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

Chapter 1

Introduction

1.1 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

Chapter 2

Classical theory of Linear Regression

Chapter 3

Theory for LASSO in high dimensions

3.1 Section 6.2

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \tag{3.1}$$

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

$$\hat{\beta} := \arg\min_{\beta} \left\{ \frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda\|\beta\|_1 \right\} \tag{3.2}$$

Lemma 3.1 (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 (3.1).

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

□

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 3.2 (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

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

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

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

Let Z be a standard normal random variable and $x>0$. Then

$$P(Z>x) = 1/\sqrt{2\pi}\int_x^{\infty} exp(-y^2/2)dy < 1/(x*\sqrt{2\pi})\int_x^{\infty} y*exp(-y^2/2)dy$$
$$= 1/(x\sqrt{2\pi})exp(-x^2/2)$$

Since in our case $x\sqrt{2\pi} > 1$ (if p is at least =2), we finally get $P(Z > x) < exp(-x^2/2)$ Since the normal distribution is symmetric we also have $P(|Z| > x) < 2exp(-x^2/2)$ □

Corollary 3.3 (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

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

Lemma 3.4 (Lemma 6.3.). *We have on $\mathscr{T}$, with $\lambda \geq 2\lambda_0$,*

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

Proof. We start with the Basic Inequality

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

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

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

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

Let $\beta_{j,S} := \beta_j 1\{j \in S\}$. We use the triangle inequality on the left hand side

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

whereas on the right hand side

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

Injecting these two results, we get that

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

□

Definition 3.5 (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*

$$\|\beta_{S_0}\|_1^2 \leq \left(\beta^T\hat{\Sigma}\beta\right) s_0/\phi_0^2 \tag{3.3}$$

Theorem 3.6 (Theorem 6.1.)**.** *Suppose the compatibility condition holds for S_0. Then on $\mathscr{T}$, 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 3.4 we have that

$$
\begin{aligned}
&2\frac{\|\mathbf{X}(\hat\beta-\beta^0)\|_2^2}{n}+\lambda\|\hat\beta-\beta^0\|_1\\
&=2\frac{\|\mathbf{X}(\hat\beta-\beta^0)\|_2^2}{n}+\lambda\|\hat\beta_{S_0}+\hat\beta_{S_0^c}-\beta^0_{S_0}-\underbrace{\beta^0_{S_0^c}}_{=0}\|_1\\
&=2\frac{\|\mathbf{X}(\hat\beta-\beta^0)\|_2^2}{n}+\lambda\|\hat\beta_{S_0}-\beta^0_{S_0}\|_1+\lambda\|\hat\beta_{S_0^c}\|_1 \quad \textit{(by lemma 3.4)}\\
&\le 4\lambda\|\hat\beta_{S_0}-\beta^0_{S_0}\|_1\\
&\le 4\lambda\sqrt{\left(\hat\beta_{S_0}-\beta^0_{S_0}\right)^T\hat\Sigma\left(\hat\beta_{S_0}-\beta^0_{S_0}\right)s_0/\phi_0^2}\\
&\le \sqrt{\left(\hat\beta_{S_0}-\beta^0_{S_0}\right)^T\mathbf{X}^T\mathbf{X}\left(\hat\beta_{S_0}-\beta^0_{S_0}\right)}\frac{4\lambda\sqrt{s_0}}{\phi_0\sqrt{n}}\\
&\le \|\mathbf{X}(\hat\beta_{S_0}-\beta^0_{S_0})\|_2\frac{4\lambda\sqrt{s_0}}{\phi_0\sqrt{n}}\\
&\le \|\mathbf{X}(\hat\beta_{S_0}-\beta^0_{S_0})\|_2^2+\frac{4\lambda^2 s_0}{\phi_0^2 n}
\end{aligned}
$$

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

3.2 Section 6.3

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

Lemma 3.7 (New version of the Basic Inequality). *$\forall\beta^*\in\mathbb{R}^p$ we have*

$$\frac{\|\mathbf{X}\hat\beta-\mathbf{f}^0\|_2^2}{n}+\lambda\|\hat\beta\|_1\le\frac{2\boldsymbol{\varepsilon}^T\mathbf{X}(\hat\beta-\beta^*)}{n}+\lambda\|\beta^*\|_1+\frac{\|\mathbf{X}\beta^*-\mathbf{f}^0\|_2^2}{n} \tag{3.4}$$

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

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

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

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

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

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

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

□

Lemma 3.8 (New version of the Lemma 6.3.). *We have on $\mathscr{T}$, with $\lambda \geq 4\lambda_0$,*

$$
\frac{4\|\mathbf{X}\hat{\beta} - \mathbf{f}^0\|_2^2}{n} + 3\lambda\|\hat{\beta}_{S_*^c}\|_1 \leq 5\lambda\|\hat{\beta}_{S_*} - \beta^*_{S_*}\|_1 + \frac{4\|\mathbf{X}\beta^* - \mathbf{f}^0\|_2^2}{n} \tag{3.5}
$$

Proof. We start with the Basic Inequality

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

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

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

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

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

□

Definition 3.9 (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*

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