Questions 2 and 7 will be marked. In all of the below, assume that any design matrices X are $n \times p$ and have their columns centred and then scaled to have ℓ_2-norm $\sqrt{n}$.

1. In the setting of Theorem 23, assume instead of the compatibility condition that for some $c \in (0,1)$ there exists $\phi > 0$ such that for all $\delta \in \mathbb{R}^p$ with $(1-c)\|\delta_N\|_1 \leq (1+c)\|\delta_S\|_1$, we have
$$\|\delta_S\|_1^2 \leq \frac{s\|X\delta\|_2^2}{n\phi^2}.$$
Let $\hat{\beta}$ be a Lasso estimator with regularisation parameter $\lambda > 0$. Show that with probability at least $1 - 2p^{-(c^2A^2/2-1)}$, we have both
$$\frac{1}{n}\|X(\hat{\beta} - \beta^0)\|_2^2 \leq (1+c)^2\lambda^2\frac{s}{\phi^2} \qquad \text{and} \qquad \|\hat{\beta} - \beta^0\|_1 \leq \lambda\frac{2(1+c)}{1-c}\frac{s}{\phi^2}.$$
[*Hint: Start with the improved version of the basic inequality from Qu. 11 of Sheet 2.*]

2. Let $Y = \mu^0\mathbf{1} + X\beta^0 + \varepsilon$ and let $S = \{k : \beta^0 \neq 0\}$, $N := \{1, \dots, p\} \setminus S$. Without loss of generality assume $S = \{1, \dots, |S|\}$. Assume that X_S has full column rank and let $\Omega = \{\|X^\top\varepsilon\|_\infty/n \leq \lambda_0\}$. Show that, when $\lambda > \lambda_0$, if the following two conditions hold
$$\sup_{\tau:\|\tau\|_\infty \leq 1} \|X_N^\top X_S(X_S^\top X_S)^{-1}\tau\|_\infty < \frac{\lambda - \lambda_0}{\lambda + \lambda_0}$$
$$(\lambda + \lambda_0)\|\{(\tfrac{1}{n}X_S^\top X_S)^{-1}\}_k\|_1 < |\beta_k^0| \qquad \text{for } k \in S,$$
then on Ω, there exists a Lasso solution that satisfies $\mathrm{sgn}(\hat{\beta}_\lambda^{\mathrm{L}}) = \mathrm{sgn}(\beta^0)$. Show moreover that the Lasso solution is unique.

3. Consider the setup of Question 1 with $c = 1/2$ and write $\hat{S} := \{j : \hat{\beta}_j \neq 0\}$ and set $\hat{s} := |\hat{S}|$.
   (a) Show that on the event Ω, for any non-empty subset B of $\hat{S}$, we have
   $$\frac{1}{n}\mathrm{sgn}(\hat{\beta}_B)^\top X_B^\top X(\beta^0 - \hat{\beta}) \geq \frac{\lambda|B|}{2}.$$
   [*Hint: Start with the KKT conditions.*]
   (b) Let κ_m^2 be the maximum eigenvalue of $X_M^\top X_M/n$ over all $M \subset \{1, \dots, p\}$ with $|M| \leq m$. Prove that on Ω, any $B \subseteq \hat{S}$ satisfies
   $$|B| \leq 9s\kappa_{|B|}^2/\phi^2.$$
   Let
   $$m^* = \min\{m \geq 1 : m > 9\kappa_m^2 s/\phi^2\},$$
   with $m^* = \infty$ if there does not exist any m satisfying the condition defining the set above. Deduce that on Ω, $\hat{s} < m^*$, and moreover that $\hat{s} \leq 9\kappa_{m^*}^2 s/\phi^2$.

4. (a) Show that
   $$\max_{\theta:\|X^\top\theta\|_\infty \leq \lambda} G(\theta) = \frac{1}{2n}\|Y - X\hat{\beta}_\lambda^{\mathrm{L}}\|_2^2 + \lambda\|\hat{\beta}_\lambda^{\mathrm{L}}\|_1,$$

where

$$G(\theta) = \frac{1}{2n}\|Y\|_2^2 - \frac{1}{2n}\|Y - n\theta\|_2^2.$$

Show that the unique θ maximising G is $\theta^* = (Y - X\hat{\beta}^{\mathrm{L}}_\lambda)/n$. [*Hint: Treat the Lasso optimisation problem as minimising* $\|Y - z\|_2^2/(2n) + \lambda\|\beta\|_1$ *subject to* $z - X\beta = 0$ *over* $(\beta, z) \in \mathbb{R}^p \times \mathbb{R}^n$ *and consider the Lagrangian.*]

(b) Let $\tilde{\theta}$ be such that $\|X^\top\tilde{\theta}\|_\infty \le \lambda$. Explain why if

$$\max_{\theta: G(\theta) \ge G(\tilde{\theta})} |X_k^\top \theta| < \lambda,$$

then we know that $\hat{\beta}^{\mathrm{L}}_{\lambda,k} = 0$. By considering $\tilde{\theta} = Y\lambda/(n\lambda_{\max})$ with $\lambda_{\max} = \|X^\top Y\|_\infty/n$, show that $\hat{\beta}^{\mathrm{L}}_{\lambda,k} = 0$ if

$$\frac{1}{n}|X_k^\top Y| < \lambda - \frac{\|Y\|_2}{\sqrt{n}}\frac{\lambda_{\max} - \lambda}{\lambda_{\max}}.$$

5. Suppose $\hat{\beta}$ is a square-root Lasso solution from a regression of Y onto X with regularisation parameter $\gamma > 0$. Show that provided $Y - \bar{Y}\mathbf{1} \neq X\hat{\beta}$, we have

$$\frac{1}{\sqrt{n}}\frac{(Y - \bar{Y}\mathbf{1})^\top(Y - \bar{Y}\mathbf{1} - X\hat{\beta})}{\|Y - \bar{Y}\mathbf{1} - X\hat{\beta}\|_2} = \frac{1}{\sqrt{n}}\|Y - \bar{Y}\mathbf{1} - X\hat{\beta}\|_2 + \gamma\|\hat{\beta}\|_1.$$

6. The elastic net estimator in the linear model minimises

$$\frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda(\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2/2)$$

over $\beta \in \mathbb{R}^p$, where $\alpha \in [0,1]$ is fixed.

(a) Suppose X has two columns X_j and X_k that are identical and $\alpha < 1$. Explain why the minimising β^* above is unique and has $\beta^*_k = \beta^*_j$.

(b) Let $\hat{\beta}^{(0)}, \hat{\beta}^{(1)}, \ldots$ be the solutions from iterations of a coordinate descent procedure to minimise the elastic net objective. For a fixed variable index k, let $A = \{1, \ldots, k-1\}$ and $B = \{k+1, \ldots, p\}$ and find the form of $\hat{\beta}^{(m)}_k$ in terms of $\hat{\beta}^{(m)}_A$ and $\hat{\beta}^{(m-1)}_B$.

7. Consider the model

$$Y = X(\beta^0 + \delta^0) + \varepsilon,$$

where $\delta^0 \in \mathbb{R}^p$ represents a dense perturbation of the usual sparse linear model defined by $\beta^0 \in \mathbb{R}^p$ alone. The *Lava estimator* $(\hat{\beta}_\lambda, \hat{\delta}_\lambda)$ with tuning parameter $\lambda = (\lambda_1, \lambda_2)^\top \in (0, \infty)^2$ is defined by

$$(\hat{\beta}_\lambda, \hat{\delta}_\lambda) := \operatorname*{argmin}_{(\beta,\delta) \in \mathbb{R}^p \times \mathbb{R}^p} \left\{\frac{1}{2n}\|Y - X(\beta + \delta)\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\delta\|_2^2\right\}.$$

Find an expression for $\hat{\delta}_\lambda$ involving X, Y, $\hat{\beta}_\lambda$, λ_2 and n. Deduce that $\hat{\beta}_\lambda$ is the minimiser of a Lasso objective with transformed design matrix $\tilde{X} := AX$ and transformed response AY, where $A := (I - XQX^\top)^{1/2}$ and $Q := (X^\top X + 2n\lambda_2 I)^{-1} \in \mathbb{R}^{p\times p}$.

Now let $\Omega := \left\{\left\|\tilde{X}^\top(\tilde{X}\delta^0 + A\varepsilon)\right\|_\infty/n \le \lambda_1\right\}$. Show that on Ω, we have

$$\frac{1}{n}\left\|\tilde{X}(\hat{\beta}_\lambda - \beta^0)\right\|_2^2 \le 4\lambda_1\|\beta^0\|_1.$$

Conclude that on Ω we have

$$\frac{1}{n}\|X(\hat{\beta}_\lambda - \beta^0)\|_2^2 \leq \frac{4\lambda_1\|\beta^0\|_1}{1-\kappa}$$

when the maximum eigenvalue of $XQX^\top$ is $\kappa < 1$.

8. By using applications of the weak union and contraction properties or otherwise, answer the following.

   (a) Suppose $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are sequences of random vectors satisfying

   $$Y_t \perp\!\!\!\perp X_{t-1}, \ldots, X_1 \,|\, Y_{t-1}, \ldots, Y_1$$

   for all $t \in \mathbb{N}$. Show that for all $r \in \mathbb{N}$, we also have

   $$Y_t, \ldots, Y_{t+r} \perp\!\!\!\perp X_{t-1}, \ldots, X_1 \,|\, Y_{t-1}, \ldots, Y_1.$$

   (b) Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are independent triples satisfying $X_i \perp\!\!\!\perp Y_i \,|\, Z_i$. Show that

   $$X_1, \ldots, X_n \perp\!\!\!\perp Y_1, \ldots, Y_n \,|\, Z_1, \ldots, Z_n.$$

   [*Hint: Argue that it suffices to show the result for* $n = 2$.]

9. Let $Z \sim N_p(\mu, \Sigma)$ with Σ positive definite. Show that for any $A, B \subset [p]$,

   $$Z_A | Z_B = z_B \sim N_{|A|}(\mu_A + \Sigma_{A,B}\Sigma_{B,B}^{-1}(z_B - \mu_B),\, \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}).$$

   Here $\Sigma_{A,B}$, for example, is the submatrix of Σ formed of those rows and columns indexed by A and B respectively. [*Hint: Find a matrix* $M \in \mathbb{R}^{|A|\times|B|}$ *such that* $Z_A - MZ_B$ *and* Z_B *are independent.*]