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 Introduction **1**
1.1 The first section 1
1.1.1 The first subsection 1

Chapter 1

Introduction

1.1 The first section

1.1.1 The first subsection

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

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).

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

□

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

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

□

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 $\left\|\beta_{S_0^c}\right\|_1 \leq 3\left\|\beta_{S_0}\right\|_1$, it holds that*

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

Theorem 1.4 (Theorem 6.1.)**.** *Suppose the compatibility condition holds for S_0. Then on $\mathscr{T}$, 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 $\left\|\beta_{S^c}\right\|_1 \leq 3\left\|\beta_S\right\|_1$, one has*

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

To be added

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