6 Appendix

The appendix contains a collection of known results as well as the technical proofs.

6.1 Tail bounds for Chi-squared variables

Throughout the paper we will often use one of the following tail bounds for central χ^2 random variables. These are well known and proofs can be found in the original papers.

Lemma 6 ([25]). *Let $X \sim \chi^2_d$. For all $x \geq 0$,*

$$\mathbb{P}[X - d \geq 2\sqrt{dx} + 2x] \leq \exp(-x) \tag{24}$$

$$\mathbb{P}[X - d \leq -2\sqrt{dx}] \leq \exp(-x). \tag{25}$$

Lemma 7 ([21]). *Let $X \sim \chi^2_d$, then*

$$\mathbb{P}[|d^{-1}X - 1| \geq x] \leq \exp(-\frac{3}{16}dx^2), \quad x \in [0, \frac{1}{2}). \tag{26}$$

The following result provide a tail bound for non-central χ^2 random variable with non-centrality parameter ν.

Lemma 8 ([4]). *Let $X \sim \chi^2_d(\nu)$, then for all $x > 0$*

$$\mathbb{P}[X \geq (d + \nu) + 2\sqrt{(d + 2\nu)x} + 2x] \leq \exp(-x) \tag{27}$$

$$\mathbb{P}[X \leq (d + \nu) - 2\sqrt{(d + 2\nu)x}] \leq \exp(-x). \tag{28}$$

6.2 Spectral norms for random matrices

The following results can be found in literature on random matrix theory. We collect some useful results that we use throughout the paper.

Lemma 9 ([9]). *Let $\mathbf{A} \in \mathbb{R}^{n\times k}$ be a random matrix from the standard Gaussian ensemble with $k < n$. Then for all $t > 0$*

$$\mathbb{P}[\Lambda_{\max}(n^{-1}\mathbf{A}'\mathbf{A} - \mathbf{I}_k) \geq f(n, k, t)] \leq 2\exp(-nt^2/2) \tag{29}$$

where $f(n, k, t) = 2(\sqrt{\frac{k}{n}} + t) + (\sqrt{\frac{k}{n}} + t)^2$.

The above results holds for random matrices whose elements are independent and identically distributed $\mathcal{N}(0, 1)$. The result can be extended to random matrices with correlated elements in each row.

Lemma 10 ([39]). *Let $A \in \mathbb{R}^{n\times k}$ be a random matrix with rows sampled iid from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Then for all $t > 0$*

$$\mathbb{P}[\Lambda_{\max}(n^{-1}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}) \geq \Lambda_{\max}(\mathbf{\Sigma})f(n, k, t)] \leq 2\exp(-nt^2/2). \tag{30}$$

Corollary 11. *Let $A \in \mathbb{R}^{n\times k}$ be a random matrix with rows sampled iid from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Then*

$$\mathbb{P}\left[|\Lambda_{\max}\left(n^{-1}\mathbf{A}'\mathbf{A}\right)| \geq 9\Lambda_{\max}\left(\mathbf{\Sigma}\right)\right] \leq 2\exp(-n/2). \tag{31}$$

6.3 Sample covariance matrix

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ be a random matrix whose rows are independent and identically distributed $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. The matrix $\Sigma = (\sigma_{ab})$ and denote $\rho_{ab} = (\sigma_{aa}\sigma_{bb})^{-1/2}\sigma_{ab}$. The following result provides element-wise deviation of the empirical covariance matrix $\widehat{\mathbf{\Sigma}} = n^{-1}\mathbf{X}'\mathbf{X}$ from the population quantity $\mathbf{\Sigma}$.

Lemma 12. *Let $\nu_{ab} = \max\{(1-\rho_{ab})\sqrt{\sigma_{aa}\sigma_{bb}}, (1+\rho_{ab})\sqrt{\sigma_{aa}\sigma_{bb}}\}$. Then or all $t \in [0, \nu_{ab}/2)$*

$$\mathbb{P}\left[|\widehat{\sigma}_{ab} - \sigma_{ab}| \geq t\right] \leq 4\exp\left(-\frac{3nt^2}{16\nu_{ab}^2}\right). \tag{32}$$

The proof is based on Lemma A.3. in [3] with explicit constants.

Proof. Let $x'_{ia} = x_{ia}/\sqrt{\sigma_{aa}}$. Then using (26)

$$\begin{aligned}
&\mathbb{P}[|\frac{1}{n}\sum_{i=1}^{n} x_{ia}x_{ib} - \sigma_{ab}| \geq t] \\
&= \mathbb{P}[|\frac{1}{n}\sum_{i=1}^{n} x'_{ia}x'_{ib} - \rho_{ab}| \geq \frac{t}{\sqrt{\sigma_{aa}\sigma_{bb}}}] \\
&= \mathbb{P}[|\sum_{i=1}^{n}((x'_{ia}+x'_{ib})^2 - 2(1+\rho_{ab})) - ((x'_{ia}-x'_{ib})^2 - 2(1-\rho_{ab}))| \geq \frac{4nt}{\sqrt{\sigma_{aa}\sigma_{bb}}}] \\
&\leq \mathbb{P}[|\sum_{i=1}^{n}((x'_{ia}+x'_{ib})^2 - 2(1+\rho_{ab}))| \geq \frac{2nt}{\sqrt{\sigma_{aa}\sigma_{bb}}}] \\
&\quad + \mathbb{P}[|\sum_{i=1}^{n}((x'_{ia}-x'_{ib})^2 - 2(1-\rho_{ab}))| \geq \frac{2nt}{\sqrt{\sigma_{aa}\sigma_{bb}}}] \\
&\leq 2\mathbb{P}[|\chi_n^2 - n| \geq \frac{nt}{\nu_{ab}}] \leq 4\exp(-\frac{3nt^2}{16\nu_{ab}^2}),
\end{aligned}$$

where $\nu_{ab} = \max\{(1-\rho_{ab})\sqrt{\Sigma_{aa}\Sigma_{bb}}, (1+\rho_{ab})\sqrt{\Sigma_{aa}\Sigma_{bb}}\}$ and $t \in [0, \nu_a/2)$. □

This result implies that, for any $\delta \in (0,1)$, we have

$$\mathbb{P}\left[\sup_{0\leq a<b\leq p} |\widehat{\sigma}_{ab} - \sigma_{ab}| \leq 4\max_{ab}\nu_{ab}\sqrt{\frac{2\log 2d + \log(1/\delta)}{3n}}\right] \geq 1-\delta.$$

As a corollary of Lemma 12, we have a tail bound for sum of product-normal random variables.

Corollary 13. *Let Z_1 and Z_2 be two independent Gaussian random variables and let $X_i \overset{iid}{\sim} Z_1Z_2$, $i = 1\ldots n$. Then for $t \in [0, 1/2)$*

$$\mathbb{P}[|n^{-1}\sum_{i\in[n]} X_i| > t] \leq 4\exp(-\frac{3nt^2}{16}). \tag{33}$$

6.4 Proof of Theorem 1

We introduce some notation before providing the proof of Theorem 1. Consider a $p+1$ dimensional random vector $(Y, \mathbf{X}') = (Y, X_1, \ldots, X_p)$ and assume that

$$\begin{pmatrix} Y \\ \mathbf{X} \end{pmatrix} \sim \mathcal{N}\left(0, \mathbf{\Sigma}_F\right), \qquad \mathbf{\Sigma}_F = \begin{pmatrix} \sigma_{00} & \mathbf{C}' \\ \mathbf{C} & \mathbf{\Sigma} \end{pmatrix}$$

with $\mathbf{C} = (\sigma_{0b})_{b=1}^p = \mathbb{E}Y\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{\Sigma} = (\sigma_{ab})_{a,b=1}^p = \mathbb{E}\mathbf{X}\mathbf{X}'$. Define

$$\mathbf{\Sigma}_F^{-1} = \mathbf{\Omega}_F = \begin{pmatrix} \omega_{00} & \mathbf{P}' \\ P & \mathbf{\Omega} \end{pmatrix},$$

with $\mathbf{P} = (\omega_{0b})_{b=1}^p$ and $\mathbf{\Omega} = (\omega_{ab})_{a,b=1}^p$. The partial correlation between Y and X_j is defined as

$$\rho_j \equiv \text{Corr}\left(Y, X_j \mid X_{\setminus\{j\}}\right) = -\frac{\omega_{0j}}{\sqrt{\omega_{00}\omega_{jj}}} \tag{34}$$

Therefore, nonzero entries of the inverse covariance matrix correspond to nonzero partial correlation coefficients. For Gaussian models, $\rho_j = 0$ correspond to Y and X_j are conditionally independent given $X_{\setminus\{j\}}$. The relationship between the partial correlation estimation and a regression problem can be formulated by the following well-known proposition [26].

Proposition 14. *Consider the following regression model:*

$$Y = \sum_{j=1}^p \beta_j X_j + \epsilon, \quad \epsilon \sim N(0, \text{Var}(\epsilon)) \tag{35}$$

Then ϵ is independent of $X_1, \ldots, X_d$ if and only if for all $j = 1, \ldots, p$

$$\beta_j = -\frac{\omega_{0j}}{\omega_{00}} = \rho_j \sqrt{\frac{\omega_{jj}}{\omega_{00}}}.$$

Furthermore, $\text{Var}(\epsilon) = 1/\omega_{00}$.

Let $\mathbf{\Sigma}_{S^C|S} = \mathbf{\Sigma}_{S^C S^C} - \mathbf{\Sigma}_{S^C S}(\mathbf{\Sigma}_{SS})^{-1}\mathbf{\Sigma}_{SS^C}$ be the conditional covariance of $(X_{S^C}|X_S)$. We are now ready to prove Theorem 1.

Theorem 1. *Consider the regression model in* (1) *with* $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, $\mathbf{x}_i \overset{iid}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$, *and* $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ *with known* $\sigma > 0$, $\mathbf{X}$ *independent of* $\boldsymbol{\epsilon}$. *Assume that*

$$\max_{j \in S^C} |\mathbf{\Sigma}_{jS}\boldsymbol{\beta}_S| + \gamma_n(p, s, \boldsymbol{\beta}, \mathbf{\Sigma}, \delta) < \min_{j \in S} |\mathbf{\Sigma}_{jS}\boldsymbol{\beta}_S|$$

with

$$\begin{aligned}\gamma_n(p, s,\boldsymbol{\beta}, \boldsymbol{\Sigma}, \delta) &= 8\Lambda_{\max}(\boldsymbol{\Sigma}_{SS})\sqrt{\frac{s}{n}}||\boldsymbol{\beta}_S||_2 \max_{j\in S^C}(1 + ||\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}||_2) \\ &+ 4\left(\max_{j\in S^C}\sqrt{\frac{\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\boldsymbol{\Sigma}_{Sj}}{\omega_{00}}} + \max_{j\in S^C}\sqrt{[\boldsymbol{\Sigma}_{S^C|S}]_{jj}\sigma_{00}}\right)\sqrt{\frac{\log\frac{4(p-s)}{\delta}}{3n}} \\ &+ 4\max_{j\in S}\sqrt{\frac{\sigma_{jj}}{\omega_{00}}}\sqrt{\frac{\log\frac{4s}{\delta}}{3n}}\end{aligned} \tag{36}$$

then

$$\mathbb{P}[\widehat{S}(s) = S] \geq 1 - 3\delta - 2\exp(-s/2).$$

Proof. Denoting $\widehat{c}_j = n^{-1}\sum_{i=1}^n y_i x_{ij}$, we would like to establish that

$$\max_{j\notin S}|\widehat{c}_j| \leq \min_{j\in S}|\widehat{c}_j|.$$

Using Proposition 14, for $j \in S^C$ we have $\mathbf{X}'_j = \boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\mathbf{X}'_S + \mathbf{E}'_j$ with $\mathbf{E}_j = (e_{ij})$, $e_{ij} \sim \mathcal{N}(0, [\boldsymbol{\Sigma}_{S^C|S}]_{jj})$. Now

$$\begin{aligned}\widehat{c}_j &= n^{-1}\mathbf{X}_j\mathbf{X}_S\boldsymbol{\beta}_S + n^{-1}\mathbf{X}_j\boldsymbol{\epsilon} \\ &= n^{-1}\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\mathbf{X}'_S(\mathbf{X}_S\boldsymbol{\beta}_S + \boldsymbol{\epsilon}) + n^{-1}\mathbf{E}'_j(\mathbf{X}_S\boldsymbol{\beta}_S + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\Sigma}_{jS}\boldsymbol{\beta}_S + \boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}(\widehat{\boldsymbol{\Sigma}}_{SS} - \boldsymbol{\Sigma}_{SS})\boldsymbol{\beta}_S \\ &\quad + n^{-1}\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\mathbf{X}'_S\boldsymbol{\epsilon} + n^{-1}\mathbf{E}'_j(\mathbf{X}_S\boldsymbol{\beta}_S + \boldsymbol{\epsilon}),\end{aligned} \tag{37}$$

where $\widehat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{X}'\mathbf{X}$ is the empirical covariance matrix. Using (30) with $t = \sqrt{s/n}$ we have that

$$\begin{aligned}&\max_{j\in S^C}|\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}(\widehat{\boldsymbol{\Sigma}}_{SS} - \boldsymbol{\Sigma}_{SS})\boldsymbol{\beta}_S| \\ &\qquad \leq 8\Lambda_{\max}(\boldsymbol{\Sigma}_{SS})\sqrt{\frac{s}{n}}||\boldsymbol{\beta}_S||_2 \max_{j\in S^C}||\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}||_2\end{aligned} \tag{38}$$

with probability at least $1 - 2\exp(-s/2)$. From (33) it follows that

$$\max_{j\in S^C}|n^{-1}\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\mathbf{X}'_S\boldsymbol{\epsilon}| \leq 4\max_{j\in S^C}\sqrt{\frac{\boldsymbol{\Sigma}_{jS}(\boldsymbol{\Sigma}_{SS})^{-1}\boldsymbol{\Sigma}_{Sj}}{\omega_{00}}}\sqrt{\frac{\log\frac{4(p-s)}{\delta}}{3n}} \tag{39}$$

with probability $1 - \delta$ and

$$\max_{j\in S^C}|n^{-1}\mathbf{E}'_j(\mathbf{X}_S\boldsymbol{\beta}_S + \boldsymbol{\epsilon})| \leq 4\max_{j\in S^C}\sqrt{[\boldsymbol{\Sigma}_{S^C|S}]_{jj}\sigma_{00}}\sqrt{\frac{\log\frac{4(p-s)}{\delta}}{3n}} \tag{40}$$

with probability $1-\delta$. Combining (37)-(40)

$$\max_{j\in S^C}|\widehat{c}_j| \leq |\mathbf{\Sigma}_{jS}\boldsymbol{\beta}_S| + 8\Lambda_{\max}(\mathbf{\Sigma}_{SS})\sqrt{\frac{s}{n}}||\boldsymbol{\beta}_S||_2 \max_{j\in S^C}||\mathbf{\Sigma}_{jS}(\mathbf{\Sigma}_{SS})^{-1}||_2$$
$$+ 4\max_{j\in S^C}\sqrt{\frac{\mathbf{\Sigma}_{jS}(\mathbf{\Sigma}_{SS})^{-1}\mathbf{\Sigma}_{Sj}}{\omega_{00}}}\sqrt{\frac{\log\frac{4(p-s)}{\delta}}{3n}} \tag{41}$$
$$+ 4\max_{j\in S^C}\sqrt{[\mathbf{\Sigma}_{S^C|S}]_{jj}\sigma_{00}}\sqrt{\frac{\log\frac{4(p-s)}{\delta}}{3n}}$$

with probability $1-2\delta-2\exp(-s/2)$.

Similarly we can show for $j \in S$ that

$$\min_{j\in S}|\widehat{c}_j| \geq \min|\mathbf{\Sigma}_{SS}\boldsymbol{\beta}_S| - \Lambda_{\max}(\widehat{\mathbf{\Sigma}}_{SS}-\mathbf{\Sigma}_{SS})||\boldsymbol{\beta}_S||_2 - \max|n^{-1}\mathbf{X}_S'\boldsymbol{\epsilon}|$$
$$\geq \min|\mathbf{\Sigma}_{SS}\boldsymbol{\beta}_S| - 8\Lambda_{\max}(\mathbf{\Sigma}_{SS})\sqrt{\frac{s}{n}}||\boldsymbol{\beta}_S||_2 - 4\max_{j\in S}\sqrt{\frac{\sigma_{jj}}{\omega_{00}}}\sqrt{\frac{\log\frac{4s}{\delta}}{3n}} \tag{42}$$

with probability $1-\delta-2\exp(-s/2)$. The theorem now follows from (41) and (42). □

6.5 Proof of Theorem 2

In this section we prove Theorem 2. Define $S_{-j} := S\backslash\{j\}$ and let

$$\widetilde{\sigma}_j^2 := \sigma_{jj} - \mathbf{\Sigma}_{jS_{-j}}(\mathbf{\Sigma}_{S_{-j}S_{-j}})^{-1}\mathbf{\Sigma}_{S_{-j}j}$$

denote the variance of $(X_{j_s}|\mathbf{X}_{S_{-j_s}})$, $j \in S$. The theorem is restated below.

Theorem 2. *Assume that the conditions of Theorem 1 are satisfied. Let*

$$\iota = \sqrt{\frac{16\log(16/\delta)}{3(n-s+1)}}$$

and assume that $\iota < \frac{1}{2}$. *Furthermore, assume that*

$$\max_{j\in S}\left\{\frac{2\sigma^2\log(4n/\delta)}{\beta_j^2\widetilde{\sigma}_j^2(1-\iota)} + \frac{2\sigma\sqrt{2(1+\iota)\log(8n/\delta)}}{\beta_j\widetilde{\sigma}_j(1-\iota)}\right\} < 1.$$

Then

$$\mathbb{P}[\widehat{S}(\widehat{s}_n) = S] \geq 1-4\delta-2\exp(-s/2).$$

Proof. Define the event

$$\mathcal{E}_n = \{\widehat{S}(s) = S\}. \tag{43}$$

From Theorem 1,

$$\mathbb{P}[\mathcal{E}_n^C] \leq 3\delta + 2\exp(-s/2). \tag{44}$$

We proceed to show that for some small $\delta' > 0$

$$\mathbb{P}[\widehat{s}_n \neq s] \leq \mathbb{P}[\widehat{s}_n \neq s|\mathcal{E}_n]\mathbb{P}[\mathcal{E}_n] + \mathbb{P}[\mathcal{E}_n^C] \leq \delta', \tag{45}$$

which will prove the theorem together with (44). An upper bound on $\mathbb{P}[\widehat{s}_n \neq s|\mathcal{E}_n]$ is constructed by combining upper bounds on $\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n]$ and $\mathbb{P}[\widehat{s}_n < s|\mathcal{E}_n]$.

Let $\tau = 2\sigma^2 \log \frac{4n}{\delta}$. From $\{\widehat{s}_n > s|\mathcal{E}_n\} \subseteq \cup_{k=s}^{p-1}\{\widehat{\xi}_n(k) \geq \tau|\mathcal{E}_n\}$ follows that

$$\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n] \leq \sum_{k=s}^{p-1} \mathbb{P}[\widehat{\xi}_n(k) \geq \tau|\mathcal{E}_n]. \tag{46}$$

Recalling definitions of $\widehat{V}_n(k)$ and $\widehat{\mathbf{H}}_n(k)$ from p. 3, for a fixed $s \leq k \leq p-1$, $\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k)$ is the projection matrix from $\mathbb{R}^n$ to $\widehat{V}_n(k+1) \cap \widehat{V}_n(k)^{\perp}$. Recall also that we are using the second half of the sample to estimate $\widehat{s}_n$, which implies that the projection matrix $\widehat{\mathbf{H}}(k)$ is independent of $\boldsymbol{\epsilon}$ for all k. Now, exactly one of the two events $\{\widehat{V}_n(k) = \widehat{V}_n(k+1)\}$ and $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\}$ occur. On the event $\{\widehat{V}_n(k) = \widehat{V}_n(k+1)\}$, $\widehat{\xi}_n(k) = 0$. We analyze the event $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \cap \mathcal{E}_n$ by conditioning on $\mathbf{X}$. Since $\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k)$ is a rank one projection matrix

$$\widehat{\xi}_n(k) = ||(\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k))\mathbf{y}||_2^2 = ||(\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k))\boldsymbol{\epsilon}||_2^2 \stackrel{d}{=} \sigma^2\chi_1^2.$$

Furthermore, $(\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k))\boldsymbol{\epsilon} \perp\!\!\!\perp (\widehat{\mathbf{H}}(k'+1) - \widehat{\mathbf{H}}(k'))\boldsymbol{\epsilon}$, $k \neq k'$. It follows that for any realization of the sequences $\widehat{V}_n(1), \ldots, \widehat{V}_n(p)$,

$$\begin{aligned}
&\sum_{k=s}^{p-1} \mathbb{P}[\widehat{\xi}_n(k) \geq \tau|\mathcal{E}_n] \\
&\quad = \sum_{k=s}^{p-1} \mathbb{P}[\widehat{\xi}_n(k) \geq \tau|\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \cap \mathcal{E}_n]\mathbb{P}[\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)] \\
&\quad = \mathbb{P}[\sigma^2\chi_1^2 \geq \tau]\mathbb{E}\sum_{k=s}^{p-1} \mathbb{I}\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \\
&\quad \leq n\mathbb{P}[\sigma^2\chi_1^2 \geq \tau],
\end{aligned}$$

where the first equality follows since $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\}$ is independent of $\mathcal{E}_n$. Combining with (46) gives

$$\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n] \leq n\mathbb{P}[\sigma^2\chi_1^2 \geq \tau] \leq \delta/2 \tag{47}$$

using a standard normal tail bound.

Next, we focus on bounding $\mathbb{P}[\widehat{s}_n < s|\mathcal{E}_n]$. Since $\{\widehat{s}_n < s|\mathcal{E}_n\} \subset \{\widehat{\xi}_n(s-1) < \tau|\mathcal{E}_n\}$, we can bound $\mathbb{P}[\widehat{\xi}_n(s-1) < \tau|\mathcal{E}_n]$. Using the definition of $\widehat{\mathbf{H}}(s)$ it is straightforward to obtain that

$$(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\mathbf{y} = (\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))(\mathbf{X}_{j_s}\beta_{j_S} + \boldsymbol{\epsilon}).$$

Using Proposition 14, we can write $\mathbf{X}'_{j_s} = \boldsymbol{\Sigma}_{j_s S_{-j_s}} (\boldsymbol{\Sigma}_{S_{-j_s} S_{-j_s}})^{-1} \mathbf{X}'_{S_{-j_s}} + \mathbf{E}'$ where $\mathbf{E} = (e_i)$, $e_i \overset{iid}{\sim} \mathcal{N}(0, \widetilde{\sigma}^2_{j_s})$. Then

$$
\begin{aligned}
(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\mathbf{y} &= (\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))(\mathbf{E}\beta_{j_S} + \boldsymbol{\epsilon}) \\
&= (\mathbf{I}_n - \widehat{\mathbf{H}}(s-1))\mathbf{E}\beta_{j_S} + (\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\boldsymbol{\epsilon}.
\end{aligned}
$$

Define

$$
T_1 = \beta^2_{j_S} \mathbf{E}'(\mathbf{I}_n - \widehat{\mathbf{H}}(s-1))\mathbf{E}
$$

and

$$
T_2 = \boldsymbol{\epsilon}'(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\boldsymbol{\epsilon}.
$$

Conditional on $\mathbf{X}_{S_{-j_s}}$, $T_1 \overset{d}{=} \beta^2_{j_S} \widetilde{\sigma}^2_{j_s} \chi^2_{n-s+1}$ since $\mathbf{E} \perp\!\!\!\perp \mathbf{X}_{S_{-j_s}}$, and conditional on $\mathbf{X}_S$, $T_2 \overset{d}{=} \sigma^2 \chi^2_1$. Define the events

$$
\mathcal{A}_1 = \{\beta^2_{j_S} \widetilde{\sigma}^2_{j_s}(1-\iota) \le T_1 \le \beta^2_{j_S} \widetilde{\sigma}^2_{j_s}(1+\iota)\}
$$

and

$$
\mathcal{A}_2 = \{T_2 \le 2\sigma^2 \log \frac{8n}{\delta}\}.
$$

From Eq. (26), $\mathbb{P}[\mathcal{A}_1(\iota)^C] \le \delta/4$, and using a normal tail bound, $\mathbb{P}[\mathcal{A}_2^C] < \delta/4$. Setting

$$
\widetilde{\tau} = \tau + 2\beta_{j_S} \widetilde{\sigma}_{j_s} \sigma \sqrt{2(1+\iota) \log \frac{8n}{\delta}},
$$

under the assumptions of theorem

$$
\begin{aligned}
\mathbb{P}[\widehat{\xi}_n(s-1) < \tau | \mathcal{E}_n] &\le \mathbb{P}[T_1 + T_2 < \tau + 2\sqrt{T_1 T_2} | \mathcal{E}_n] \\
&\le \mathbb{P}[\beta^2_{j_S} \widetilde{\sigma}^2_{j_s}(1-\iota) < \widetilde{\tau}] + \mathbb{P}[\mathcal{A}_1^C] + \mathbb{P}[\mathcal{A}_2^C] \\
&\le \frac{\delta}{2}.
\end{aligned} \tag{48}
$$

Combining (44)-(48), we have that $\mathbb{P}[\widehat{S}(\widehat{s}_n) = S] \ge 1 - 4\delta - 2\exp(-s/2)$, which completes the proof.

□

6.6 Proof of Theorem 3

We proceed to show that (11) holds with high probability under the assumptions of the theorem. We start with the case when $\Phi(\cdot) = ||\cdot||_2$. Let $\sigma_n^2 = \sigma^2/n$ and $\nu_j = \sigma_n^{-2} \sum_{k \in [T]} (\Sigma_{jS_k} \beta_{kS_k})^2$. With this notation, it is easy to observe that $\Phi^2(\{\widehat{\mu}_{kj}\}_k) \sim \sigma_n^2 \chi^2_T(\nu_j)$ where $\chi^2_T(\nu_j)$ is a non-central chi-squared random variable with T degrees of freedom and non-centrality parameter ν_j. From (27),

$$
\sigma_n^{-2} \max_{j \in S^C} \Phi^2(\{\widehat{\mu}_{kj}\}_k) \le T + 2\log \frac{2(p-s)}{\delta} + \max_{j \in S^C} \nu_j + 2\sqrt{(T + 2\nu_j) \log \frac{2(p-s)}{\delta}}
$$

with probability at least $1-\delta/2$. Similarly, from (28),

$$\sigma_n^{-2} \min_{j \in S} \Phi^2(\{\widehat{\mu}_{kj}\}_k) \geq T + \min_{j \in S} \nu_j - \max_{j \in S} 2\sqrt{(T+2\nu_j)\log\frac{2s}{\delta}}$$

with probability at least $1-\delta/2$. Combining the last two displays we have shown that (12) is sufficient to show that $\mathbb{P}[\widehat{S}_{\ell_2}(s) = S] \geq 1-\delta$.

Next, we proceed with $\Phi(\cdot) = ||\cdot||_1$, which can be dealt with similarly as the previous case. Using (24) together with $||\mathbf{a}||_1 \leq \sqrt{p}||\mathbf{a}||_2$, $a \in \mathbb{R}^p$,

$$\max_{j \in S^C} \sum_{k \in [T]} |\widehat{\mu}_{kj}| \leq \max_{j \in S^C} \sum_{k \in [T]} |\mathbf{\Sigma}_{jS_k}\boldsymbol{\beta}_{kS_k}| + \sigma_n \sqrt{T^2 + 2T\sqrt{T\log\frac{2(p-s)}{\delta}} + 2T\log\frac{2(p-s)}{\delta}}$$

with probability at least $1-\delta/2$. Similarly,

$$\min_{j \in S} \sum_{k \in [T]} |\widehat{\mu}_{kj}| \geq \min_{j \in S} \sum_{k \in [T]} |\mathbf{\Sigma}_{jS_k}\boldsymbol{\beta}_{kS_k}| - \sigma_n \sqrt{T^2 + 2T\sqrt{T\log\frac{2s}{\delta}} + 2T\log\frac{2s}{\delta}}$$

with probability $1-\delta/2$. Combining the last two displays we have shown that (13) is sufficient to show that $\mathbb{P}[\widehat{S}_{\ell_1}(s) = S] \geq 1-\delta$.

We complete the proof with the case when $\Phi(\cdot) = ||\cdot||_\infty$. Using a standard normal tail bound together with union bound

$$\max_{j \in S^C} \Phi(\{\widehat{\mu}_{kj}\}_k) \leq \max_{j \in S^C} \max_{k \in [T]} |\mathbf{\Sigma}_{jS_k}\boldsymbol{\beta}_{kS_k}| + \sigma_n \sqrt{2\log\frac{2(p-s)T}{\delta}}$$

with probability $1-\delta/2$ and

$$\min_{j \in S} \Phi(\{\widehat{\mu}_{kj}\}_k) \geq \min_{j \in S} \max_{k \in [T]} |\mathbf{\Sigma}_{jS_k}\boldsymbol{\beta}_{kS_k}| - \sigma_n \sqrt{2\log\frac{2sT}{\delta}}$$

with probability $1-\delta/2$, where $\sigma_n^2 = \sigma^2/n$. This shows that (14) is sufficient to show that $\mathbb{P}[\widehat{S}_{\ell_\infty}(s) = S] \geq 1-\delta$.

6.7 Proof of Theorem 4

We proceed as in the proof of 2. Define the event

$$\mathcal{E}_n = \{\widehat{S}_\phi(s) = S\}.$$

Irrespective of which scoring function Φ is used, Theorem 1 provides the sufficient conditions under which $\mathbb{P}[\mathcal{E}_n^C] \leq \delta$. It remains to upper bound $\mathbb{P}[\widehat{s}_n \neq s|\mathcal{E}_n]$, since

$$\mathbb{P}[\widehat{s}_n \neq s] \leq \mathbb{P}[\widehat{s}_n \neq s|\mathcal{E}_n]\mathbb{P}[\mathcal{E}_n] + \mathbb{P}[\mathcal{E}_n^C]. \tag{49}$$

An upper bound on $\mathbb{P}[\widehat{s}_n \neq s|\mathcal{E}_n]$ is constructed by combining upper bounds on $\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n]$ and $\mathbb{P}[\widehat{s}_n < s|\mathcal{E}_n]$.

Let $\tau = (T+2\sqrt{T\log(2/\delta)}+2\log(2/\delta))\sigma^2$. From $\{\widehat{s}_n > s|\mathcal{E}_n\} \subseteq \cup_{k=s}^{p-1}\{\widehat{\xi}_{\ell_2,n}(k) \geq \tau|\mathcal{E}_n\}$ follows that

$$\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n] \leq \sum_{k=s}^{p-1}\mathbb{P}[\widehat{\xi}_{\ell_2,n}(k) \geq \tau|\mathcal{E}_n]. \tag{50}$$

For a fixed $s \leq k \leq p-1$, $\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k)$ is the projection matrix from $\mathbb{R}^n$ to $\widehat{V}_n(k+1) \cap \widehat{V}_n(k)^{\perp}$. Since we are estimating $\widehat{s}_n$ on the second half of the samples, the projection matrix $\widehat{\mathbf{H}}(k)$ is independent of $\boldsymbol{\epsilon}$ for all k. Now, exactly one of the two events $\{\widehat{V}_n(k) = \widehat{V}_n(k+1)\}$ and $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\}$ occur. On the event $\{\widehat{V}_n(k) = \widehat{V}_n(k+1)\}$, $\widehat{\xi}_{\ell_2,n}(k) = 0$. Next we analyze the event $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \cap \mathcal{E}_n$. Since $\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k)$ is a rank one projection matrix

$$\widehat{\xi}_{\ell_2,n}(k) = \sum_{t\in[T]} ||(\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k))\mathbf{y}_t||_2^2 = \sum_{t\in[T]} ||(\widehat{\mathbf{H}}(k+1) - \widehat{\mathbf{H}}(k))\boldsymbol{\epsilon}_t||_2^2 \overset{d}{=} \sigma^2\chi_T^2.$$

Furthermore, $\widehat{\xi}_{\ell_2,n}(k) \perp\!\!\!\perp \widehat{\xi}_{\ell_2,n}(k')$, $k \neq k'$. It follows that for any realization of the sequences $\widehat{V}_n(1), \ldots, \widehat{V}_n(p)$,

$$\begin{aligned}
&\sum_{k=s}^{p-1}\mathbb{P}[\widehat{\xi}_{\ell_2,n}(k) \geq \tau|\mathcal{E}_n] \\
&\quad= \sum_{k=s}^{p-1}\mathbb{P}[\widehat{\xi}_{\ell_2,n}(k) \geq \tau|\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \cap \mathcal{E}_n]\mathbb{P}[\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)] \\
&\quad= \mathbb{P}[\sigma^2\chi_T^2 \geq \tau]\mathbb{E}\sum_{k=s}^{p-1}\mathbb{I}\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\} \\
&\quad\leq n\mathbb{P}[\sigma^2\chi_T^2 \geq \tau],
\end{aligned}$$

where the first equality follows since $\{\widehat{V}_n(k) \subsetneq \widehat{V}_n(k+1)\}$ is independent of $\mathcal{E}_n$. Combining with (50) gives

$$\mathbb{P}[\widehat{s}_n > s|\mathcal{E}_n] \leq n\mathbb{P}[\sigma^2\chi_T^2 \geq \tau] \leq \delta/2 \tag{51}$$

using (24).

Next, we focus on bounding $\mathbb{P}[\widehat{s}_n < s|\mathcal{E}_n]$. Since $\{\widehat{s}_n < s|\mathcal{E}_n\} \subset \{\widehat{\xi}_{\ell_2,n}(s-1) < \tau|\mathcal{E}_n\}$, it is sufficient to bound $\mathbb{P}[\widehat{\xi}_{\ell_2,n}(s-1) < \tau|\mathcal{E}_n]$. Using the definition of $\widehat{\mathbf{H}}(s)$ it is straightforward to obtain that

$$(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\mathbf{y}_t = (\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))(\mathbf{X}_{j_s}\beta_{tj_s} + \boldsymbol{\epsilon}_t).$$

Write $\mathbf{X}_{j_s} = \mathbf{X}_{j_s}^{(1)} + \mathbf{X}_{j_s}^{(2)}$ where $\mathbf{X}_{j_s}^{(1)} \in \widehat{V}_n(s-1)$ and $\mathbf{X}_{j_s}^{(2)} \in \widehat{V}_n(s) \cap \widehat{V}_n(s-1)^{\perp}$. Then

$$(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))\mathbf{y}_t = (\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))(\mathbf{X}_{j_s}^{(2)}\beta_{tj_s} + \boldsymbol{\epsilon}_t).$$

Furthermore we have that

$$||(\widehat{\mathbf{H}}(s) - \widehat{\mathbf{H}}(s-1))(\mathbf{X}_{j_s}^{(2)}\beta_{tj_s} + \boldsymbol{\epsilon}_t)||_2^2 = (||\mathbf{X}_{j_s}^{(2)}\beta_{tj_s}||_2 + Z_t)^2$$

where $Z_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. It follows that $\widehat{\xi}_{\ell_2,n}(s-1) \sim \sigma^2\chi_T^2(\nu)$ with $\nu = \sigma^{-2}\sum_{t\in[T]} ||\mathbf{X}_{j_s}^{(2)}\beta_{tj_s}||_2^2$. It is left to show that

$$\mathbb{P}[\sigma^2\chi_T^2(\nu) < \tau] \leq \delta/2. \tag{52}$$

Using (28) and following the proof of Theorem 2 in [2], we have that (52) holds if

$$\nu > 2\sqrt{5}\log^{1/2}\left(\frac{4}{\delta^2}\right)\sqrt{T} + 8\log\left(\frac{4}{\delta^2}\right).$$

Under the assumptions, we have that

$$\min_{j\in S}\sum_{t\in[T]} ||\mathbf{X}_j^{(2)}\beta_{tj}||_2^2 > \left[2\sqrt{5}\log^{1/2}\left(\frac{4}{\delta^2}\right)\sqrt{T} + 8\log\left(\frac{4}{\delta^2}\right)\right]\sigma^2$$

which shows (52). Combining (51) and (52), we obtain (49) which completes the proof.

6.8 Proof of Theorem 5

We have

$$H_p(\widehat{S}, S \,|\, \mathbf{X}) \geq \sum_{j=1}^{p}\left[\mathbb{P}\left(\|\beta_{\cdot j}\|_2 = 0, \|\widehat{\beta}_{\cdot j}\|_2 \neq 0\right) + \mathbb{P}\left(\|\beta_{\cdot j}\|_2 \neq 0, \|\widehat{\beta}_{\cdot j}\|_2 = 0\right)\right]. \tag{53}$$

For $1 \leq j \leq p$, we consider the hypothesis testing:

$$H_{0,j}: \quad \|\beta_{\cdot j}\|_2 = 0 \;\text{ vs. }\; \|\beta_{\cdot j}\|_2 \neq 0. \tag{54}$$

For $1 \leq t \leq T$, we denote by $\boldsymbol{\beta}_t$ any empirical realization of the coefficient vector. Let $\widetilde{\boldsymbol{\beta}}_t := \boldsymbol{\beta}_t - \beta_{tj}e_j$ where e_j is the j-th canonical basis of $\mathbb{R}^p$. We define $h(\mathbf{y}; \widetilde{\beta}, \boldsymbol{\alpha}) := h(\mathbf{y}_1, \ldots, \mathbf{y}_T; \widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_T, \alpha_1, \ldots, \alpha_T)$ be the joint distribution of

$$\mathbf{y}_1, \ldots, \mathbf{y}_T \sim \prod_{t=1}^{T}\mathcal{N}\left(\mathbf{X}\left(\widetilde{\boldsymbol{\beta}}_t + \alpha_t e_j\right), \mathbf{I}_n\right). \tag{55}$$

We then have

$$h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}) = h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, 0) \cdot \exp\left(\sum_{t=1}^{T}\alpha_t x_j'(\mathbf{y}_t - \mathbf{X}\widetilde{\boldsymbol{\beta}}_t) - \sum_{t=1}^{T}\frac{\alpha_t^2}{2}\right). \tag{56}$$

Let $\max_{1\leq t\leq T}|\alpha_t| \leq \tau_p$. We define

$$h(\mathbf{y}; \widetilde{\beta}, \tau_p) = h(\mathbf{y}; \widetilde{\beta}, 0) \cdot \exp\left(\tau_p\sum_{t=1}^{T}x_j'(\mathbf{y}_t - \mathbf{X}\widetilde{\boldsymbol{\beta}}_t) - \frac{T\tau_p^2}{2}\right). \tag{57}$$

Let $G(\widetilde{\boldsymbol{\beta}})$ be the joint distribution of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T$. Using Neyman-Pearson Lemma, Fubinni's Theorem and some basic calculus, we have

$$\mathbb{P}\left(\|\beta_{\cdot j}\|_2 = 0, \|\widehat{\beta}_{\cdot j}\|_2 \neq 0\right) + \mathbb{P}\left(\|\beta_{\cdot j}\|_2 \neq 0, \|\widehat{\beta}_{\cdot j}\|_2 = 0\right) \tag{58}$$

$$\geq \frac{1}{2} - \frac{1}{2}\int \left[\int \left|(1-\eta_p)h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, 0) - \eta_p h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha})\right| d\mathbf{y}\right] d\pi_p(\boldsymbol{\alpha}) dG(\widetilde{\boldsymbol{\beta}}) \tag{59}$$

$$= \frac{1}{2} - \frac{1}{2}\int H(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}) d\pi_p(\boldsymbol{\alpha}) dG(\widetilde{\boldsymbol{\beta}}), \tag{60}$$

where

$$H(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}) \equiv \int \left|(1-\eta_p)h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, 0) - \eta_p h(\mathbf{y}; \widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha})\right| d\mathbf{y}. \tag{61}$$

It can be seen that

$$H(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}) \leq H(\widetilde{\boldsymbol{\beta}}, \tau_p). \tag{62}$$

We then have

$$\mathbb{P}\left(\|\beta_{\cdot j}\|_2 = 0, \|\widehat{\beta}_{\cdot j}\|_2 \neq 0\right) + \mathbb{P}\left(\|\beta_{\cdot j}\|_2 \neq 0, \|\widehat{\beta}_{\cdot j}\|_2 = 0\right) \geq \frac{1}{2} - \frac{1}{2}\int H(\widetilde{\boldsymbol{\beta}}, \tau_p) dG(\widetilde{\boldsymbol{\beta}}). \tag{63}$$

For any realization of $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_p$, we define

$$D_p(\widetilde{\boldsymbol{\beta}}) := \left\{\mathbf{y}_1, \ldots, \mathbf{y}_T : \eta_p \cdot \exp\left(\tau_p \sum_{t=1}^{T} x_j'(\mathbf{y}_t - \mathbf{X}\widetilde{\boldsymbol{\beta}}_t) - \frac{T\tau_p^2}{2}\right) > (1-\eta_p)\right\}. \tag{64}$$

We know that $\mathbf{y}_1, \ldots, \mathbf{y}_T \in D_p(\widetilde{\boldsymbol{\beta}})$ if and only if

$$W_j = \sum_{t=1}^{T} x_j^{'}(\mathbf{y}_t - \mathbf{X}\widetilde{\boldsymbol{\beta}}_t) > \lambda_p. \tag{65}$$

It is then easy to see that

$$W_j \sim \mathcal{N}(0, T) \text{ under } H_{0,j} \tag{66}$$

$$W_j \sim \mathcal{N}(T\tau_p, T) \text{ under } H_{1,j}. \tag{67}$$

Following exactly the same argument as in Lemma 6.1 from Ji and Jin (2011), we obtain the lower bound:

$$\frac{1}{2} - \frac{1}{2}H(\widetilde{\boldsymbol{\beta}}, \tau_p) \geq (1-\eta_p)\bar{F}\left(\frac{\lambda_p}{\sqrt{T}}\right) + \eta_p F\left(\frac{\lambda_p}{\sqrt{T}} - \sqrt{T}\tau_p\right).$$

Thus we finish the proof of the main argument (21).

To obtain more detailed rate, we have

$$\frac{1}{\eta_p} - 1 = p^v - 1. \tag{68}$$

Also,

$$\bar{\Phi}\left(\frac{\lambda_p}{\sqrt{T}}\right) \geq \frac{\sqrt{T}}{2\lambda_p}\phi\left(\frac{\lambda_p}{\sqrt{T}}\right) \tag{69}$$

$$\geq \frac{\sqrt{rT}}{(v+Tr)\sqrt{2\log p}} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(v+Tr)^2\log p}{4rT}\right) \tag{70}$$

$$= \frac{\sqrt{rT}}{2(v+Tr)\sqrt{\pi \log p}} \cdot p^{-(v+Tr)^2/(4rT)}. \tag{71}$$

Therefore

$$\frac{1-\eta_p}{\eta_p}\bar{F}\left(\frac{\lambda_p}{\sqrt{T}}\right) \asymp \frac{\sqrt{rT}}{2(v+Tr)\sqrt{\pi\log p}} \cdot p^{v-(v+Tr)^2/(4rT)} \tag{72}$$

$$= \frac{\sqrt{rT}}{2(v+Tr)\sqrt{\pi\log p}} \cdot p^{-(v-Tr)^2/(4rT)}. \tag{73}$$

We then evaluate the second term

$$F\left(\frac{\lambda_p}{\sqrt{T}} - \sqrt{T}\tau_p\right) = \bar{F}\left(\sqrt{T}\tau_p - \frac{\lambda_p}{\sqrt{T}}\right). \tag{74}$$

First, we have that

$$\frac{\lambda_p}{\sqrt{T}} - \sqrt{T}\tau_p = \frac{(v+Tr)\sqrt{\log p}}{\sqrt{2Tr}} - \sqrt{2rT\log p}. \tag{75}$$

If $v > Tr$, we have

$$\frac{\lambda_p}{\sqrt{T}} - \sqrt{T}\tau_p \to \infty,$$

which implies that

$$F\left(\frac{\lambda_p}{\sqrt{T}} - \sqrt{T}\tau_p\right) \geq 1 + o(1). \tag{76}$$

Now, we consider the case that $v < Tr$,

$$\bar{F}\left(\sqrt{T}\tau_p - \frac{\lambda_p}{\sqrt{T}}\right) = \bar{F}\left(\frac{(Tr-v)\sqrt{\log p}}{\sqrt{2Tr}}\right) \tag{77}$$

$$\geq \frac{\sqrt{2Tr}}{(Tr-v)\sqrt{\log p}}\frac{1}{\sqrt{2\pi}} \cdot p^{-(v-Tr)^2/(4rT)} \tag{78}$$

$$= \frac{\sqrt{Tr}}{(Tr-v)\sqrt{\pi\log p}} \cdot p^{-(v-Tr)^2/(4rT)} \tag{79}$$

This finishes the whole proof.

6.9 Extended empirical results

6.9.1 Extended results for Simulation 1

Simulation 1: $(n, p, s, T) = (500, 20000, 18, 500)$, $T_{\text{non-zero}} = 500$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 15	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	91.0	18.1
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	92.0	18.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	92.0	18.1
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	63.0	18.5
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	68.0	18.4
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	68.0	18.4
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	87.0	18.1
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	87.0	18.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	88.0	18.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	99.9	0.0	0.0
	$\widehat{S}_{\ell_1}$	0.0	100.0	100.0	0.0	0.0
	$\widehat{S}_{\ell_2}$	0.0	100.0	99.9	0.0	0.0

Simulation 1: $(n, p, s, T) = (500, 20000, 18, 500)$, $T_{\text{non-zero}} = 300$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 15	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	18.0
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	99.0	18.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	18.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	76.0	18.3
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	91.0	18.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	92.0	18.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	98.1	0.0	0.3
	$\widehat{S}_{\ell_1}$	0.0	100.0	98.4	0.0	0.3
	$\widehat{S}_{\ell_2}$	0.0	100.0	98.2	0.0	0.3

Simulation 1: $(n, p, s, T) = (500, 20000, 18, 500)$, $T_{\text{non-zero}} = 100$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 15	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	18.0
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	18.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	18.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	60.0	18.5
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	96.0	18.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	97.0	18.0

6.9.2 Extended results for Simulation 2

Simulation 2.a: $(n, p, s, T) = (200, 5000, 10, 500)$, $T_{\text{non-zero}} = 400$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	83.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	88.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	93.0	10.1
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	82.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	91.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	91.0	10.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	98.4	0.0	0.2
	$\widehat{S}_{\ell_1}$	0.0	100.0	98.3	0.0	0.2
	$\widehat{S}_{\ell_2}$	0.0	100.0	98.2	0.0	0.2

Simulation 2.a: $(n, p, s, T) = (200, 5000, 10, 500)$, $T_{\text{non-zero}} = 250$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	86.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	98.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	97.0	10.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	1.0	100.0	41.0	1.0	5.9
	$\widehat{S}_{\ell_1}$	4.0	100.0	41.6	4.0	5.8
	$\widehat{S}_{\ell_2}$	2.0	100.0	41.9	2.0	5.8

Simulation 2.a: $(n, p, s, T) = (200, 5000, 10, 500)$, $T_{\text{non-zero}} = 100$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	77.0	10.3
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	97.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	96.0	10.0

Simulation 2.b: $(n, p, s, T) = (200, 5000, 10, 750)$, $T_{\text{non-zero}} = 600$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	88.0	10.1
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	87.0	10.2
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	89.0	10.1
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	72.0	10.3
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	89.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	90.0	10.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	98.1	0.0	0.2
	$\widehat{S}_{\ell_1}$	0.0	100.0	97.9	0.0	0.2
	$\widehat{S}_{\ell_2}$	0.0	100.0	98.1	0.0	0.2

Simulation 2.b: $(n, p, s, T) = (200, 5000, 10, 750)$, $T_{\text{non-zero}} = 375$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	91.0	10.1
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	94.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	95.0	10.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	9.0	100.0	28.6	9.0	7.1
	$\widehat{S}_{\ell_1}$	12.0	100.0	27.4	12.0	7.3
	$\widehat{S}_{\ell_2}$	9.0	100.0	28.4	9.0	7.2

Simulation 2.b: $(n, p, s, T) = (200, 5000, 10, 750)$, $T_{\text{non-zero}} = 150$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	77.0	10.3
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	98.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	96.0	10.0

Simulation 2.c: $(n, p, s, T) = (200, 5000, 10, 1000)$, $T_{\text{non-zero}} = 800$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	82.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	89.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	85.0	10.2
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	76.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	83.0	10.2
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	83.0	10.2
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	97.6	0.0	0.2
	$\widehat{S}_{\ell_1}$	0.0	100.0	97.4	0.0	0.3
	$\widehat{S}_{\ell_2}$	0.0	100.0	97.5	0.0	0.2

Simulation 2.c: $(n, p, s, T) = (200, 5000, 10, 1000)$, $T_{\text{non-zero}} = 500$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	85.0	10.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	94.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	93.0	10.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	14.0	100.0	21.2	14.0	7.9
	$\widehat{S}_{\ell_1}$	15.0	100.0	20.9	15.0	7.9
	$\widehat{S}_{\ell_2}$	16.0	100.0	20.1	16.0	8.0

Simulation 2.c: $(n, p, s, T) = (200, 5000, 10, 1000)$, $T_{\text{non-zero}} = 200$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	10.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	10.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	79.0	10.3
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	94.0	10.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	93.0	10.1

6.9.3 Extended results for Simulation 3

Simulation 3: $(n, p, s, T) = (100, 5000, 3, 150)$, $T_{\text{non-zero}} = 80$, $\rho = 0.2$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	3.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	3.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	97.0	3.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	99.0	3.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	99.0	3.0

Simulation 3: $(n, p, s, T) = (100, 5000, 3, 150)$, $T_{\text{non-zero}} = 80$, $\rho = 0.5$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	3.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	3.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	3.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	79.0	3.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	79.0	3.2
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	72.0	3.3

Simulation 3: $(n, p, s, T) = (100, 5000, 3, 150)$, $T_{\text{non-zero}} = 80$, $\rho = 0.7$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	97.0	100.0	1.0	97.0	3.0
	$\widehat{S}_{\ell_1}$	99.0	100.0	0.3	99.0	3.0
	$\widehat{S}_{\ell_2}$	99.0	100.0	0.3	99.0	3.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	96.0	100.0	1.3	95.0	3.0
	$\widehat{S}_{\ell_1}$	99.0	100.0	0.3	97.0	3.0
	$\widehat{S}_{\ell_2}$	97.0	100.0	1.0	95.0	3.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	94.0	100.0	2.0	67.0	3.3
	$\widehat{S}_{\ell_1}$	98.0	100.0	0.7	71.0	3.3
	$\widehat{S}_{\ell_2}$	94.0	100.0	2.0	63.0	3.3

6.9.4 Extended results for Simulation 4

Simulation 4: $(n, p, s, T) = (150, 4000, 8, 150)$, $T_{\text{non-zero}} = 80$, $\rho = 0.2$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	8.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	8.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	8.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	95.0	8.1
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	8.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	8.0
SNR = 1	$\widehat{S}_{\ell_\infty}$	0.0	100.0	77.4	0.0	1.8
	$\widehat{S}_{\ell_1}$	0.0	100.0	77.6	0.0	1.8
	$\widehat{S}_{\ell_2}$	0.0	100.0	78.0	0.0	1.8

Simulation 4: $(n, p, s, T) = (150, 4000, 8, 150)$, $T_{\text{non-zero}} = 80$, $\rho = 0.5$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
SNR = 10	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	8.0
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	100.0	8.0
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	100.0	8.0
SNR = 5	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	84.0	8.2
	$\widehat{S}_{\ell_1}$	100.0	100.0	0.0	87.0	8.1
	$\widehat{S}_{\ell_2}$	100.0	100.0	0.0	87.0	8.1
SNR = 1	$\widehat{S}_{\ell_\infty}$	1.0	100.0	56.2	1.0	3.5
	$\widehat{S}_{\ell_1}$	0.0	100.0	57.2	0.0	3.4
	$\widehat{S}_{\ell_2}$	0.0	100.0	57.0	0.0	3.4

6.9.5 Extended results for Simulation 5

Simulation 5: $(n, p, s, T) = (200, 10000, 5, 500)$, $T_{\text{non-zero}} = 400$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
$\sigma = 1.5$	$\widehat{S}_{\ell_\infty}$	0.0	100.0	20.0	0.0	8.6
	$\widehat{S}_{\ell_1}$	0.0	99.9	83.8	0.0	11.4
	$\widehat{S}_{\ell_2}$	0.0	99.9	74.2	0.0	10.9
$\sigma = 2.5$	$\widehat{S}_{\ell_\infty}$	0.0	99.9	20.0	0.0	13.4
	$\widehat{S}_{\ell_1}$	0.0	99.8	83.6	0.0	16.8
	$\widehat{S}_{\ell_2}$	0.0	99.8	74.4	0.0	16.9
$\sigma = 4.5$	$\widehat{S}_{\ell_\infty}$	0.0	99.7	32.6	0.0	35.6
	$\widehat{S}_{\ell_1}$	0.0	99.7	83.2	0.0	29.3
	$\widehat{S}_{\ell_2}$	0.0	99.7	74.6	0.0	29.7

Simulation 5: $(n, p, s, T) = (200, 10000, 5, 500)$, $T_{\text{non-zero}} = 250$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
$\sigma = 1.5$	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	99.0	5.0
	$\widehat{S}_{\ell_1}$	0.0	99.9	92.0	0.0	10.7
	$\widehat{S}_{\ell_2}$	0.0	99.9	54.4	0.0	8.7
$\sigma = 2.5$	$\widehat{S}_{\ell_\infty}$	87.0	100.0	2.6	39.0	5.9
	$\widehat{S}_{\ell_1}$	0.0	99.9	90.6	0.0	14.8
	$\widehat{S}_{\ell_2}$	0.0	99.9	55.0	0.0	12.5
$\sigma = 4.5$	$\widehat{S}_{\ell_\infty}$	0.0	99.9	20.2	0.0	16.2
	$\widehat{S}_{\ell_1}$	0.0	99.8	86.4	0.0	22.2
	$\widehat{S}_{\ell_2}$	0.0	99.8	56.2	0.0	19.9

Simulation 5: $(n, p, s, T) = (200, 10000, 5, 500)$, $T_{\text{non-zero}} = 100$

	$\widehat{S}$	Prob. (%) of $S \subseteq \widehat{S}$	Fraction (%) of Correct zeros	Fraction (%) of Incorrect zeros	Fraction (%) of $S = \widehat{S}$	$\lvert\widehat{S}\rvert$
$\sigma = 1.5$	$\widehat{S}_{\ell_\infty}$	100.0	100.0	0.0	100.0	5.0
	$\widehat{S}_{\ell_1}$	0.0	99.9	95.8	0.0	6.7
	$\widehat{S}_{\ell_2}$	9.0	100.0	18.2	9.0	4.4
$\sigma = 2.5$	$\widehat{S}_{\ell_\infty}$	99.0	100.0	0.2	91.0	5.1
	$\widehat{S}_{\ell_1}$	0.0	99.9	93.0	0.0	7.7
	$\widehat{S}_{\ell_2}$	0.0	100.0	21.6	0.0	4.5
$\sigma = 4.5$	$\widehat{S}_{\ell_\infty}$	9.0	100.0	18.2	4.0	5.1
	$\widehat{S}_{\ell_1}$	0.0	99.9	85.2	0.0	8.0
	$\widehat{S}_{\ell_2}$	0.0	100.0	29.6	0.0	5.3

References

[1] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3):243–272, 2008.

[2] Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[3] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Annals of Statistics*, 36(1):199–227, 2008.

[4] L. Birgé. An alternative point of view on Lepski's method. *Lecture Notes-Monograph Series*, 36:113–133, 2001.

[5] T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *Signal Processing*, 58(3):1300–1308, 2010.

[6] T.T. Cai, J. Jin, and M.G. Low. Estimation and confidence sets for sparse normal mixtures. *The Annals of Statistics*, 35(6):2421–2449, 2007.

[7] E. Candes and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6):2313–2351, 2007.

[8] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[9] K.R. Davidson and S.J. Szarek. Local operator theory, random matrices and Banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

[10] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32(3):962–994, 2004.

[11] D.L. Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[12] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *PNAS*, 100(5):2197, 2003.

[13] L. El Ghaoui, V. Viallon, and T. Rabbani. Safe feature elimination in sparse supervised learning. Technical Report UC/EECS-2010-126, EECS Dept., University of California at Berkeley, September 2010.

[14] J. Fan, Y. Feng, and R. Song. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *JASA*, 106(495):544–557, 2011.

[15] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *JASA*, 96:1348–1360, 2001.

[16] J. Fan and J. Lv. Sure independence screening for ultrahigh dimensional feature space. *JRSS: B*, 70(5):849–911, 2008.

[17] J. Fan, R. Samworth, and Y. Wu. Ultrahigh dimensional feature selection: beyond the linear model. *JMLR*, 10:2013–2038, 2009.

[18] J.J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, 2005.

[19] C. Genovese, J. Jin, and L. Wasserman. Revisiting marginal regression. *arXiv:0911.4080*, 2009.

[20] P. Ji and J. Jin. UPS Delivers Optimal Phase Diagram in High Dimensional Variable Selection. *ArXiv e-prints*, October 2010.

[21] I.M. Johnstone. Chi-square oracle inequalities. *Lecture Notes-Monograph Series*, 36:399–418, 2001.

[22] S. Kim and E. P. Xing. Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet*, 5(8):e1000587, 2009.

[23] M. Kolar, J. Lafferty, and L. Wasserman. Union support recovery in multi-task learning. *J. Mach. Learn. Res.*, 12:2415–2435, July 2011.

[24] M. Kolar and E. P. Xing. Ultra-high dimensional multiple output learning with simultaneous orthogonal matching pursuit: Screening approach. In *AISTATS*, pages 413–420, 2010.

[25] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5):1302–1338, 2000.

[26] S. L. Lauritzen. *Graphical Models*. Oxford University Press, USA, July 1996.

[27] H. Liu, M. Palatucci, and J. Zhang. Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In *ICML*, pages 649–656, New York, NY, USA, 2009. ACM.

[28] K. Lounici, M. Pontil, A. B. Tsybakov, and S. van de Geer. Taking advantage of sparsity in Multi-Task learning. In *COLT*, 2009.

[29] K. Lounici, M. Pontil, A. B. Tsybakov, and S. van de Geer. Oracle inequalities and optimal inference under group sparsity. *arXiv 1007.1771*, 2010.

[30] N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics*, 37(1):246–270, 2009.

[31] S. Negahban and M. Wainwright. Phase transitions for high-dimensional joint support recovery. *NIPS*, pages 1161–1168, 2009.

[32] G. Obozinski, M.J. Wainwright, and M.I. Jordan. Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*, 39(1):1–47, 2011.

[33] J. M. Robins, R. Scheines, P. Spirtes, and L. Wasserman. Uniform consistency in causal inference. *Biometrika*, 90(3):491–515, 2003.

[34] P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2000.

[35] R. Tibshirani. Regression shrinkage and selection via the lasso. *JRSS: B*, 58:267–288, 1996.

[36] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R.J. Tibshirani. Strong rules for discarding predictors in lasso-type problems. *Arxiv preprint arXiv:1011.2234*, 2010.

[37] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[38] B.A. Turlach, W.N. Venables, and S.J. Wright. Simultaneous variable selection. *Technometrics*, 47(3):349–363, 2005.

[39] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.

[40] H. Wang. Forward regression for ultra-high dimensional variable screening. *JASA*, 104(488):1512–1524, 2009.

[41] L. Wasserman and K. Roeder. High dimensional variable selection. *Annals of statistics*, 37(5A):2178, 2009.

[42] C.H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38(2):894–942, 2010.

[43] J. Zhang. *A probabilistic framework for multitask learning (Technical Report CMU-LTI-06-006).* PhD thesis, Carnegie Mellon University, 2006.

[44] P. Zhao and B. Yu. On model selection consistency of lasso. *J. Mach. Learn. Res.*, 7:2541–2563, 2006.

[45] H. Zou. The adaptive lasso and its oracle properties. *JASA*, 101:1418–1429, 2006.

[46] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *JRSS: B*, 67(2):301–320, 2005.

[47] H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, 36(4):1509–1533, 2008.

[48] H. Zou and M. Yuan. The $\mathtt{F}_\infty$-norm support vector machine. *Stat. Sin*, 18:379–398, 2008.