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
\[
r(rI_p + X^TX)^{-1} = I_p - X^T(I_n + \frac{1}{r}XX^T)^{-1}X\frac{1}{r} = I_p - X^T(rI_n + XX^T)^{-1}X.
\]
Multiplying $X^TY$ on both sides, we get
\[
r(rI_p + X^TX)^{-1}X^TY = X^TY - X^T(rI_n + XX^T)^{-1}XX^TY.
\]
The right hand side can be further simplified as
\[
X^TY - X^T(rI_n + XX^T)^{-1}XX^TY
= X^TY - X^T(rI_n + XX^T)^{-1}(rI_n + XX^T - rI_n)Y
= X^TY - X^TY + r(rI_n + XX^T)^{-1}Y = rX^T(rI_n + XX^T)^{-1}Y.
\]
Therefore, we have
\[
(rI_p + X^TX)^{-1}X^TY = X^T(rI_n + XX^T)^{-1}Y.
\]

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 three lemmas will be used to bound $\xi$ and $\eta$ respectively.

**Lemma 2.** Let $\Phi = X^T(XX^T)^{-1}X$. Assume $p > c_0n$ 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
\[
P \left( |\Phi_{ij}| > c_4\kappa t\sqrt{\frac{n}{p}} \right) \leq 5e^{-Cn} + 2e^{-t^2/2},
\]
where $c_4 = \frac{\sqrt{c_2(c_0-c_1)}}{\sqrt{c_3(c_0-1)}}$. 

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 \( \mathcal{E}(L, \Sigma) \). 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

\[
P\left( \|\eta\|_\infty \leq \frac{c_1 \kappa^{-1} \tau^* n}{6p} \right) \geq 1 - O\left( \frac{\sigma^2 \kappa^4 \log p}{\tau^* n^{1-\alpha}} \right),
\]

where \( \tau^* \) is defined as the minimum value for the important signals 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

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

(3)

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

As each row of \( X \) can be represented as \( X = \bar{L}Z\Sigma^{-1/2} \), where \( \bar{L} = \text{diag}(\sqrt{p}L_1/\|z_1\|, \cdots, \sqrt{p}L_n/\|z_n\|) \) 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

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

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 \( A = pX^T(XX^T)^{-1} \bar{L} \) and \( Z = \bar{L}^{-1}X\Sigma^{-1/2} \). Then \( \eta = p^{-1} A\bar{L}^{-1}\epsilon \).

**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 \( T \in \mathcal{O}(n) \), we have

\[
TVDU^T \overset{(d)}{=} VDU^T,
\]
Now according to (4), we can further bound
\[ A = pX^T (XX^T)^{-1}L = p\Sigma \tilde{Z}^T L (LZ \Sigma Z^T L)^{-1} L = p\Sigma \tilde{Z}^T UDV^T L (LV DU^T \Sigma UD V^T L)^{-1} L \]
\[ = p\Sigma \tilde{Z}^T U(U^T \Sigma U)^{-1} D^{-1} V^T = \sqrt{p\Sigma \tilde{Z}^T U(U^T \Sigma U)^{-1}(D/\sqrt{p})^{-1}} V^T. \]

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
\[ A \overset{(d)}{=} AT. \] (5)

Our first goal is to bound the magnitude of individual entries \(A_{ij}\). Let \(v_i = e_i^T A A^T e_i\), which is a function of \(U \) and \(D\) (see below). From (5), we know that \(e_i^T A\) is uniformly distributed on the sphere \(S^{n-1}(\sqrt{v_i})\) if conditional on \(v_i\) (i.e., conditional on \(U, D\)), which implies that
\[ e_i^T A \overset{(d)}{=} \sqrt{v_i} \left( \frac{x_1}{\sqrt{\sum_{j=1}^n x_j^2}}, \frac{x_2}{\sqrt{\sum_{j=1}^n x_j^2}}, \ldots, \frac{x_n}{\sqrt{\sum_{j=1}^n x_j^2}} \right), \]
(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
\[ v_i = e_i^T A A^T e_i = p e_i^T \Sigma \tilde{Z}^T U(U^T \Sigma U)^{-1}(D/\sqrt{p})^{-1}(U^T \Sigma U)^{-1} U^T \Sigma \tilde{Z}^T e_i. \]
\[ = p e_i^T H(U^T \Sigma U)^{-1/2}(D/\sqrt{p})^{-1}(U^T \Sigma U)^{-1/2} H^T e_i \]
\[ \leq p e_i^T H H^T e_i \cdot \lambda_{\text{min}}^{-1}(U^T \Sigma U) \cdot \lambda_{\text{min}}^{-1}(D/\sqrt{p}). \]

Here \(H = \Sigma^{1/2} U(U^T \Sigma U)^{-1/2}\) is defined the same as in Wang and Leng (2015) and can be bounded as \(e_i H H^T e_i \leq c_2 \kappa n/p\) with probability \(1 - 2 \exp(-Cn)\) (see the proof of Lemma 3 in Wang et al. (2015)). Therefore, we have
\[ P \left( v_i \leq c_2 \kappa^2 \lambda_{\text{min}}^{-1}(D/\sqrt{p}) n \right) \geq 1 - 2 \exp(-Cn) \]

Now applying the tail bound and the concentration inequality to (6) we have for any \(t > 0\) and any \(C > 0\)
\[ P(|x_j| > t) \leq 2 \exp(-t^2/2) \quad P \left( \frac{\sum_{j=1}^n x_j^2}{n} \leq c_3 \right) \leq \exp(-Cn). \] (7)

Putting the pieces all together, we have for any \(t > 0\) and any \(C > 0\) that
\[ P \left( \max_{ij} |A_{ij}| \leq \kappa t \sqrt{c_2} \lambda_{\text{min}}^{-1/2}(D/\sqrt{p}) \right) \geq 1 - 2n p \exp(-t^2/2) - 3p \exp(-Cn). \]

Now according to (4), we can further bound \(\lambda_{\text{min}}(D^2/p)\) and obtain that
\[ P \left( \max_{ij} |A_{ij}| \leq \sqrt{c_2} \kappa t \right) \geq 1 - 2n p \exp(-t^2/2) - 3p \exp(-Cn) - 2 \exp(-n/2). \] (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 $\|z_i\|^2$ follows a chi-square distribution $\chi^2(p)$. We have for any $t$

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

from which we know

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

Now define $W_j = (A_{1j}p^{-1/2}\|z_j\|_2L_j^{-1}\epsilon_j, A_{2j}p^{-1/2}\|z_j\|_2L_j^{-1}\epsilon_j, \cdots, A_{pj}p^{-1/2}\|z_j\|_2L_j^{-1}\epsilon_j)$. 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

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

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

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

To match our previous result, we take $r = c_1\sqrt{n}/\tau^k/6$ and $t = n^{(1-\alpha)/2}$ for some small $\alpha$,

$$P\left(\eta \geq \frac{c_1\kappa^{-1}\tau n}{6 p}\right) \geq 1 - \frac{342c_2\sigma^2\kappa^4 M_1 \log p}{n^\alpha} - 2np \exp(-n^{1-\alpha}/2) - 3p \exp(-Cn) - 2 \exp(-n/2) \geq 1 - \frac{\sigma^2\kappa^4 \log p}{\tau^2 n^\alpha}.$$  

Lemma 4. Assume $\text{var}(Y) \leq M_0$. Define $\Phi = X^T(XX^T)^{-1}X$. If $p > c_0 n$ for some $c_0 > 1$, then we have for any $t > 0$

$$P\left(\max_i \sum_{j \neq i} |\Phi_{ij}\beta_j| \geq c_4 \sqrt{M_0}\kappa^{3/2} \sqrt{n} \frac{\sqrt{n}}{p}\right) \leq 2pe^{-t^2/2} + 5pe^{-Cn}.$$  

where $c_4, \kappa$ are defined in Lemma 2.

Proof of Lemma 4. Following Wang and Leng (2015); Wang et al. (2015), we define $H = X^T(XX^T)^{-1/2}$. When $X \sim N(0, \Sigma)$, $H$ follows the $\text{MACG}(\Sigma)$ distribution as indicated in Lemma 3 in Wang et al. (2015) and Theorem 1 in Wang and Leng (2015). For simplicity, we only consider a particular case where $i = 1$.  

For vector $v$ with $v_1 = 0$, we define $v' = (v_2, v_3, \cdots, v_p)^T$ and we can always identify a $(p-1) \times (p-1)$ orthogonal matrix $T'$ such that $T'v' = \|v'\|_2 e_1'$ where $e_1'$ is a $(p-1) \times 1$ unit vector with the first coordinate being 1. Now we define a new orthogonal matrix $T$ as

$$T = \begin{pmatrix} 1 & 0 \\ 0 & T' \end{pmatrix}.$$  

4
and we have
\[ Tv = \begin{pmatrix} 1 & 0 \\ 0 & T' \end{pmatrix} \begin{pmatrix} 0 \\ v' \end{pmatrix} = \|v\|_2 e'_1 = \|v\|_2 e_2. \]

Therefore, we have
\[ \epsilon_1^T H H^T v = \epsilon_1^T T' T H H^T T' T v = e_1^T T' T H H^T T' e_2 = \|v\|_2 e_1^T \tilde{H} \tilde{H}^T e_2. \]

Since \( H \) follows MACG(\( \Sigma \)), \( \tilde{H} = T' H \) follows MACG(\( T' \Sigma T \)) for any fixed \( T \). Therefore, we can apply Lemma 2 again to obtain that
\[ P\left( \left| e_1^T X^T (XX^T)^{-1} X v\right| \geq \|v\|_2 c_4 \kappa t \frac{\sqrt{n}}{p} \right) = P\left( \left| e_1^T H H^T v\right| \geq \|v\|_2 c_4 \kappa \frac{\sqrt{n}}{p} \right) \]
\[ = P\left( \|v\|_2 |\epsilon_1^T \tilde{H} \tilde{H}^T e_2| \geq \|v\|_2 c_4 \kappa \frac{\sqrt{n}}{p} \right) = P\left( \|v\|_2 |\Phi_{12}| \geq \|v\|_2 c_4 \kappa \frac{\sqrt{n}}{p} \right) \]
\[ = P\left( \|\Phi_{12}\| \geq c_4 \kappa \frac{\sqrt{n}}{p} \right) \leq 5e^{-Cn} + 2e^{-t^2/2}. \]

Applying the above result to \( v = (0, \beta^{(-1)}_s) \) we have
\[ \sum_{j \neq 1} |\Phi_{1j} \beta_j| \leq c_4 \kappa t \|\beta\|_2 \frac{\sqrt{n}}{p} \]
with probability at least \( 1 - 5e^{-Cn} - 2e^{-t^2/2} \).

In addition, we know that \( \text{var}(Y) = \beta^T \Sigma \beta + \sigma^2 \leq M_0 \) and thus
\[ \|\beta\|_2 \leq \sqrt{M_0 \kappa}. \]

Consequently, we have
\[ P\left( \max_{i} \sum_{j \neq i} |\Phi_{ij} \beta_j| \geq c_4 \sqrt{M_0 \kappa^2 t} \frac{\sqrt{n}}{p} \right) \leq 2pe^{-t^2/2} + 5pe^{-Cn}. \]

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 \( i \) we have
\[ \xi_i = \epsilon_i^T X^T (XX^T)^{-1} X \beta = \sum_{j \in S} \Phi_{ii} \beta_i + \sum_{j \neq i} \Phi_{ij} \beta_j, \]

For the first term, we have
\[ |\min_{ii} \beta_i| \geq c_1 \kappa^{-1} \tau^* \frac{n}{p} \quad \forall i \in S^* \]
with probability \( 1 - |S^*|e^{-Cn} \) and
\[ |\min_{ii} \beta_i| \leq c_1 \kappa \tau^* \frac{n}{p} \quad \forall i \in S. \]
with probability $1 - |S| e^{-Cn}$. Now, for the second term, using Lemma 4, we have

$$\sum_{j \neq i} |\Phi_{ij} \beta_j| \leq \frac{c_1 \kappa^{-1} \tau^*}{6} \quad \forall i = 1, 2, \cdots, p$$

with probability at least $1 - 2p \exp\left\{-\frac{c_2 \kappa^{-4} \tau^*}{12c_4^2 M_0} n\right\} - 5pe^{-Cn}$. Therefore, we have for any $i \in S^*$

$$|\xi_i| \geq c_1 \kappa^{-1} \tau^* \frac{n}{p} - \frac{c_1 \kappa^{-1} \tau^* n}{6} \geq \frac{5c_1 \kappa^{-1} \tau^* n}{6}.$$

and for $i \in S^*$ we have

$$|\xi_i| \leq c_1 \kappa \tau^* \frac{n}{p} + \frac{c_1 \kappa^{-1} \tau^* n}{6} \leq \frac{7c_1 \kappa^{-1} \tau^* n}{12},$$

where we use the assumption that $\tau^* > 4 \kappa^2 \tau_*$. Now combining the result from Lemma 3, we can obtain

$$P\left(\min_{i \in S^*} |\hat{\beta}_i| \geq \frac{2c_1 \kappa^{-1} \tau^* n}{3} \frac{1}{p}\right) \geq 1 - O\left(\frac{\sigma^2 \kappa^4 \log p}{\tau^* 2p^\alpha}\right),$$

and

$$P\left(\max_{i \in S_*} |\hat{\beta}_i| \leq \frac{7c_1 \kappa^{-1} \tau^* n}{12} \frac{1}{p}\right) \geq 1 - O\left(\frac{\sigma^2 \kappa^4 \log p}{\tau^* 2p^\alpha}\right).$$

Taking $\gamma = \frac{2c_1 \kappa^{-1} \tau^* n}{3} np$, we have

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

\[\square\]

**Proof of Theorem 2 and 3**

For the selected submodel $\hat{\mathcal{M}}_d$, we define $X_d$ to be the variables contained in $\hat{\mathcal{M}}_d$ and $X_{d,c}$ to be variables that are excluded from $\hat{\mathcal{M}}_d$. It is clear that

$$\hat{\beta}_d^{(OLS)} = (X_d^T X_d)^{-1} X_d^T Y = \beta_d + (X_d^T X_d)^{-1} X_d^T \varepsilon + (X_d^T X_d)^{-1} X_d^T X_{d,c} \beta_{d,c} = \beta_d + \eta + \omega.$$

To prove Theorem 2 is essentially to bound $\eta$ and $\omega$. Thus, we need following three lemmas.

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

$$\frac{\|Zv\|_2}{\sqrt{n}} \geq \frac{1}{4} \|\Sigma v\|_2 - 9\rho(\Sigma) \sqrt{\frac{\log p}{n}} \|v\|_1, \quad \forall v \in \mathcal{R}^p,$$

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
\[ \|X^1 v\|_2 \geq \lambda_{\text{min}}^2(\Sigma) \|v\|_2. \] Therefore,
\[
\frac{\|Zv\|_2}{\sqrt{n}} \geq \left( \frac{\lambda_{\text{min}}^2(\Sigma)}{4} - 9 \sqrt{\frac{d \log p}{n}} \right) \|v\|_2, \quad \|v\|_0 \leq d.
\]

Thus, as long as \( n \geq 6^4 k d \log p \), we have
\[
\min_{|\mathcal{M}| \leq d} \lambda_{\text{min}}^1(Z^T_{\mathcal{M}} Z_{\mathcal{M}} / n) \geq \frac{\lambda_{\text{min}}^2(\Sigma)}{8}.
\]

**Lemma 6.** Assume \( E[L^{-12}] \leq M_1 \) and \( e[L^{12}] \leq M_2 \). For any \( \mathcal{M} \) such that \( S^* \subset \mathcal{M} \) and \( |\mathcal{M}| \leq d \), we have for any \( \alpha > 0 \)
\[
P \left( \max_{|\mathcal{M}| \leq d} \|\eta_d\|_\infty \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) = 1 - O \left( \frac{\lambda_+^{-2} d \log d}{n^{\frac{3}{2}(1-\alpha)}} + \frac{M_1 + M_2}{n^{\frac{3}{2}(1-4\alpha)}} \right),
\]
where \( \lambda_+ = \lambda_{\text{min}}(\Sigma) \).

**Proof of Lemma 6.** Define \( A = (X^T_d X_d)^{-1} X^T_d \), we have
\[
\eta = (X^T_d X_d)^{-1} X^T_d \epsilon = A \epsilon.
\]
For \( A \), we can bound its entries as
\[
\max_{ij} \left| A_{ij} \right| \leq \max_{ij} \|e^T_i (X^T_d X_d)^{-1} X^T_d e_j \| \leq \max_{ij} \|e^T_i (X^T_d X_d)^{-1} \| \|X^T_d e_j \|_\infty
\]
\[
\leq \sqrt{d} \max_{ij} \|e^T_i (X^T_d X_d)^{-1}\|_2 \max_{ij} |X^T_d| \leq \frac{\sqrt{d} \lambda^{-1}_{\text{min}}(X^T_d X_d)}{n} \max_{ij} |X^T_d|.
\]
Recall that \( X = LZ^{-1/2} \), where \( L = \text{diag}(\sqrt{p}L_1/\|z_1\|_2, \cdots, \sqrt{p}L_n/\|z_n\|_2) \) and thus \( X_d \) possesses a representation as \( X_d = LZ^{-1/2} \), where \( \Sigma^{-1/2} \) is a \( p \times d \) matrix formed by the selected \( d \) columns of \( \Sigma^{-1/2} \). We can now further bound \( \lambda_{\text{min}}^{-1} \left( \frac{X^T_d X_d}{n} \right) \) as
\[
\lambda_{\text{min}}^{-1} \left( \frac{X^T_d X_d}{n} \right) = \lambda_{\text{min}}^{-1} \left( \frac{\Sigma^T_d Z^T L^T L Z \Sigma^\frac{1}{2}_d}{n} \right)
\]
\[
\leq \left( \lambda_{\text{min}}(L^T L) \lambda_{\text{min}}(\Sigma^T_d Z^T Z \Sigma^\frac{1}{2}_d / n) \right)^{-1}.
\]
Using Lemma 5, it is clear that
\[
\min_{|\mathcal{M}| \leq d} \lambda_{\text{min}}(\Sigma^T_d Z^T Z \Sigma^\frac{1}{2}_d / n) \geq \frac{\lambda_{\text{min}}(\Sigma)}{64} \geq \frac{\lambda_+}{64},
\]
with probability at least \( 1 - O(e^{-c n}) \). In addition, since \( E[L^{-12}] \leq M_1 \) and \( E[L^{12}] \leq M_2 \), we have for any \( k_1 > 0, k_2 > 0 \)
\[
P(L^2 \leq k_1) \leq k_1^6 M_1 \quad \text{and} \quad P(L \geq k_2) \leq \frac{M_2}{k_2^{12}}.
\]
Combining with equation (9) implies that

$$\lambda_{\min}(\hat{L}^T \hat{L}) \geq \frac{2k_1}{5},$$

with probability at least $1 - pe^{-p/4} - nk_1^6 M_1$. Therefore, we have

$$\max_{|\mathcal{A}| \leq d} \lambda_{\min}^{-1} \left( \frac{X_d^T X_d}{n} \right) \leq \frac{162}{\lambda_* k_1},$$

with probability $1 - O(nk_1^6 M_1)$.

For $\max_{ij} |X_d^T|$, we just need to bound $\max_{ij} X_{ij}$. Using the representation $X = \bar{L} Z \Sigma^{1/2}$, we know that

$$X_{ij} = \frac{\sqrt{p} L_i}{\|z\|_2} Z_i \Sigma^{1/2} e_j.$$

It is easy to see that $Z_i \Sigma^{1/2} e_j$ is a Gaussian random variable with mean zero and variance 1, thus for any $t > 0$

$$P(|Z_i \Sigma^{1/2} e_j| \geq t) \leq 2e^{-t^2/2}.$$ 

In addition, $\|z\|_2^2 / p$ follows a $\chi^2(p)$ and we have

$$P\left( \left\| z_i \right\|_2^2 \geq 1 - 2 \sqrt{\frac{t}{p}} \right) \geq 1 - e^{-t}.$$ 

Taking $t = p/4$, we have $\max_i \|z_i\|_2 \sqrt{p} \geq 1/2$ with probability at least $1 - ne^{-p/4}$ and thus

$$P(\max_{ij} |X_{ij}| \leq 4k_2 \sqrt{\log p}) \geq 1 - \frac{M_2 n}{k_2} - 2p - ne^{-p/4}.$$

Combining all pieces of results, we obtain that

$$P\left( \min_{|\mathcal{A}| \leq d} \max_{ij} |A_{ij}| \leq \frac{648k_2 \sqrt{d \log p}}{\lambda_* k_1 n} \right) \geq 1 - O\left( nk_1^6 M_1 + \frac{nM_2}{k_1^2} \right).$$

Following a similar argument in proving Lemma 3, we define $W_j = (A_{1j} \epsilon_j, A_{2j} \epsilon_j, \cdots, A_{dj} \epsilon_j)$ and then

$$\eta = \sum_{j=1}^n W_j.$$

Using Proposition 1, we have

$$E\|\eta\|_\infty^2 = E\left\| \sum_{j=1}^n W_j \right\|_\infty^2 \leq \tilde{C} \log d \sum_{j=1}^n E\|W_j\|_\infty^2 \leq O\left( \frac{\sigma^2 k_2^2 d \log d \log p}{\lambda_1^2 k_1 n} \right).$$

Using the Markov inequality implies that for any $r > 0$

$$P\left( \max_{|\mathcal{A}| \leq d} \|\eta\|_\infty > r \right) \leq \frac{\|\eta\|_\infty^2}{r^2} = O\left( \frac{\sigma^2 k_2^2 d \log d \log p}{\lambda_1^2 k_1^2 r^2} \right) + O\left( nk_1^6 M_1 + \frac{nM_2}{k_1^2} \right).$$
Let \( r = \sigma \sqrt{\frac{\log p}{n^\alpha}} \), \( k_1 = n^{-2(1-\alpha)/\alpha} \) and \( k_2 = n^{1-\alpha} \), we have
\[
P \left( \max_{|\mathcal{M}| \leq d} \|\eta\|_\infty \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) = 1 - O \left( \frac{\lambda_2^* d \log d}{n^\alpha} + \frac{M_1 + M_2}{n^\frac{\alpha}{2}(1 - 4\alpha)} \right)
\]

Lemma 7. Assume \( E[L^{-12}] \leq M_1 \) and \( e[L^{12}] \leq M_2 \). For any \( \hat{\mathcal{M}} \) such that \( S^* \subset \hat{\mathcal{M}} \) and \( |\hat{\mathcal{M}}| \leq d \).
Assume that \( d - |S^*| \leq \hat{c} \) and \( \sum_{i \in S^*} |\beta_i|^4 \leq R \) for some \( \hat{c} \in (0, 1) \), then for any \( \alpha > 0 \), we have
\[
P \left( \max_{|\mathcal{M}| \leq d} \|w\|_2 \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) \leq 1 - O \left( \frac{(M_1 + M_2)R^3}{(\log p)^2 n^{3-4\alpha - \alpha^2}} \right).
\]

Proof of Lemma 7. According to our definition that \( \omega = (X_d^T X_d)^{-1} X_d^T X_{d,c} \beta_{d,c} \), we can directly bound the \( l_2 \) norm of \( \omega \) as
\[
\|\omega\|_2^2 = \frac{1}{n^2} \beta_{d,c}^T X_{d,c}^T X_d (X_d^T X_d)^{-2} X_d^T X_{d,c} \beta_{d,c} \leq \frac{1}{n} \beta_{d,c}^T X_{d,c}^T X_d \beta_{d,c} \lambda_{\min}^{-1} \left( \frac{X_d^T X_d}{n} \right)
\]
where \( \lambda_{\min}^{-1} \left( \frac{X_d^T X_d}{n} \right) \) has already obtained a bound in Lemma 6 as
\[
\max_{|\mathcal{M}| \leq d} \lambda_{\min}^{-1} \left( \frac{X_d^T X_d}{n} \right) \leq \frac{162}{\lambda_* k_1}.
\]

with probability \( 1 - O(nk_1^6 M_1) \). Now for \( \frac{1}{n} \beta_{d,c}^T X_{d,c}^T X_d \beta_{d,c} \) we have
\[
\frac{1}{n} \beta_{d,c}^T X_{d,c}^T X_d \beta_{d,c} = \frac{1}{n} \beta_{d,c}^T \Sigma_{d,c}^{T/2} Z^T \tilde{Z} \Sigma_{d,c}^{1/2} \beta_{d,c} \leq \frac{1}{n} \beta_{d,c}^T \Sigma_{d,c}^{T/2} Z^T Z \Sigma_{d,c}^{1/2} \beta_{d,c} \max_{i} \frac{pL_i^2}{\|z_i\|_2^2}
\]

Since \( Z \sim N(0, I_p) \), we can choose an orthogonal matrix \( Q \) such that \( \beta_{d,c} \Sigma_{d,c}^{1/2} = e_1 Q \beta_{d,c} \Sigma_{d,c}^{1/2} \) and
\[
\frac{1}{n} \beta_{d,c}^T \Sigma_{d,c}^{T/2} Z^T Z \Sigma_{d,c}^{1/2} \beta_{d,c} = \|\beta_{d,c} \Sigma_{d,c}^{1/2} \|_2^2 e_1^T Z^T Z e_1 \alpha_1 \leq \|\beta_{d,c} \|_2^2 \alpha_1 \| e_1 Z^T Z e_1 \|
\]

where \( Z \sim N(0, I_p) \). It is easy to see that for any \( t > 0 \)
\[
P \left( \frac{e_1^T Z^T Z e_1}{n} \leq 1 + 2 \sqrt{\frac{t}{n}} \right) \geq 1 - e^{-t}.
\]

and \( \|\beta_{d,c}\|_2^2 \leq \tau_*^2 t \). Thus, taking \( t = (1 + \hat{c}) \log p \), we have
\[
\max_{|\mathcal{M}| \leq d} \frac{1}{n} \beta_{d,c}^T \Sigma_{d,c}^{T/2} Z^T Z \Sigma_{d,c}^{1/2} \beta_{d,c} \leq 5 \tau_*^2 t R \lambda_*
\]

with probability \( 1 - p^{-1} \) as long as \( n \geq (1 + \hat{c}) \log p \) where \( \hat{c} \) is the upper bound on \( d - |S^*| \). For \( \max_{i} \frac{pL_i^2}{\|z_i\|_2^2} \), we follow the same argument in Lemma 6
\[
P \left( \max_{i} \frac{pL_i^2}{\|z_i\|_2^2} \leq 2k_2^2 \right) \geq 1 - ne^{-p/4} - \frac{nM_2}{k_2^2}.
\]
Putting all pieces together, we have

\[
\max_{|\mathcal{M}| \leq d} \|w\|_2 \leq 36\tau_*^{1/2} R^2 k_1^{1/2} \sqrt{\frac{k_2^2}{k_1}},
\]

with probability at least \(1 - O\left(\frac{nM_2^2}{k_1^2} + nk_1^6M_1\right)\). According to our assumption that \(\tau_* \leq \frac{\sigma}{\sqrt{n}} \sqrt{\log p}\) and taking \(k_1 = \frac{n^{1/4}R^{1/2}}{(\log p)^{1/4}n^{1/4-\alpha/2}}\) and \(k_2 = 1/\sqrt{k_1}\) we have

\[
P\left(\max_{|\mathcal{M}| \leq d} \|w\|_2 \leq \sigma \sqrt{\frac{\log p}{n^\alpha}}\right) \geq 1 - O\left(\frac{(M_1 + M_2)R^3}{(\log p)n^{2\alpha/3 - 4\alpha/2 - 2\alpha}}\right).
\]

We are now ready to prove Theorem 2.

**Proof of Theorem 2.** We just need to combine the results of Lemma 6 and 7, i.e.,

\[
\hat{\beta}_d^{(OLS)} = \beta_d + \eta + \omega,
\]

where

\[
P\left(\max_{|\mathcal{M}| \leq d} \|\eta\|_\infty \leq \sigma \sqrt{\frac{\log p}{n^\alpha}}\right) = 1 - O\left(\frac{\lambda_*^{-2}d \log d + M_1 + M_2}{n^{\alpha/3}(1-\alpha)}\right)
\]

and

\[
P\left(\max_{|\mathcal{M}| \leq d} \|w\|_2 \leq \sigma \sqrt{\frac{\log p}{n^\alpha}}\right) \geq 1 - O\left(\frac{(M_1 + M_2)R^3}{(\log p)n^{2\alpha/3 - 4\alpha/2 - 2\alpha}}\right).
\]

Therefore, we have

\[
P\left(\max_{|\mathcal{M}| \leq d, S \subset \mathcal{M}} \|\hat{\beta}_d^{(OLS)} - \beta_d\|_\infty \leq 2\sigma \sqrt{\frac{\log p}{n^\alpha}}\right) = 1 - O\left(\frac{\lambda_*^{-2}d \log d + M_1 + M_2}{n^{\alpha/3}(1-\alpha)} + \frac{(M_1 + M_2)R^3}{(\log p)n^{2\alpha/3 - 4\alpha/2 - 2\alpha}}\right).
\]

**Proof of Theorem 3.** Recall that \(X_d\) consists of variables contained in \(\hat{\mathcal{M}}_d\), the definition of \(\hat{\beta}(r)^{(\text{Ridge})}\) becomes

\[
\hat{\beta}(r)^{(\text{Ridge})} = (X_d^T X_d + rI_d)^{-1}X_d^T \varepsilon + (X_d^T X_d + rI_d)^{-1}X_d^T X_d c \hat{\beta}_{d,c} = \beta - r(X_d^T X_d + rI_d)^{-1} \beta + (X_d^T X_d + rI_d)^{-1}X_d^T \varepsilon + (X_d^T X_d + rI_d)^{-1}X_d^T X_d c \hat{\beta}_{d,c}
\]

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

\[
\|
\tilde{\xi}(r)
\|_2 \leq r^2 \beta^T (X_d^T X_d + rI_d)^{-2} \beta \leq \frac{r^2 \|
\beta
\|_2^2}{n^2 \lambda_{\min}^2(X_d^T X_d/n + r/n)} \leq \frac{8^4 r^2 \kappa^3 M_0}{n^2}
\]

As proved in Lemma 6, we know that

\[
\max_{|\mathcal{M}| \leq d} \lambda_{\min}(X_d^T X_d/n) \geq \frac{\lambda_* k_1}{162}.
\]
with probability $1 - O(nk_1^6 M_1)$. Adding $r/n$ to the above matrix will only increase the smallest eigenvalue. Thus, we have
\[
\|\tilde{\xi}(r)\|_2 \leq r^2 \beta^T (X_d^T X_d + r I_d)^{-2} \beta \leq \frac{162 r \lambda^* M_0}{n \lambda_4 k_1} = \frac{162 r \kappa M_0}{n k_1}.
\]
Where we used $M_0 \geq \text{var}(Y) \geq \|\beta\|_2^2 \lambda^{-1}(\Sigma)$. Choosing $k_1 = n^{-\frac{2(1-\alpha)}{\alpha}}$, we have
\[
P\left( \max_{|\mathcal{M}| \leq d} \|\tilde{\xi}(r)\|_2 \leq \frac{162 r \kappa M_0}{n^{\frac{1}{2}(7+2\alpha)}} \right) = 1 - O\left( \frac{M_1}{n^{\frac{1}{3}(1-4\alpha)}} \right),
\]
which implies that as long as $r \leq \frac{\sigma n^{(7/9-5\alpha/18)} \sqrt{\log p}}{162 \kappa M_0}$, we have
\[
P\left( \max_{|\mathcal{M}| \leq d} \|\tilde{\xi}(r)\|_2 \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) = 1 - O\left( \frac{M_1}{n^{\frac{1}{3}(1-4\alpha)}} \right).
\]
In addition, the proof for $\|\tilde{\eta}\|_\infty$ and $\|\tilde{\omega}\|_2$ shows that the only key quantity that has changed is $\max_{|\mathcal{M}| \leq d} \lambda_{\min}\left(\frac{X_d^2 X_d}{n}\right)$ which is replaced by $\max_{|\mathcal{M}| \leq d} \lambda_{\min}\left(\frac{X_d^2 X_d + r I_d}{n}\right)$ for $\beta^{(\text{ridge})}$. While the latter is trivially lower bounded by the former, we thus have
\[
P\left( \max_{|\mathcal{M}| \leq d} \|\tilde{\eta}(r)\|_\infty \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) = 1 - O\left( \frac{\lambda_{\min}^{-2} d \log d}{n^{\frac{1}{3}(1-\alpha)}} + \frac{M_1 + M_2}{n^{\frac{1}{3}(1-4\alpha)}} \right)
\]
and
\[
P\left( \max_{|\mathcal{M}| \leq d} \|\tilde{\omega}(r)\|_2 \leq \sigma \sqrt{\frac{\log p}{n^\alpha}} \right) \geq 1 - O\left( \frac{(M_1 + M_2) R^3}{(\log p)^{2\alpha} n^{3-4\alpha-2\alpha}} \right).
\]
Consequently, we have
\[
P\left( \max_{|\mathcal{M}| \leq d, \delta^* \in \mathcal{M}} \|\tilde{\beta}^{(\text{ridge})}_d - \beta_d\|_\infty \leq 3\sigma \sqrt{\frac{\log p}{n^\alpha}} \right) = 1 - O\left( \frac{\lambda_{\min}^{-2} d \log d}{n^{\frac{1}{3}(1-\alpha)}} + \frac{2 M_1 + M_2}{n^{\frac{1}{3}(1-4\alpha)}} + \frac{(M_1 + M_2) R^3}{(\log p)^{2\alpha} n^{3-4\alpha-2\alpha}} \right),
\]
as long as
\[
r \leq \frac{\sigma n^{(7/9-5\alpha/18)} \sqrt{\log p}}{162 \kappa M_0}.
\]

\[\square\]

**Proof of Corollary 1.** As mentioned before, we have $\hat{\beta}^{(\text{OLS})} = \beta_{\mathcal{M}_d} + (X_{\mathcal{M}_d}^T X_{\mathcal{M}_d})^{-1} X_{\mathcal{M}_d} \varepsilon$. Because $\varepsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \cdots, n$, we have for any $i \in \mathcal{M}_d$,
\[
\tilde{\eta}_i = e_i^T (X_{\mathcal{M}_d}^T X_{\mathcal{M}_d})^{-1} X_{\mathcal{M}_d} \varepsilon \sim N(0, \sigma^2 e_i^T (X_{\mathcal{M}_d}^T X_{\mathcal{M}_d})^{-1} e_i) \overset{(d)}{=} \sigma \sqrt{e_i^T (X_{\mathcal{M}_d}^T X_{\mathcal{M}_d})^{-1} e_i} N(0, 1).
\]
Likewise in the proof of Lemma 5, we know that as long as $n \geq 64 \kappa d \log p$
\[
\lambda_{\min}(X_{\mathcal{M}_d}^T X_{\mathcal{M}_d}/n) \geq \frac{1}{64 \kappa}.
\]
Thus, we have
\[
\max_{i \in \tilde{M}_d} e_i^T (X_{\tilde{M}_d}^T X_{\tilde{M}_d})^{-1} e_i \leq 64 \kappa / n.
\]

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

Then for any \( \delta > 0 \), if \( n > \log(2^{c''}/\delta)/c'' \), then with probability at least \( 1 - \delta \) we have
\[
\max_{i \in \tilde{M}_d} |\tilde{\eta}_i| \leq 8 \sqrt{\frac{2 \kappa \log(4d/\delta)}{n}}.
\] (11)

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
\[
\hat{\sigma}^2 = \sigma^2 e^T (I_n - X_{\tilde{M}_d} (X_{\tilde{M}_d}^T X_{\tilde{M}_d})^{-1} X_{\tilde{M}_d}) e \sim \frac{\sigma^2 \chi^2(n - d)}{n - d},
\]
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 \( K = \|\chi^2(1) - 1\|_{\psi_1} \). There exists some \( c_5 > 0 \) such that for any \( t \geq 0 \) we have,
\[
P\left( \frac{\chi^2(n - d)}{n - d} - 1 \geq t \right) \leq 2 \exp \left\{ -c_5 \min \left( \frac{t^2}{K^2}, \frac{t(n - d)}{K} \right) \right\}.
\]

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

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 \( i \notin S \)
\[
|\hat{\beta}_i| = |\tilde{\eta}_i| \leq 8 \sqrt{\frac{2 \kappa \log(4d/\delta)}{n}} \leq \gamma',
\]
and for \( i \in S \) we have
\[
|\hat{\beta}_i| \geq \tau - \max_{i \in \tilde{M}_d} |\tilde{\eta}_i| \geq 16 \sigma \sqrt{\frac{2 \kappa \log(4d/\delta)}{n}} \geq \gamma'.
\]
References


