



□ FACULTY OF SCIENCE,
TECHNOLOGY
AND MEDICINE

UNIVERSITY OF LUXEMBOURG

THESIS FOR THE BACHELOR OF MATHEMATICS

High Dimensional Regression Models

BY JORIS LIMONIER

Supervisor: Mark PODOLSKIJ

Submitted: June 4, 2021

Abstract

Abstract goes here.

Contents

0	Notes	1
0.1	Notes for chapter 1	1
0.2	Notes for chapter 2	1
0.2.1	Least squares	1
0.3	Notes for chapter 3	1
0.3.1	Section 6.2	1
0.3.2	Section 6.3	7
1	Introduction	11
2	Classical theory of Linear Regression	13
3	Theory for LASSO in high dimensions	15

Chapter 0

Notes

0.1 Notes for chapter 1

To be added

- how to get \hat{b} on page 101.
- where the χ^2 distribution comes from in page 101

0.2 Notes for chapter 2

0.2.1 Least squares

Definition 0.1 (Linear regression model). *We define the linear regression model as*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

Least-squares consists of solving the minimizing the following function

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (2)$$

0.3 Notes for chapter 3

0.3.1 Section 6.2

We define $\hat{\boldsymbol{\beta}}$ as follows

$$\hat{\boldsymbol{\beta}} := \arg \min_{\boldsymbol{\beta}} \left\{ \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}\|_1 \right\} \quad (3)$$

Lemma 0.2 (Basic Inequality).

$$\frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq 2 \frac{\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)}{n} + \lambda \|\boldsymbol{\beta}^0\|_1$$

Proof. By definition of $\hat{\boldsymbol{\beta}}$, we have that

$$\forall \boldsymbol{\beta} \quad \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}\|_1$$

In particular for $\boldsymbol{\beta} = \boldsymbol{\beta}^0$ we have

$$\frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^0\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1$$

We now replace \mathbf{Y} using equation (1).

$$\begin{aligned} & \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^0\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\|(\mathbf{X}\boldsymbol{\beta}^0 + \boldsymbol{\varepsilon}) - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|(\mathbf{X}\boldsymbol{\beta}^0 + \boldsymbol{\varepsilon}) - \mathbf{X}\boldsymbol{\beta}^0\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\|\mathbf{X}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{X}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}^0) + \boldsymbol{\varepsilon}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\langle \mathbf{X}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}, \mathbf{X}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon} \rangle}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\boldsymbol{\varepsilon}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\|\mathbf{X}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})\|_2^2 + \|\boldsymbol{\varepsilon}\|_2^2 + 2\langle \mathbf{X}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}), \boldsymbol{\varepsilon} \rangle}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\boldsymbol{\varepsilon}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\langle \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0), \boldsymbol{\varepsilon} \rangle}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \\ \implies & \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)}{n} + \lambda \|\boldsymbol{\beta}^0\|_1 \end{aligned}$$

□

Let

$$\mathcal{T} := \left\{ \max_{1 \leq j \leq p} 2 \frac{|\boldsymbol{\varepsilon}^T \mathbf{X}^{(j)}|}{n} \leq \lambda_0 \right\}$$

Lemma 0.3 (Lemma 6.2.). *For all $t > 0$ and*

$$\lambda_0 := 2\sigma \sqrt{\frac{t^2 + 2 \log p}{n}}$$

we have

$$\mathbb{P}(\mathcal{T}) \geq 1 - 2 \exp[-t^2/2]$$

Proof. We define

$$V_j := \frac{\varepsilon^T \mathbf{X}^{(j)}}{\sqrt{n\sigma^2}}$$

Then we have

$$\begin{aligned} \mathbb{P}(\mathcal{T}) &= \mathbb{P}\left(\max_{1 \leq j \leq p} 2 \frac{|\varepsilon^T \mathbf{X}^{(j)}|}{n} \leq 2\sigma \sqrt{\frac{t^2 + 2 \log p}{n}}\right) \\ &= \mathbb{P}\left(\max_{1 \leq j \leq p} \left| \frac{\varepsilon^T \mathbf{X}^{(j)}}{\sqrt{n\sigma^2}} \right| \leq \sqrt{t^2 + 2 \log p}\right) \\ &= \mathbb{P}\left(\max_{1 \leq j \leq p} |V_j| \leq \sqrt{t^2 + 2 \log p}\right) \\ &= 1 - \mathbb{P}\left(\max_{1 \leq j \leq p} |V_j| > \sqrt{t^2 + 2 \log p}\right) \\ &= 1 - \mathbb{P}\left(\bigcup_{j=1}^p |V_j| > \sqrt{t^2 + 2 \log p}\right) \\ &\geq 1 - \sum_{j=1}^p \mathbb{P}\left(|V_j| > \sqrt{t^2 + 2 \log p}\right) \\ &\geq 1 - p \mathbb{P}\left(|V_j| > \sqrt{t^2 + 2 \log p}\right) \end{aligned} \tag{4}$$

Now, let us define $\zeta := \sqrt{t^2 + 2 \log p}$.
Since V_j is $\mathcal{N}(0, 1)$ -distributed and $\zeta > 0$.

$$\begin{aligned} \mathbb{P}(V_j > \zeta) &= \frac{1}{\sqrt{2\pi}} \int_{\zeta}^{\infty} e^{-y^2/2} dy \\ &< \frac{1}{\sqrt{2\pi}} \int_{\zeta}^{\infty} \frac{y}{\zeta} * e^{-y^2/2} dy \\ &= \frac{1}{\zeta \sqrt{2\pi}} \int_{\zeta}^{\infty} y * e^{-y^2/2} dy \\ &= \frac{1}{\zeta \sqrt{2\pi}} e^{-\zeta^2/2} \end{aligned}$$

We note that $p \geq 2 \implies \zeta \sqrt{2\pi} \geq 1$ therefore

$$\mathbb{P}(V_j > \zeta) < e^{-\zeta^2/2}$$

Moreover by symmetry of the $\mathcal{N}(0, 1)$ distribution,

$$\begin{aligned}\mathbb{P}(|V_j| > \zeta) &= \mathbb{P}(V_j > \zeta) + \mathbb{P}(-V_j < -\zeta) \\ &= 2\mathbb{P}(V_j > \zeta) \\ &< 2e^{-\zeta^2/2}\end{aligned}$$

Thus by definition of ζ

$$\begin{aligned}\mathbb{P}(|V_j| > \zeta) &< 2e^{-\zeta^2/2} \\ &= 2 \exp \left[\frac{-\sqrt{t^2 + 2 \log p}^2}{2} \right] \\ &= 2 \exp \left[\frac{-t^2}{2} - \log p \right] \\ &= 2 \exp \left[\frac{-t^2}{2} \right] \exp \left[\log \frac{1}{p} \right] \\ &= \frac{2}{p} \exp \left[\frac{-t^2}{2} \right]\end{aligned}$$

Inserting this result into (4) we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{T}) &\geq 1 - p \mathbb{P} \left(|V_j| > \sqrt{t^2 + 2 \log p} \right) \\ &\geq 1 - p \frac{2}{p} \exp \left[\frac{-t^2}{2} \right] \\ &= 1 - 2 \exp \left[\frac{-t^2}{2} \right]\end{aligned}$$

□

Corollary 0.4 (Consistency of the LASSO). *Assume $\sigma^2 = 1$ for all j . We define the regularization parameter as*

$$\lambda = 4\hat{\sigma}^2 \sqrt{\frac{t^2 + 2 \log p}{n}}$$

where $\hat{\sigma}$ is some estimator of σ .

Then with probability at least $1 - \alpha$, where $\alpha := 2 \exp(-t^2/2) + \mathbb{P}(\hat{\sigma} \leq \sigma)$ we have

$$2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} \leq 3\lambda \|\boldsymbol{\beta}^0\|_1$$

Lemma 0.5 (Lemma 6.3.). *We have on \mathcal{T} , with $\lambda \geq 2\lambda_0$,*

$$2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \leq 3\lambda \|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\|_1 \quad (5)$$

Proof. We start with the Basic Inequality

$$\frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq 2 \frac{\varepsilon^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)}{n} + \lambda \|\boldsymbol{\beta}^0\|_1$$

Now since we are on \mathcal{T} and since $2\lambda_0 \leq \lambda$

$$\frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \lambda_0 \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 + \lambda \|\boldsymbol{\beta}^0\|_1$$

$$2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + 2\lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \lambda \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 + 2\lambda \|\boldsymbol{\beta}^0\|_1$$

Let $\boldsymbol{\beta}_{j,S} := \boldsymbol{\beta}_j 1\{j \in S\}$. We use the triangle inequality on the left hand side

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}\|_1 &= \|\hat{\boldsymbol{\beta}}_{S_0}\|_1 + \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \\ &= \|\boldsymbol{\beta}_{S_0}^0 - \boldsymbol{\beta}_{S_0}^0 + \hat{\boldsymbol{\beta}}_{S_0}\|_1 + \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \\ &\geq \|\boldsymbol{\beta}_{S_0}^0\|_1 - \|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\|_1 + \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \end{aligned}$$

whereas on the right hand side

$$\begin{aligned} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 &= \|(\hat{\boldsymbol{\beta}}_{S_0} + \hat{\boldsymbol{\beta}}_{S_0^c}) - (\boldsymbol{\beta}_{S_0}^0 + \underbrace{\boldsymbol{\beta}_{S_0^c}^0}_{=0})\|_1 \\ &= \|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\|_1 + \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \end{aligned}$$

Injecting these two results, we get that

$$\begin{aligned}
& 2 \frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + 2\lambda \|\hat{\beta}\|_1 \leq \lambda \|\hat{\beta} - \beta^0\|_1 + 2\lambda \|\beta^0\|_1 \\
\implies & 2 \frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + 2\lambda \left(\|\beta_{S_0}^0\|_1 - \|\hat{\beta}_{S_0} - \beta_{S_0}^0\|_1 + \|\hat{\beta}_{S_0^c}\|_1 \right) \\
& \leq \lambda \left(\|\hat{\beta}_{S_0} - \beta_{S_0}^0\|_1 + \|\hat{\beta}_{S_0^c}\|_1 \right) + 2\lambda \|\beta^0\|_1 \\
\implies & 2 \frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + 2\lambda \underbrace{\|\beta_{S_0}^0\|_1}_{=\beta^0} + \lambda \|\hat{\beta}_{S_0^c}\|_1 \leq 3\lambda \|\hat{\beta}_{S_0} - \beta_{S_0}^0\|_1 + 2\lambda \|\beta^0\|_1 \\
\implies & 2 \frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + \lambda \|\hat{\beta}_{S_0^c}\|_1 \leq 3\lambda \|\hat{\beta}_{S_0} - \beta_{S_0}^0\|_1
\end{aligned}$$

□

Definition 0.6 (Compatibility condition). *We say that the compatibility condition is met for the set S_0 , if for some $\phi_0 > 0$, and for all β satisfying $\|\beta_{S_0^c}\|_1 \leq 3\|\beta_{S_0}\|_1$, it holds that*

$$\|\beta_{S_0}\|_1^2 \leq \left(\beta^T \hat{\Sigma} \beta \right) \frac{s_0}{\phi_0^2} \quad (6)$$

Theorem 0.7 (Theorem 6.1.). *Suppose the compatibility condition holds for S_0 . Then on \mathcal{T} , we have for $\lambda \geq 2\lambda_0$,*

$$\frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + \lambda \|\hat{\beta} - \beta^0\|_1 \leq 4\lambda^2 \frac{s_0}{\phi_0^2}$$

Proof. Using lemma 5 we have that

$$\begin{aligned}
& 2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 \\
&= 2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}_{S_0} + \hat{\boldsymbol{\beta}}_{S_0^c} - \boldsymbol{\beta}_{S_0}^0 - \underbrace{\boldsymbol{\beta}_{S_0^c}^0}_{=0}\|_1 \\
&= 2 \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\|_1 + \lambda \|\hat{\boldsymbol{\beta}}_{S_0^c}\|_1 \quad (\text{by lemma 5}) \\
&\leq 4\lambda \|\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\|_1 \\
&\leq 4\lambda \sqrt{\left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\right)^T \hat{\Sigma} \left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\right) s_0 / \phi_0^2} \\
&\leq \sqrt{\left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\right)^T \mathbf{X}^T \mathbf{X} \left(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0\right)} \frac{4\lambda \sqrt{s_0}}{\phi_0 \sqrt{n}} \\
&\leq \|\mathbf{X}(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0)\|_2 \frac{4\lambda \sqrt{s_0}}{\phi_0 \sqrt{n}} \\
&\leq \|\mathbf{X}(\hat{\boldsymbol{\beta}}_{S_0} - \boldsymbol{\beta}_{S_0}^0)\|_2^2 + \frac{4\lambda^2 s_0}{\phi_0^2 n}
\end{aligned}$$

Where the last inequality follows from $4uv \leq u^2 + 4v^2$. □

0.3.2 Section 6.3

Now $\mathbf{Y} := \mathbf{f}^0 + \boldsymbol{\varepsilon}$, therefore $\mathbb{E}[\mathbf{Y}] := \mathbf{f}^0$.

Lemma 0.8 (New version of the Basic Inequality). $\forall \boldsymbol{\beta}^* \in \mathbb{R}^p$ we have

$$\frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)}{n} + \lambda \|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n} \quad (7)$$

Proof. By definition of $\hat{\boldsymbol{\beta}}$, we have that

$$\forall \boldsymbol{\beta} \quad \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{n} + \lambda \|\boldsymbol{\beta}\|_1$$

In particular for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ we have

$$\forall \boldsymbol{\beta}^* \quad \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda \|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*\|_2^2}{n} + \lambda \|\boldsymbol{\beta}^*\|_1$$

We since $\mathbf{Y} = \mathbf{f}^0 + \boldsymbol{\varepsilon}$

$$\begin{aligned}
& \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*\|_2^2}{n} + \lambda\|\boldsymbol{\beta}^0\|_1 \\
\Rightarrow & \frac{\|(\mathbf{f}^0 + \boldsymbol{\varepsilon}) - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|(\mathbf{f}^0 + \boldsymbol{\varepsilon}) - \mathbf{X}\boldsymbol{\beta}^*\|_2^2}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 \\
\Rightarrow & \frac{\|(\mathbf{f}^0 - \mathbf{X}\hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{\|(\mathbf{f}^0 - \mathbf{X}\boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}\|_2^2}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 \\
\Rightarrow & \frac{\langle (\mathbf{f}^0 - \mathbf{X}\hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}, (\mathbf{f}^0 - \mathbf{X}\hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon} \rangle}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \\
& \leq \frac{\langle (\mathbf{f}^0 - \mathbf{X}\boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}, (\mathbf{f}^0 - \mathbf{X}\boldsymbol{\beta}^*) + \boldsymbol{\varepsilon} \rangle}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 \\
\Rightarrow & \frac{\|\mathbf{f}^0 - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2 + \|\boldsymbol{\varepsilon}\|_2^2 + 2\langle \mathbf{f}^0 - \mathbf{X}\hat{\boldsymbol{\beta}}, \boldsymbol{\varepsilon} \rangle}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \\
& \leq \frac{\|\mathbf{f}^0 - \mathbf{X}\boldsymbol{\beta}^*\|_2^2 + \|\boldsymbol{\varepsilon}\|_2^2 + 2\langle \mathbf{f}^0 - \mathbf{X}\boldsymbol{\beta}^*, \boldsymbol{\varepsilon} \rangle}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 \\
\Rightarrow & \frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\langle \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*), \boldsymbol{\varepsilon} \rangle}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow & \frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n}
\end{aligned}$$

□

Lemma 0.9 (New version of Lemma 6.3.). *We have on \mathcal{T} , with $\lambda \geq 4\lambda_0$,*

$$\frac{4\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + 3\lambda\|\hat{\boldsymbol{\beta}}_{S_*^c}\|_1 \leq 5\lambda\|\hat{\boldsymbol{\beta}}_{S_*} - \boldsymbol{\beta}_{S_*}^*\|_1 + \frac{4\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n} \quad (8)$$

where $S_* := \{j : \beta_j^* \neq 0\}$.

Proof. We start with the Basic Inequality

$$\frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n}$$

Now since we are on \mathcal{T} and since $4\lambda_0 \leq \lambda$

$$\begin{aligned}
& \frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)}{n} + \lambda\|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow & \frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \lambda_0\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 + \lambda\|\boldsymbol{\beta}^*\|_1 + \frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow & 4\frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{f}^0\|_2^2}{n} + 4\lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \lambda\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 + 4\lambda\|\boldsymbol{\beta}^*\|_1 + 4\frac{\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{f}^0\|_2^2}{n}
\end{aligned}$$

We use the triangle inequality on the left hand side

$$\begin{aligned}
\|\hat{\beta}\|_1 &= \|\hat{\beta}_{S_*}\|_1 + \|\hat{\beta}_{S_*^c}\|_1 \\
&= \|\beta_{S_*}^* - \beta_{S_*}^* + \hat{\beta}_{S_*}\|_1 + \|\hat{\beta}_{S_*^c}\|_1 \\
&\geq \|\beta_{S_*}^*\|_1 - \|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + \|\hat{\beta}_{S_*^c}\|_1
\end{aligned}$$

whereas on the right hand side

$$\begin{aligned}
\|\hat{\beta} - \beta^*\|_1 &= \|(\hat{\beta}_{S_*} + \hat{\beta}_{S_*^c}) - (\beta_{S_*}^* + \underbrace{\beta_{S_*^c}^*}_{=0})\|_1 \\
&= \|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + \|\hat{\beta}_{S_*^c}\|_1
\end{aligned}$$

Injecting these two results, we get that

$$\begin{aligned}
&4 \frac{\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 4\lambda\|\hat{\beta}\|_1 \leq \lambda\|\hat{\beta} - \beta^*\|_1 + 4\lambda\|\beta^*\|_1 + 4 \frac{\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow &4 \frac{\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 4\lambda \left(\|\beta_{S_*}^*\|_1 - \|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + \|\hat{\beta}_{S_*^c}\|_1 \right) \\
&\leq \lambda \left(\|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + \|\hat{\beta}_{S_*^c}\|_1 \right) + 4\lambda\|\beta^*\|_1 + 4 \frac{\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow &4 \frac{\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 4\lambda \underbrace{\|\beta_{S_*}^*\|_1}_{=\beta^*} + 3\lambda\|\hat{\beta}_{S_*^c}\|_1 \\
&\leq 5\lambda\|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + 4\lambda\|\beta^*\|_1 + 4 \frac{\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} \\
\Rightarrow &4 \frac{\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 3\lambda\|\hat{\beta}_{S_*^c}\|_1 \leq 5\lambda\|\hat{\beta}_{S_*} - \beta_{S_*}^*\|_1 + 4 \frac{\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n}
\end{aligned}$$

□

Definition 0.10 (Compatibility condition for general sets). *We say that the compatibility condition holds for the set S , if for some constant $\phi(S) > 0$, and for all β , with $\|\beta_{S^c}\|_1 \leq 3\|\beta_S\|_1$, one has*

$$\|\beta_S\|_1^2 \leq \left(\beta^T \hat{\Sigma} \beta \right) \frac{|S|}{\phi^2(S)}$$

We define \mathcal{S} as the collection of sets S for which the compatibility condition holds.

Definition 0.11 (The oracle). *We define the oracle β^* as*

$$\beta^* = \arg \min_{\beta: S_\beta \in \mathcal{S}} \left\{ \frac{\|\mathbf{X}\beta - \mathbf{f}^0\|_2^2}{n} + \frac{4\lambda^2 s_\beta}{\phi^2(S_\beta)} \right\}$$

where $S_\beta := \{j : \beta_j \neq 0\}$, $s_\beta := |S_\beta|$ denotes the cardinality of S_β and the factor 4 in the right hand side comes from choosing $\lambda \geq \lambda_0$.

Chapter 1

Introduction

Chapter 2

Classical theory of Linear Regression

Chapter 3

Theory for LASSO in high dimensions

