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

Contents

Chapter 1

Notes

1.1 Section 6.2

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \tag{1.1}$$

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

$$\hat{\beta} := \arg\min_{\beta} \left\{ \frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda\|\beta\|_1 \right\} \tag{1.2}$$

Lemma 1.1 (Basic Inequality)**.**

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

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

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

In particular for $\beta = \beta^0$ we have

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

We now replace $\mathbf{Y}$ using equation (1.1).

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

□

Let

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

Lemma 1.2 (Lemma 6.2.)**.** *For all $t > 0$ and*

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

we have

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

Proof. We define

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

Then we have

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

□

Lemma 1.3 (Lemma 6.3.). *We have on $\mathscr{T}$, with $\lambda \ge 2\lambda_0$,*

$$2\left\|\mathbf{X}\left(\hat\beta-\beta^0\right)\right\|_2^2/n + \lambda\left\|\hat\beta_{S_0^c}\right\|_1 \le 3\lambda\left\|\hat\beta_{S_0}-\beta^0_{S_0}\right\|_1$$

Proof. □

Theorem 1.4 (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 $\left\|\beta_{S_0^c}\right\|_1 \le 3\left\|\beta_{S_0}\right\|_1$, it holds that*

$$\|\beta_{S_0}\|_1^2 \le \left(\beta^T\hat\Sigma\beta\right) s_0/\phi_0^2$$

Theorem 1.5 (Theorem 6.1.). *Suppose the compatibility condition holds for S_0. Then on $\mathscr{T}$, we have for $\lambda \ge 2\lambda_0$,*

$$\left\|\mathbf{X}\left(\hat\beta-\beta^0\right)\right\|_2^2/n + \lambda\left\|\hat\beta-\beta^0\right\|_1 \le 4\lambda^2 s_0/\phi_0^2$$

Theorem 1.6 (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 $\|\beta_{S^c}\|_1 \le 3\|\beta_S\|_1$, one has*

$$\|\beta_S\|_1^2 \le \left(\beta^T\hat\Sigma\beta\right)|S|/\phi^2(S)$$

1.2 Section 6.3

Now $\mathbb{E}[\mathbf{Y}] := \mathbf{f}^0$

Lemma 1.7 (New version of the Basic Inequality). *$\forall \beta^* \in \mathbb{R}^p$ we have*

$$\frac{\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + \lambda\|\hat{\beta}\|_1 \leq \frac{\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} + \lambda\|\beta^*\|_1 + \frac{2\boldsymbol{\varepsilon}^T\mathbf{X}(\hat{\beta} - \beta^*)}{n} \tag{1.3}$$

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

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

In particular for $\beta = \beta^*$ we have

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

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

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

□

Lemma 1.8 (New version of the Lemma 6.3.).

$$\frac{4\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 3\lambda\|\hat{\beta}_{S_*^c}\|_1 \leq 5\lambda\|\hat{\beta}_{S_*} - \beta^*_{S_*}\|_1 + \frac{4\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} \tag{1.4}$$

Chapter 2

Introduction

2.1 The first section

2.1.1 The first subsection

To be added

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