1. In this question we will outline an algorithm to compute the graphical Lasso.

   (a) Let
$$Q(\Omega) = -\log\det(\Omega) + \mathrm{tr}(S\Omega) + \lambda\|\Omega\|_1$$
be the graphical Lasso objective with $\hat{\Omega} = \operatorname*{argmin}_{\Omega \succ 0} Q(\Omega)$ assumed unique. Consider the following version of the graphical Lasso objective:
$$\min_{\Omega,\Theta\succ 0}\{-\log\det(\Omega) + \mathrm{tr}(S\Omega) + \lambda\|\Theta\|_1\}$$
subject to $\Omega = \Theta$. By introducing the Lagrangian for this objective, show that
$$p + \max_{U:S+U\succ 0,\, \|U\|_\infty \le \lambda} \log\det(S+U) \le Q(\hat{\Omega}).$$
Here $\|U\|_\infty = \max_{j,k}|U_{jk}|$ and p is the number of columns in the underlying data matrix X. *Hint: Write the additional term in the Lagrangian as* $\mathrm{tr}(U(\Omega - \Theta))$.
Solution: We know that for all symmetric $U \in \mathbb{R}^{p\times p}$,
$$\min_{\Omega,\Theta\succ 0}\{-\log\det(\Omega) + \mathrm{tr}(S\Omega) + \lambda\|\Theta\|_1 + \mathrm{tr}(U(\Omega-\Theta))\} \le Q(\hat{\Omega}^L).$$
Subdifferentiating the LHS, we see that for a minimiser of the LHS, Ω^*, Θ^* we have $S+U = \Omega^{*,-1}$ provided $S+U \succ 0$, and $U = \lambda\Gamma$ where $\|\Gamma\|_\infty \le 1$ and $\Gamma_{jk} = \mathrm{sgn}(\Theta^*_{jk})$ when $\Theta^*_{jk} \neq 0$. The latter implies that $\mathrm{tr}(\Theta^* U) = \lambda\|\Theta^*\|_1$. Thus we get that the LHS is
$$\log\det(S+U) + \mathrm{tr}((S+U)(S+U)^{-1}) = \log\det(S+U) + p,$$
and as the inequality is true for all U, we may take the maximum over U.

   (b) Suppose that U^* is the unique maximiser of the LHS. Show that $\hat{\Omega} = (S+U^*)^{-1}$.
Solution: The KKT conditions for the original objective Q tell us that
$$\hat{\Omega}^{-1} - S = \lambda\hat{\Gamma}$$
where $\|\hat{\Gamma}\|_\infty \le 1$ and $\Gamma_{jk} = \mathrm{sgn}(\Omega_{jk})$. Note that $\|\hat{\Omega}^{-1} - S\|_\infty \le \lambda$, so this is a feasible value of U. We will show that it is the optimal U. We see that
$$\mathrm{tr}(\hat{\Omega}(\hat{\Omega}^{-1} - S)) = \lambda\|\Omega\|_1$$
so
$$\begin{aligned} Q(\hat{\Omega}) &= \log\det(\hat{\Omega}^{-1}) + \mathrm{tr}(S\hat{\Omega}) + \mathrm{tr}(\hat{\Omega}(\hat{\Omega}^{-1} - S)) \\ &= \log\det(\hat{\Omega}^{-1}) + p. \end{aligned}$$
This show that taking $U = \hat{\Omega}^{-1} - S$ gives equality. So by uniqueness we must have $U^* = \hat{\Omega}^{-1} - S$.

(c) Now consider

$$\hat{\Sigma} = \underset{W: W \succ 0, \, \|W - S\|_\infty \leq \lambda}{\operatorname{argmin}} - \log\det(W). \tag{1}$$

Let $\hat{\Sigma}_{-j,j}$ be a block of $\hat{\Sigma}$ containing all but the jth row of $\hat{\Sigma}$ and only the jth column. Use the formula for the determinant in terms of Schur complements to show that $(\hat{\Sigma}_{jj}, \hat{\Sigma}_{-j,j}) = (\alpha^*, \beta^*)$, where (α^*, β^*) solve the following optimisation problem over (α, β):

$$\begin{aligned} &\text{minimise} && -\alpha + \beta^T \hat{\Sigma}_{-j,-j}^{-1} \beta, \\ &\text{such that} && \|\beta - S_{-j,j}\|_\infty \leq \lambda, \; |\alpha - S_{jj}| \leq \lambda. \end{aligned}$$

Conclude that $\alpha^* = S_{jj} + \lambda$. ($\beta^*$ can be found by standard quadratic programming techniques, or by converting the optimisation to a standard Lasso optimisation problem; thus we can perform block coordinate descent on the optimisation problem in (1), updating a row and corresponding column of W at each iteration.)
Solution: Follows from noting that

$$\log\det(W) = \log(W_{k,k,} - W_{k,-k} W_{-k,-k}^{-1} W_{-k,k}) + \log\det(W_{-k,-k}).$$

2. Consider the matrix

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1.01 \end{bmatrix}$$

and its perturbation

$$\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1.01 \end{bmatrix} + \begin{bmatrix} 0 & 0.01 \\ 0.01 & 0 \end{bmatrix}.$$

Show that the eigenvalues are stable to the perturbation, but the top eigenvector is not.
Solution: Q has eigenvalues 1.01 and 1 with eigenvectors $(0, 1)$ and $(1, 0)$, respectively. A numerical calculation shows that $\hat{Q}$ has eigenvalues approximately 1.017 and 0.994, which are close to those of Q, but the top eigenvector is $(0.53, 0.85)$ which is quite far from $(0, 1)$.

3. Prove the following inequality for two rank-r matrices S, U with orthonormal columns,

$$\min_{O \in \mathbb{R}^{d \times d} \text{ orthogonal}} \|S - UO\|_F \leq \sqrt{2} \sin\Theta(S, U).$$

Hint: Use $O = VW^T$ where $W\Sigma V^T$ is a singular value decomposition of $S^T U$.
Solution: With the O defined in the hint,

$$\begin{aligned} \|S - UO\|_F^2 = \operatorname{Tr}((S - UO)^T (S - UO)) &= \|S\|_F^2 + \|U\|_F^2 - 2\operatorname{Tr}(OS^T U) \\ &= 2r - 2\operatorname{Tr}(\Sigma), \end{aligned}$$

where r is the rank of S and U. On the other hand, by Pythagoras' theorem

$$\begin{aligned} \sin\Theta(S, U)^2 = \|\Pi_S (I - \Pi_U)\|_F^2 &= \|\Pi_S\|_F^2 - \|\Pi_S \Pi_U\|_F^2 \\ &= r - \|\Pi_S \Pi_U\|_F^2 \\ &= r - \|SS^T UU^T\|_F^2 \\ &= r - \operatorname{Tr}(\Sigma^2). \end{aligned}$$

We claim that the entries of Σ are bounded above by 1, such that $\operatorname{Tr}(\Sigma) \leq \operatorname{Tr}(\Sigma^2)$, then

$$\min_{O \in \mathbb{R}^{d \times d} \text{ orthogonal}} \|S - UO\|_F^2 \leq 2r - 2\operatorname{Tr}(\Sigma) \leq 2r - 2\operatorname{Tr}(\Sigma^2) = 2\sin\Theta(S, U)^2.$$

Taking the square root of both sides yields the desired inequality. To prove the claim, let $w = [S, S']$ and $\tilde{U} = [U, U']$ be orthogonal matrices. Then $S^T U$ is a diagonal block in $w^T \tilde{U}$. It follows that $\max_i \Sigma_{i,i} = \|S^T U\|_{op} \leq \|w^T \tilde{U}\|_{op} = 1$.

4. The SCoTTLASS estimator for Sparse PCA is obtained by solving

$$\underset{v \in S^{d-1}}{\text{maximise}}\; v^T \hat{\Sigma} v \quad \text{subject to } \|v\|_1 \leq \lambda.$$

Show that this is equivalent to the optimisation problem

$$\underset{\Theta \in \mathcal{S}_+^{d\times d}}{\text{maximise}}\; \text{Tr}(\Theta\hat{\Sigma}) \quad \text{subject to } \text{Tr}(\Theta) = 1, \sum_{i,j} |\Theta_{i,j}| \leq \lambda^2, \text{rank}(\Theta) = 1,$$

where $\mathcal{S}_+^{d\times d}$ is the cone of positive semidefinite matrices. Which of the constraints in this problem are convex? Dropping the rank constraint yields the problem

$$\underset{\Theta \in \mathcal{S}_+^{d\times d}}{\text{maximise}}\; \text{Tr}(\Theta\hat{\Sigma}) \quad \text{subject to } \text{Tr}(\Theta) = 1, \sum_{i,j} |\Theta_{i,j}| \leq R^2.$$

What happens when the maximum is achieved by a rank 1 matrix?
Solution: Identify $\Theta = vv^T$. Then, the two optimisation objectives are equal by the cyclic property of the trace. Furthermore, $\Theta = vv^T$ for some $v \in S^{d-1}$ with $\|v\|_1 \leq \lambda$ if and only if Θ is a rank 1 positive semidefinite matrix, $\text{Tr}(\Theta) = \|v\| = 1$, and $\sum_{i,j} |\Theta_{i,j}| = \sum_{i,j} |v_i v_j| = \sum_{i,j} |v_i||v_j| = \|v\|_1^2 \leq \lambda^2$. All the constraints in the problem are convex except for the rank constraint.

If the relaxed optimisation problem is solved by a rank 1 matrix, Θ^*, then this is also a solution of the previous problem, as Θ^* satisfies its more stringent constraints. Therefore, if $\Theta^* = v^* v^{*T}$, we can certify that v^* is a solution of the SCoTLASS problem.

5. Consider the spiked covariance model $\Sigma = \theta vv^T + I_d$ with a sparse spike $v \in S^{d-1}$, $\|v\|_0 = k$. Suppose that $k \leq d/2$ and $k \leq k'$ and define the Sparse PCA estimator

$$\hat{v} = \underset{v \in S^{d-1}, \|v\|_0 \leq k'}{\text{argmax}}\; v^T \hat{\Sigma} v.$$

In this problem you shall prove a non-asymptotic error bound for this estimator (Proposition 34 in the notes).

(a) Let $s \subseteq \{1, \ldots, d\}$ be the random subset of entries where either v or $\hat{v}$ is non-zero. Prove that

$$\sin \Theta(v, \hat{v})^2 \leq \frac{1}{\theta} \langle \hat{\Sigma}_s - \Sigma_s, \hat{v}_s \hat{v}_s^T - v_s v_s^T \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product, and subscripting by s indicates rows (and columns, for a matrix) are subsetted by s.
Solution: The right hand side is the same if we don't subscript by s, so it suffices to prove

$$\theta \sin \Theta(v, \hat{v})^2 \leq \langle \hat{\Sigma} - \Sigma, \hat{v}\hat{v}^T - vv^T \rangle.$$

Expanding the right hand side,

$$\begin{aligned}
\hat{v}^T \hat{\Sigma} \hat{v} - \hat{v}^T \Sigma \hat{v} - v \hat{\Sigma} v + v^T \Sigma v &\geq -\hat{v}^T \Sigma \hat{v} + v^T \Sigma v \\
&= -\theta(\hat{v}^T v)^2 - \hat{v}^T \hat{v} + \theta(v^T v)^2 + v^T v \\
&= \theta(1 - (\hat{v}^T v)^2) \\
&= \theta \sin \Theta(v, \hat{v})^2.
\end{aligned}$$

where the inequality is because by definition $\hat{v}^T\hat{\Sigma}\hat{v} \geq v^T\hat{\Sigma}v$, and the first identity follows from plugging in the value of Σ.

(b) Use Problem 3 and the matrix inequality $\langle A, B\rangle \leq \|A\|_{op}\|B\|_1$, where $\|B\|_1$ is the Schatten-1 norm, to show that

$$\min_{\epsilon\in\{-1,1\}} \|v - \epsilon\hat{v}\| \leq \frac{2\sqrt{2}}{\theta}\|\hat{\Sigma}_s - \Sigma_s\|_{op}.$$

Solution: The matrix inequality implies

$$\begin{aligned}
\langle \hat{\Sigma}_s - \Sigma_s, \hat{v}_s\hat{v}_s^T - v_s v_s^T\rangle &\leq \|\hat{\Sigma}_s - \Sigma_s\|_{op}\|\hat{v}_s\hat{v}_s^T - v_s v_s^T\|_1 \\
&\leq \sqrt{2}\|\hat{\Sigma}_s - \Sigma_s\|_{op}\|\hat{v}_s\hat{v}_s^T - v_s v_s^T\|_F \\
&\leq \sqrt{2}\|\hat{\Sigma}_s - \Sigma_s\|_{op}\|\hat{v}\hat{v}^T - vv^T\|_F \\
&= 2\|\hat{\Sigma}_s - \Sigma_s\|_{op}\left(1 - (v^T\hat{v})^2\right)^{1/2},
\end{aligned}$$

where the second inequality is due to the Cauchy-Schwarz inequality and the fact that $v_s\hat{v}_s^T - v_s v_s^T$ is a rank-2 matrix. Plugging this into the result of part (a) yields

$$\begin{aligned}
\sin\Theta(v,\hat{v})^2 &\leq \frac{2}{\theta}\|\hat{\Sigma}_s - \Sigma_s\|_{op}\sin\Theta(v,\hat{v}) \\
\implies \sin\Theta(v,\hat{v}) &\leq \frac{2}{\theta}\|\hat{\Sigma}_s - \Sigma_s\|_{op} \\
\implies \min_{\epsilon\in\{-1,1\}} \|v - \epsilon\hat{v}\| &\leq \frac{2\sqrt{2}}{\theta}\|\hat{\Sigma}_s - \Sigma_s\|_{op}
\end{aligned}$$

where the last implication follows from Problem 3.

(c) Finally, apply a union bound over all subsets $w \subseteq \{1,\ldots,d\}$ of size $k + k'$ in conjunction with Theorem 29 in the notes, to show that for some constant c,

$$\mathbb{P}\left(\min_{\epsilon\in\{-1,1\}} \|v - \epsilon\hat{v}\| \geq c\frac{1+\theta}{\theta}\left[\sqrt{\eta_n} \vee \eta_n\right]\right) \leq e^{-\delta} \qquad \text{for all } \delta > 0,$$

with

$$\eta_n = \frac{(k+k')\log(de^2/(k+k')) + \delta}{n}.$$

Solution: As the subset s is random, and so is Σ_s, we cannot apply a simple deviation bound for Σ_s. Instead, we note that

$$\min_{\epsilon\in\{-1,1\}} \|v - \epsilon\hat{v}\| \leq \max_{w\subseteq\{1,\ldots,d\},|w|=k+k'} \frac{2}{\theta}\|\hat{\Sigma}_w - \Sigma_w\|_{op}. \tag{2}$$

For any fixed $w \subseteq \{1,\ldots,d\}, |w| = k + k'$, Theorem 29 tells us that

$$\mathbb{P}\left(\frac{\|\hat{\Sigma}_w - \Sigma_w\|_{op}}{\|\Sigma_w\|_{op}} \geq C'\left[\left(\frac{k+k'}{n} + \sqrt{\frac{k+k'}{n}}\right) + \left(\frac{t}{n} + \sqrt{\frac{t}{n}}\right)\right]\right) \leq e^{-t} \text{ for all } t > 0.$$

A union bound over all deterministic subsets gives us for all $t > 0$,

$$\mathbb{P}\left(\max_{w\subseteq\{1,\ldots,d\},|w|=k+k'} \frac{\|\hat{\Sigma}_w - \Sigma_w\|_{op}}{\|\Sigma_w\|_{op}} \geq C'\left[\left(\frac{k+k'}{n} + \sqrt{\frac{k+k'}{n}}\right) + \left(\frac{t}{n} + \sqrt{\frac{t}{n}}\right)\right]\right)$$
$$\leq \binom{d}{k+k'}e^{-\delta}.$$

Hence, by (2), and the fact that $\binom{a}{b} \leq (ae/b)^b$,

$$\mathbb{P}\left(\min_{\epsilon\in\{-1,1\}} \|v - \epsilon\hat{v}\| \geq C'\left[\left(\frac{k+k'}{n} + \sqrt{\frac{k+k'}{n}}\right) + \left(\frac{t}{n} + \sqrt{\frac{t}{n}}\right)\right]\right)$$
$$\leq e^{-t+(k+k')[\log(d)+1-\log(k+k')]}.$$

Finally, setting the exponent on the right hand side to δ and solving for t yields the desired inequality for C large enough.

6. Show that if all null hypotheses are true, then the FDR is equivalent to the FWER.
Solution: In this case we have

$$\mathrm{FDP} = \frac{N_{01}}{\max(R,1)} = \mathbb{1}_{\{R\geq 1\}}.$$

7. Show that the definition of Holm's procedure as the closed testing procedure with the local tests as the Bonferroni test is equivalent to the the step-down procedure definition.
Hint: It may help to first show that with the definition as a closed testing procedure, we reject $H_{(i)}$ when

$$\min_{s\in\{1,\ldots,m\}} \mathbb{1}_{\{\min_{j\in J_s} p_j \leq \alpha/s\}} = 1,$$

where

$$J_s \in \operatorname*{argmin}_{J:J\supseteq\{(i)\},\,|J|=s} \mathbb{1}_{\{\min_{j\in J} p_j \leq \alpha/s\}}.$$

Solution: We reject $H_{(i)}$ if

$$\min_{J\supseteq\{(i)\}} \mathbb{1}_{\{\min_{j\in J} p_j \leq \alpha/|J|\}} = \min_{s\in\{1,\ldots,m\}} \mathbb{1}_{\{\min_{j\in J_s} p_j \leq \alpha/s\}} = 1,$$

where

$$J_s \in \operatorname*{argmin}_{J:J\supseteq\{(i)\},\,|J|=s} \mathbb{1}_{\{\min_{j\in J} p_j \leq \alpha/s\}}.$$

Then if $s \leq m-i+1$, J_s will contain the indices of the $s-1$ largest p-values, and (i). In this case $\min_{j\in J_s} p_j = p_{(i)}$, and

$$\min_{s\leq m-i+1} \mathbb{1}_{\{p_{(i)}\leq \alpha/s\}} = \mathbb{1}_{\{p_{(i)}\leq \alpha/(m-i+1)\}}.$$

If $s > m-i+1$, then J_s will contain the indices of the s largest p-values, and so then $\min_{j\in J_s} p_j = p_{(m-s+1)}$. Thus in order to reject $p_{(i)}$, we must have

$$\begin{aligned} & p_{(i)} \leq \frac{\alpha}{m-i+1} \\ \text{and} \qquad & p_{(m-s+1)} \leq \frac{\alpha}{s} \qquad \text{for } s > m-i+1, \end{aligned}$$

or equivalently,

$$p_{(j)} \leq \frac{\alpha}{m-j+1} \qquad \text{for } j \leq i.$$

8. The Benjamini–Hochberg procedure allows us to control the FDR when the p-values of true null hypotheses are independent of each other, and independent of the false null hypotheses. The following variant of the method, known as the Benjamini–Yekutieli procedure allows us to control the FDR under arbitrary dependence of the p-values, and works as follows. Define

$$\gamma_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}.$$

Let $\hat{k} = \max\{i : p_{(i)} \leq \alpha i/(m\gamma_m)\}$ and reject $H_{(1)}, \ldots, H_{(\hat{k})}$. First show that the FDR of this procedure satisfies

$$\mathrm{FDR} = \sum_{i \in I_0} \mathbb{E}\left(\frac{1}{R}\mathbb{1}_{\{p_i \leq \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}}\right).$$

Now go on to prove that $\mathrm{FDR} \leq \alpha m_0/m \leq \alpha$. *Hint: Verify that that for any $r \in \mathbb{N}$ we have*

$$\frac{1}{r} = \sum_{j=1}^{\infty} \frac{\mathbb{1}_{\{j \geq r\}}}{j(j+1)},$$

and use this to replace $1/R$.

Solution: The first part is reasonably clear. For the next part, note that

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{\mathbb{1}_{\{j \geq r\}}}{j(j+1)} &= \frac{1}{r(r+1)} + \frac{1}{(r+1)(r+2)} + \cdots \\ &= \frac{1}{r} - \frac{1}{r+1} + \frac{1}{r+1} - \frac{1}{r+2} + \cdots \\ &= \frac{1}{r}.\end{aligned}$$

Next we argue

$$\mathrm{FDR} = \sum_{i \in I_0} \sum_{j=1}^{\infty} \frac{1}{j(j+1)} \mathbb{E}\left(\mathbb{1}_{\{j \geq R\}}\mathbb{1}_{\{p_i \leq \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}}\right).$$

The key thing to note now is that

$$\mathbb{1}_{\{j \geq R\}}\mathbb{1}_{\{p_i \leq \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}} \leq \mathbb{1}_{\{p_i \leq \alpha \min(j,m)/(m\gamma_m)\}}.$$

Taking expectations, we get

$$\mathbb{E}(\mathbb{1}_{\{j \geq R\}}\mathbb{1}_{\{p_i \leq \alpha R/(m\gamma_m)\}}\mathbb{1}_{\{R>0\}}) \leq \mathbb{P}(p_i \leq \alpha \min(j,m)/(m\gamma_m)) \leq \frac{\alpha \min(j,m)}{m\gamma_m}.$$

Thus

$$\mathrm{FDR} \leq \frac{\alpha m_0}{\gamma_m m} \sum_{j=1}^{\infty} \frac{\min(j,m)}{j(j+1)}.$$

But

$$\sum_{j=1}^{\infty} \frac{\min(j,m)}{j(j+1)} = \sum_{j=1}^{m-1} \frac{1}{j+1} + m \sum_{j=m}^{\infty} \frac{1}{j(j+1)} = \gamma_m.$$

9. Consider the closed testing procedure applied to m hypotheses $H_1, \dots, H_m$. Let $\mathcal{R}$ be the collection of all $I \subseteq \{1, \dots, m\}$ for which for all $J \supseteq I$, the local test $\phi_J = 1$. Now suppose that (perhaps after having looked at the results of the ϕ_I), we decide we want to reject a set of hypotheses indexed by $B \subseteq \{1, \dots, m\}$. Let

$$t_\alpha(B) = \max\{|I| : I \subseteq B,\ I \notin \mathcal{R}\}.$$

Show that $\{0, 1, \dots, t_\alpha(B)\}$ gives a $1 - \alpha$ confidence set for the number of false rejections in B. That is, show that

$$\mathbb{P}(|B \cap I_0| > t_\alpha(B)) \leq \alpha,$$

and that this is true no matter how B is chosen. *Hint: Argue by working on the event* $\{\phi_{I_0} = 0\}$.

Solution: On the event $\{\phi_{I_0} = 0\}$, no $I \subseteq I_0$ is in $\mathcal{R}$. Therefore

$$t_\alpha(B) = \max\{|I| : I \subseteq B,\ I \notin \mathcal{R}\} \geq \max\{|I| : I \subseteq B \cap I_0,\ I \notin \mathcal{R}\} = |B \cap I_0|.$$

But the event under consideration has probability $1 - \alpha$.

In the following questions, let all quantities be as defined in Section 4.3 of the lecture notes concerning the debiased Lasso.

10. Show that

$$(\hat{\Theta}\hat{\Sigma}\hat{\Theta}^T)_{jj} = \frac{1}{n}\|X_j - X_{-j}\hat{\gamma}^{(j)}\|_2^2/\hat{\tau}j^4.$$

Solution: It is easy to see that the LHS equals $\|X\hat{\theta}_j\|_2^2/n$, and the result follows from the equation at the bottom of page 57 in the notes.

11. Show that

$$\frac{1}{n}X_j^T(X_j - X_{-j}\hat{\gamma}^{(j)}) = \frac{1}{n}\|X_j - X_{-j}\hat{\gamma}^{(j)}\|_2^2 + \lambda_j\|\hat{\gamma}^{(j)}\|_1.$$

Solution: Let us rewrite $X_j = Y$, $X_{-j} = X$, $\hat{\gamma}^{(j)} = \hat{\beta}$, $\lambda_j = \lambda$ for notational simplicity. Also let $\hat{S} = \{k : \hat{\beta}_k \neq 0\}$. We have

$$\begin{aligned}\frac{1}{n}Y^T(Y - X\hat{\beta}) &= \frac{1}{n}(Y - X\hat{\beta} + X\hat{\beta})^T(Y - X\hat{\beta}) \\ &= \frac{1}{n}\|Y - X\hat{\beta}\|_2^2 + \frac{1}{n}\hat{\beta}^T X^T(Y - X\hat{\beta}).\end{aligned}$$

Also $X_{\hat{S}}^T(Y - X\hat{\beta})/n = \lambda \mathrm{sgn}(\hat{\beta}_{\hat{S}})$, so we see the final term above is $\lambda\|\hat{\beta}\|_1$.

12. Prove that $\mathbb{P}(\Lambda_n) \to 1$, where the sequence of events Λ_n is defined in the proof of Theorem 40.

Solution: We need to show that

(a) $\mathbb{P}(\{\phi^2_{\hat{\Sigma}, s_{\max}} \geq c_{\min}/2\} \cup_j \{\phi^2_{\hat{\Sigma}_{-j,-j}, s_j} \geq c_{\min}/2\}) \to 1$,

(b) $\mathbb{P}(\cup_j\{2\|X_{-j}^T\varepsilon^{(j)}\|_\infty/n > \lambda\} \cup \{2\|X^T\varepsilon\|_\infty/n > \lambda\}) \to 0$,

(c) $\mathbb{P}(\cup_j\{\Omega_{jj}\|\varepsilon^{(j)}\|_2^2/n < 1 - 4\sqrt{\log(p)/n}\}) \to 0$, for A sufficiently large, where $\lambda = A\sqrt{\log(p)/n}$. Consider (a) first. Firstly, we know that $\min_m \phi^2_{\Sigma,m} \geq c_{\min}$ (see the discussion preceding Theorem 23). Clearly also $\min_m \phi^2_{\Sigma_{-j,-j},m} \geq c_{\min}$.

From Lemma 24, we know that on the event

$$\Xi_n = \{\max_{jk} |\hat{\Sigma}_{jk} - \Sigma_{jk}| \leq c_{\min}/(32 s_{\max})\}$$

we have in particular that $\phi^2_{\hat{\Sigma},s} \geq \phi^2_{\Sigma,s}/2 \geq c_{\min}/2$, and also $\phi^2_{\hat{\Sigma}_{-j,-j},s_j} \geq \phi^2_{\Sigma_{-j,-j},s_j}/2 \geq c_{\min}/2$ for all j. Theorem 25 then shows that $\mathbb{P}(\Xi_n) \to 1$ provided $s_{\max}\sqrt{\log(p)/n} \to 0$ (which is true by assumption).

For (c), note that $\|\varepsilon^{(j)}\|_2^2 \Omega_{jj} \sim \chi^2_n$. From question 7 (a) on example sheet 2, we know that if $W \sim \chi^2_d$, then $\mathbb{P}(W/d \leq 1-t) \leq e^{-dt^2/8}$. Thus by a union bound, we have

$$\mathbb{P}(\cup_j \{\Omega_{jj} \|\varepsilon^{(j)}\|_2^2/n < 1 - 4\sqrt{\log(p)/n}\}) \leq p \exp(-2\log(p)) = 1/p \to 0.$$

Finally, for (b), using a union bound and then Lemma 13 we get

$$\begin{aligned}
&\mathbb{P}(\cup_j \{2\|X^T_{-j}\varepsilon^{(j)}\|_\infty/n > \lambda\} \cup \{2\|X^T\varepsilon\|_\infty/n > \lambda\}) \\
\leq &2p^{1-A^2/(8\sigma^2)} + 2\sum_j p^{1-\Omega_{jj}A^2/8}.
\end{aligned}$$

Note $\Omega_{jj} \geq \Sigma_{jj} \geq c_{\min} > 0$. Thus by taking $A > 4\max(\sigma, c_{\min}^{-1})$, the RHS of the above will tend to zero.