_d} (X^T_{\tilde{M}_d} X_{\tilde{M}_d})^{-1} X_{\tilde{M}_d}) \varepsilon \sim \frac{\sigma^2 \chi^2(n-d)}{n-d}, \]

where \( \chi^2(k) \) denotes a chi-square random variable with degree of freedom \( k \). Using Proposition 5.16 in Vershynin (2010), we can bound \( \hat{\sigma}^2 \) as follows. Let \( \bar{K} = \| \chi^2(1) - 1 \|_{\psi_1} \). There exists some \( c_5 > 0 \) such that for any \( t \geq 0 \) we have,

\[ P\left( \frac{1}{n-d} \right) \leq 2 \exp \left\{ - c_5 \min \left( \frac{t^2(n-d)}{K^2}, \frac{t(n-d)}{K} \right) \right\}. \]
Hence for any $\delta > 0$, if $n > d + 4K^2 \log(2/\delta)/c_5$, then with probability at least $1 - \delta$ we have,

$$|\hat{\sigma}^2 - \sigma^2| \leq \sigma^2/2,$$

which implies that

$$\frac{1}{2}\sigma^2 \leq \hat{\sigma}^2 \leq \frac{3}{2}\sigma^2.$$

Then we know that

$$\max_{i \in \mathcal{M}_d} |\tilde{\eta}_i| \leq 8\sigma \sqrt{\frac{2\kappa \log(4d/\delta)}{n}} \leq 8\sqrt{2}\hat{\sigma} \sqrt{\frac{2\kappa \log(4d/\delta)}{n}} \leq 8\sqrt{3}\sigma \sqrt{\frac{2\kappa \log(4d/\delta)}{n}}.$$

Now define $\gamma' = 8\sqrt{2}\hat{\sigma} \sqrt{\frac{2\kappa \log(4d/\delta)}{n}}$. If the signal $\tau = \min_{i \in S} |\beta_i|$ satisfies that

$$\tau \geq 24\sigma \sqrt{\frac{2\kappa \log(4d/\delta)}{n}},$$

then with probability at least $1 - 2\delta$, for any $i \not\in S$

$$|\hat{\beta}_i| = |\tilde{\eta}_i| \leq 8\sigma \sqrt{\frac{2\kappa \log(4d/\delta)}{n}} \leq \gamma',$$

and for $i \in S$ we have

$$|\hat{\beta}_i| \geq \tau - \max_{i \in \mathcal{M}_d} |\tilde{\eta}_i| \geq 16\sigma \sqrt{\frac{2\kappa \log(4d/\delta)}{n}} \geq \gamma'.$$

\[\square\]

**Proof of Theorem 4**

The result of Theorem 4 can be immediately implied from Theorem 1, 2, 3, (12) and (13).

**References**


