P2

- Let's assume $X\beta_1 \neq X\beta_2$.

  Let f^* be the optimal value, $\alpha = \frac{1}{2}$, $\beta_\alpha = \alpha\beta_1 + (1-\alpha)\beta_2$. Then, by the convexity of $\|.\|_2^2, \|.\|_1$:

  $$\begin{aligned} f^* &\leq ||Y - X\beta_\alpha||_2^2 + \lambda||\beta_\alpha||_1 \\ &= \|\alpha(Y - X\beta_1) + (1-\alpha)(Y - X\beta_2)\|_2^2 + \lambda\|\alpha\beta_1 + (1-\alpha)\beta_2\|_1 \\ &< \alpha\left(||Y - X\beta_1||_2^2 + \lambda||\beta_1||_1\right) + (1-\alpha)\left(||Y - X\beta_2||_2^2 + \lambda||\beta_2||_1\right) p \quad \text{(By strict convexity of } \|.\|_2^2) \\ &\leq f^* \end{aligned}$$

  Contradicition.

- $\mathcal{L}(\beta^*, \lambda) = \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$

  $\partial\|\beta\|_1 = \{\alpha \in [-1,1]^n, \alpha_j = sign(\hat{\beta}_j) \text{ when } \hat{\beta}_j \neq 0\}$

  Let (β^*, λ^*) be an optimal solution, then $0 \in \partial_\lambda L(\beta^*, \lambda^*)$

  $\partial_{\lambda^*} L(\beta, \lambda^*) = -X^T(Y - X\beta) + \lambda^*\partial\|\beta\|_1$

  Coordinate wise, this gives for all j:

  $X_j^T(Y - X\beta) = \lambda sign(\beta_j)$ if $\beta_j \neq 0$

  $-X(Y - X\beta) = \lambda\alpha_i$ if $\beta_j = 0$

  e.g

  $\lambda^* = -sign(\beta_j^*)X_j^T(Y - X\beta^*)$ if $\beta_j^* \neq 0$

  $\lambda^* \geq |2X_j^T(Y - X\beta^*)|$ if $\beta_j^* = 0$

- Let $\hat{\beta}$ be an optimal solution. Let $\chi = \{j, \hat{\beta}_j \neq 0\}$, and let's suppose it is non empty. Let j such that $\hat{\beta}_j > 0$ (If such j exists)

  By 2.2, $\lambda = X_j^T(Y - X\hat{\beta})$, but since $\lambda > \|X^TY\|_\infty \geq X_j^TY$, then $X_j^TX\hat{\beta} > 0$.

  Similarly, if there for j such that $\hat{\beta} < 0$, $X_j^TX\hat{\beta} < 0$.

  c/c $\beta_j \neq 0 \implies \beta_j X_j^T X\hat{\beta} > 0$

  $$\begin{aligned} \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 &= \frac{1}{2}\|Y\|_2^2 - \hat{\beta}^TX^TY + \frac{1}{2}\beta^TX^TX\hat{\beta} + \lambda\sum_{i\in\chi}|\hat{\beta}_i| \\ &\geq \frac{1}{2}\|Y\|_2^2 + \sum_{i\in\chi}|\hat{\beta}_i|(\lambda - |X_i^TY|) + \frac{1}{2}\underbrace{\sum_{i\in\chi}\hat{\beta}_iX_i^TX\hat{\beta}}_{>0} \\ &> \frac{1}{2}\|Y\|_2^2 \\ &= \frac{1}{2}\|Y - X0\|_2^2 + \lambda\|0\|_1 \end{aligned}$$

  Contradiction, so $\hat{\beta} = 0$

- $$\lambda \in [\lambda_0, \lambda_1]$$

  Let $\chi(\lambda) = \{j, \hat{\beta}_j(\lambda) \neq 0\} := \chi$, $r = |\chi|$ (doesn't depend on λ by assumption)

  We have proved in 2.2 that there exist $\alpha(\lambda)$

  $$X^T(Y - X\hat{\beta}(\lambda)) = \lambda\alpha(\lambda)$$

  where $\alpha(\lambda) \in \partial\|\hat{\beta}(\lambda)\|_1$.

  It is easy to see that this KKT conditions is actualy necessary and sufficient (because we are minimizing a convexe function), since we are assuming uniqueness, $\hat{\beta}(\lambda)$ is the unique solution to :

  $$(\exists\alpha(\lambda) \in \partial\|\hat{\beta}(\lambda)\|_1)\ X^T(Y - X\hat{\beta}(\lambda)) = \lambda\alpha(\lambda)$$

  Note that by uniqueness of $X\beta$ and $\hat{\beta}(\lambda)$, $\alpha(\lambda)$ is unique when $\lambda > 0$.

  Note also, that since we assumed that the signs and support are unchanged, $\partial\|\hat{\beta}(\lambda)\|_1 = \partial\|\hat{\beta}(\lambda_0)\|_1$.

  The last condition becomes:

  $$X^T(Y - X\hat{\beta}(\lambda)) \in \lambda\partial\|\hat{\beta}(\lambda_0)\|_1$$

  Notation: $\alpha(\lambda_0) = X^T \underbrace{\frac{(Y - X\hat{\beta}(\lambda_0))}{\lambda_0}}_{v} = X^T v$, $\gamma_0 = X^\dagger v$, $\delta = \hat{\beta}(\lambda_0) - (\lambda - \lambda_0)\gamma_0$.

  Note that:

  $$X^T X\gamma_0 = X^T X X^\dagger v = (V\Lambda U^T)(U\Lambda V^T)(V\Lambda^{-1}U^T)v = V\Lambda U^T v = X^T v = \alpha(\lambda_0)$$

  $$\begin{aligned} X^T(Y - X\delta) &= \underbrace{X^T(Y - X\hat{\beta}(\lambda_0))}_{\lambda_0\alpha(\lambda_0)} + (\lambda - \lambda_0)\underbrace{X^T X\alpha_0}_{\alpha(\lambda_0)} \\ &= \lambda\alpha(\lambda_0) \in \lambda\partial\|\hat{\beta}(\lambda_0)\|_1 \end{aligned}$$

  Which proves that $\hat{\beta}(\lambda) = \delta = \hat{\beta}(\lambda_0) - (\lambda - \lambda_0)\alpha(\lambda_0)$

- **Notation**: For a vector v, let $v^+ = \max(v, 0), v^- = -\min(-v, 0),\ sign(v), supp(v)$ the sign and support of v, $\phi(v) = (supp(v^+), supp(v^-))$

  The number of values $\phi(v)$ can take is finite and at most n^2 because $\phi(v) \in \mathcal{P}(\{1 \ldots n\})^2$.

  Notice that in the last part, we have proven a stronger result: if for λ_1, λ_2, $\phi(\beta(\lambda_1)) = \phi(\beta(\lambda_2))$, then $\beta(\lambda_2) = \beta(\lambda_1) - (\lambda_2 - \lambda_1)\gamma_0$, where γ_0 depend only on λ_1. This proves a segment of the path C is fully caracterized by the $\phi(v)$ where $v(C)$ is one of the element of C chosen arbitrarly.

Let $\mathcal{A}$ denote the set of segments that form the lasso path, and consider the following application:

$\mathcal{A} \to \mathcal{B}; C \to \phi(v(C))$ Where v is an arbitrary element in C.

We have proven that this application is injective, so $|\mathcal{A}| \leq n^2 < \infty$. Which proves that the number of segments in the lasso path is finite. Let λ_0 be small enough so that $(0, \lambda_0]$ corresponds to last segment, and let $0 < \lambda < \lambda_0$ then $\hat{\beta}(\lambda) = \hat{\beta}(\lambda_0) - (\lambda - \lambda_0)\gamma_0$.

This proves that the $\hat{\beta}(\lambda)$ has $\hat{\beta}(\lambda_0) + \lambda_0\gamma_0$ as limit at 0^+. Let's call $\hat{\beta}$ that limit.

Recall the definition of γ_0: $X^\dagger \frac{(Y - X\hat{\beta}(\lambda_0))}{\lambda_0}$, and sine X is full rank, $XX^\dagger = I_n$. So $X\hat{\beta} = \hat{\beta}(\lambda_0) + XX^\dagger(Y - X\hat{\beta}(\lambda_0)) = Y$

Suppose by contradiction that $\hat{\beta} \neq \hat{\beta}^{CS}$, e.g $\|\hat{\beta}^{CS}\|_1 < \|\hat{\beta}\|_1$.

By continuity of norms in finite dimensional space, $\|\hat{\beta}(\lambda)\|_1 \to \|\hat{\beta}\|_1$. Let λ be small enought so that

$$\|\hat{\beta}^{CS}\|_1 < \|\hat{\beta}(\lambda)\|$$

Which would imply that $\lambda\|\hat{\beta}^{CS}\|_1 < \lambda\|\hat{\beta}(\lambda)\| + \|Y - X\hat{\beta}(\lambda)\|_2^2$, which contradicts the minimality of $\hat{\beta}(\lambda)$.

Conclusion: $\hat{\beta} = \hat{\beta}^{CS}$

P3

- Let's consider the unconstrained optimization problem:

$$\min ||Y - X\beta||^2$$

  β is optimal iff $X^TY = X^TX\beta$.

  We check easily that $(X^TX)^\dagger X^TY$ is a solution to the last equation, therefore it minimizes the L_2 risk.

  If $t > \|(X^TX)^\dagger X^TY\|_{L_1}$, then it is also solution to the following problem: $\min_{\|\beta\|_{L_1} \leq t} ||Y - X\beta||^2$.

- We adopt the following notations:

  -
$$\gamma = (-1, \beta)^T, |\gamma|_1 = |\beta| + 1, Z_i = (Y_i, X_i^T)^T$$

  -
$$R(\beta) = \gamma^T \Sigma \gamma$$

  -
$$\hat{R}(\beta) = \gamma^T \hat{\Sigma} \gamma$$

  -
$$\hat{R}^{(V_k)}(\beta) := \frac{1}{|V_K|} \sum_{i \in V_k} (Y_i - X_i^T\beta)^2 = \gamma^T \hat{\Sigma}^{V_k} \gamma$$

–
$$\hat{R}^{(-V_k)}(\beta) := \sum_{j\neq k}\frac{1}{|V_j|}\sum_{i\in V_j}(Y_i - X_i^T\beta)^2 = \gamma^T\underbrace{(\hat{\Sigma}-\hat{\Sigma}^{V_k})}_{\hat{\Sigma}^{(-V_k)}}\gamma$$

Note that $|Z_i| \leq 2b, |Z_iZ_i^T| \leq 4b^2$, by Hoeffding inequality:

*
$$\mathbb{P}(||\Sigma-\hat{\Sigma}||_\infty \geq \varepsilon) \leq 2\exp(\frac{-n\varepsilon^2}{8b^4})$$

*
$$\mathbb{P}(||\Sigma-\hat{\Sigma}^{(V_k)}||_\infty \geq \varepsilon) \leq 2\exp(\frac{-|V_k|\varepsilon^2}{8b^4})$$

*
$$\mathbb{P}(||\Sigma-\hat{\Sigma}^{(-V_k)}||_\infty \geq \varepsilon) \leq 2\exp(\frac{-(n-|V_k|)\varepsilon^2}{8b^4}) \leq 2\exp(\frac{-|V_k|\varepsilon^2}{8b^4})$$

*
$$\mathbb{P}(||\hat{\Sigma}-\hat{\Sigma}^{(V_k)}||_\infty \geq \varepsilon) \leq \mathbb{P}(||\Sigma-\hat{\Sigma}^{(V_k)}||_\infty \geq \frac{\varepsilon}{2})+\mathbb{P}(||\Sigma-\hat{\Sigma}||_\infty \geq \frac{\varepsilon}{2}) \leq 4\exp(\frac{-|V_k|\varepsilon^2}{32b^4})$$

*
$$\mathbb{P}(||\hat{\Sigma}^{(-V_k)}-\hat{\Sigma}^{(V_k)}||_\infty \geq \varepsilon) \leq \mathbb{P}(||\Sigma-\hat{\Sigma}^{(V_k)}||_\infty \geq \frac{\varepsilon}{2})+\mathbb{P}(||\Sigma-\hat{\Sigma}^{(-V_k)}||_\infty \geq \frac{\varepsilon}{2}) \leq 4\exp(\frac{-|V_k|\varepsilon^2}{32b^4})$$

1.)

$$R(\hat{\beta}_{\hat{t}}) - \hat{R}_{CV}(\hat{t}) = R(\hat{\beta}_{\hat{t}}) - \hat{R}(\hat{\beta}_{\hat{t}}) + \hat{R}(\hat{\beta}_{\hat{t}}) - \hat{R}_{CV}(\hat{t})$$

$$R(\hat{\beta}_{\hat{t}}) - \hat{R}(\hat{\beta}_{\hat{t}}) = \hat{\beta}_{\hat{t}}(\Sigma-\hat{\Sigma})\hat{\beta}_{\hat{t}} \leq ||\Sigma-\hat{\Sigma}||_\infty|\gamma|_1 \leq ||\Sigma-\hat{\Sigma}||_\infty(1+t_n)$$

$$\begin{aligned}
\hat{R}(\hat{\beta}_{\hat{t}}) - \hat{R}_{CV}(\hat{t}) &= \frac{1}{K}\sum_k(\hat{R}(\hat{\beta}_{\hat{t}}) - \hat{R}^{V_k}(\hat{\beta}_{\hat{t}}^{(V_k)}) \\
&\leq \frac{1}{K}\sum_k \hat{R}(\hat{\beta}_{\hat{t}}^{V_k}) - \hat{R}^{V_k}(\hat{\beta}_{\hat{t}}^{(V_k)} \\
&= \frac{1}{K}\sum \hat{(}\gamma_{\hat{t}}^{V_k})^T(\hat{\Sigma}-\hat{\Sigma}^{V_k})\hat{\gamma}_{\hat{t}}^{V_k} \\
&\leq \frac{1}{K}|\gamma_{\hat{t}}^{V_k}|_{L_1}^2\sum ||\hat{\Sigma}-\hat{\Sigma}^{V_k}||_\infty \\
&\leq \frac{1}{K}(1+t_n)^2\sum ||\hat{\Sigma}-\hat{\Sigma}^{V_k}||_\infty
\end{aligned}$$

$$
\begin{aligned}
\mathbb{P}((i) \geq \varepsilon) &\leq \mathbb{P}(||\Sigma - \hat{\Sigma}||_\infty \geq \frac{\varepsilon}{(1+t_n)^2}) + \sum_k \mathbb{P}(\frac{1}{K}||\hat{\Sigma} - \hat{\Sigma}^{(V_k)}||_\infty \geq \frac{\varepsilon}{(1+t_n)^2 K}) \\
&\leq 2\exp(\frac{-n\varepsilon^2}{16b^4(1+t_n^2)^2}) + 2K\exp(\frac{-|V_1|\varepsilon^2}{16b^4(1+t_n^2)^2}) \\
&\leq 2(1+K)\exp(\frac{-|V_1|\varepsilon^2}{16b^4(1+t_n^2)^2}) \\
&\leq C_1 K \exp(\exp(\frac{-|V_1|\varepsilon^2}{C_2 b^4(1+t_n^2)^2})
\end{aligned}
$$

Where $C_1 \geq 4, C_2 \geq 32$

2.)

$$\hat{R}_{CV}(\hat{t}) - \hat{R}_{CV}(t_{\max}) \leq 0$$

3.)

$$
\begin{aligned}
&\hat{R}_{CV}(t_{\max}) - \hat{R}(\hat{\beta}_{t_{\max}}) \\
&= \frac{1}{K}\sum_k \hat{R}^{V_k}(\hat{\beta}^{V_k}_{t_{\max}}) - \hat{R}(\hat{\beta}_{t_{\max}}) \\
&= \frac{1}{K}\sum_k \hat{R}^{V_k}(\hat{\beta}^{V_k}_{t_{\max}}) + R^{-V_k}(\hat{\beta}^{V_k}_{t_{\max}}) + \underbrace{R^{-V_k}(\hat{\beta}^{V_k}_{t_{\max}}) - \hat{R}^{-V_k}(\hat{\beta}_{t_{\max}})}_{\leq 0} - \hat{R}^{V_k}(\hat{\beta}_{t_{\max}}) \\
&\leq \frac{1}{K}\sum_k \hat{R}^{V_k}(\hat{\beta}^{V_k}_{t_{\max}}) - R^{-V_k}(\hat{\beta}^{V_k}_{t_{\max}}) \\
&= \frac{1}{K}\sum_k (\hat{\gamma}^{V_k}_{t_{\max}})^T(\hat{\Sigma}^{V_k} - \hat{\Sigma}^{-V_k})\hat{\gamma}^{V_k}_{t_{\max}} \\
&\leq \frac{(1+t_n)^2}{K}\sum_k ||\hat{\Sigma}^{V_k} - \hat{\Sigma}^{-V_k}||_\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(iii \geq \varepsilon) &\leq K\mathbb{P}(||\hat{\Sigma}^{V_1} - \hat{\Sigma}^{-V_1}||_\infty \geq \frac{\varepsilon}{(1+t_n)^2}) \\
&\leq 4Kexp(\frac{-|V_1|\varepsilon^2}{32(1+t_n)^2 b^4}) \\
&\leq C_1 Kexp(\frac{-|V_1|\varepsilon^2}{C_2(1+t_n)^2 b^4})
\end{aligned}
$$

4.)

$$\hat{R}(\hat{\beta}_{t_{\max}}) = \hat{R}(\hat{\beta}_{t_n})$$

5. and 6.)

$$
\begin{aligned}
\hat{R}(\hat{\beta}_{t_n}) - R(\beta_{t_n}) & \\
&\leq \hat{R}(\beta_{t_n}) - R(\beta_{t_n}) \\
&\leq \gamma_{t_n}(\hat{\Sigma} - \Sigma)\gamma_{t_n} \\
&\leq (1+t_n)^2|\hat{\Sigma} - \Sigma|_\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(v + vi \geq \varepsilon) &\leq \mathbb{P}(||\Sigma - \hat{\Sigma}||_\infty \geq \frac{\varepsilon}{(1+t_n)^2}) \\
&\leq 2\exp(-\frac{n\varepsilon^2}{8b^4(1+t_n)^2}) \\
&\leq C_1 K \exp(-\frac{|V_1|\varepsilon^2}{C_2 b^4 (1+t_n)^2})
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(i + ii + iii + iv + v + vi \geq \varepsilon) &= \mathbb{P}(i \geq \varepsilon/3) + \mathbb{P}(i \geq \varepsilon/3) + \mathbb{P}(v + vi \geq \varepsilon/3) \\
&\leq 3C_1 K \exp(-\frac{|V_1|\varepsilon^2}{9C_2 b^4 (1+t_n)^2}) \\
&\leq CK\exp(-\frac{n\varepsilon^2}{CK(1+t_n)^2})
\end{aligned}
$$

Where $C \geq \max(3C_1, 9C_2b^4)$

Which gives the result by setting δ to $CK\exp(-\frac{n\varepsilon^2}{CK(1+t_n)^2})$, or equivalently setting ε to $C(1+t_n^2)\sqrt{\log(\frac{CK}{\delta})K/n}$