



□ 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

1	Notes	1
2	Introduction	5
2.1	The first section	5
2.1.1	The first subsection	5

Chapter 1

Notes

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad (1.1)$$

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

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

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).

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

□

Let

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

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

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

Proof. We define

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

Then we have

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

□

Theorem 1.3 (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*

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

Theorem 1.4 (Theorem 6.1.). *Suppose the compatibility condition holds for S_0 . Then on \mathcal{T} , we have for $\lambda \geq 2\lambda_0$,*

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

Theorem 1.5 (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*

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

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