Proof of Proposition 1:

$Z_j = \hat{\beta}_{OLS,j}$

$$\|Y - X\beta\|_2^2/n + \lambda\|\beta\|_1$$

$$= \|Y\|_2^2/n + \underbrace{\sum_{j=1}^{p}\left\{-2\underbrace{\frac{(X^TY)_j}{n}}_{=Z_j}\beta_j + \beta_j^2 + \lambda|\beta_j|\right\}}$$

decouples: minimize w.r.t. β_j, $j = 1, \ldots, p$

$$(*) = -2Z_j\beta_j + \beta_j^2 + \lambda|\beta_j| \overset{!}{=} \min_{\beta_j}$$

if $\beta_j = 0 \Rightarrow (*) = 0$

if $|Z_j| \leq \lambda/2$: $\lambda|\beta_j| - 2Z_j\beta_j \geq 0$

$\Rightarrow \hat{\beta}_j = 0$

if $Z_j > \lambda/2$: $\hat{\beta}_j > 0$

$\frac{\partial}{\partial \beta_j}$: $2\beta_j - 2Z_j + \lambda \stackrel{!}{=} 0$

$\Rightarrow \hat{\beta}_j = Z_j - \lambda/2$

analogously for $Z_j < -\lambda/2$

$\blacksquare$

Proof of Corollary 6.1

Basic Inequality (Lemma 6.1)

$$\|X(\hat\beta-\beta^0)\|_2^2/n + \lambda\|\hat\beta\|_1 \le 2\varepsilon^T X(\hat\beta-\beta^0)/n + \lambda\|\beta^0\|_1$$

Proof:

$$\underbrace{\|Y-X\hat\beta\|_2^2/n}_{\|X(\hat\beta-\beta^0)\|_2^2/n - 2\varepsilon^T X(\hat\beta-\beta^0)/n + \cancel{\|\varepsilon\|_2^2/n}} + \lambda\|\hat\beta\|_1 \le \overbrace{\|Y-X\beta^0\|_2^2/n}^{\cancel{\|\varepsilon\|_2^2/n}} + \lambda\|\beta^0\|_1$$

$\blacksquare$

term in Basic Inequality:

$$|2\,\varepsilon^T X(\hat\beta-\beta^0)/n| = 2\left|\sum_{i=1}^n \varepsilon_i \sum_{j=1}^p X_i^{(j)}(\hat\beta_j-\beta_j^0)\right|/n$$

Hölder

$$\leq 2\max_{j=1,\dots,p}\left|\sum_{i=1}^n \varepsilon_i X_i^{(j)}\right|/n \cdot \|\hat\beta-\beta^0\|_1$$

trick: decouple into probabilistic and analytical part

probabilistic part: consider

$$\mathcal{T} = \left\{2\max_{j=1,\dots,p}|\varepsilon^T X^{(j)}|/n \leq \lambda_0\right\}, \quad \underline{\lambda \geq 2\lambda_0}$$

"noise level"

we will show that $\mathbb{P}[\mathcal{T}]$ is large for appropriate λ_0

Then, on $\mathcal{T}$:

$$\|X(\hat{\beta}-\beta^0)\|_2^2/n + \lambda\|\hat{\beta}\|_1 \overset{\text{BI + Hölder}}{\leq} \lambda_0 \underbrace{\|\hat{\beta}-\beta^0\|_1}_{\leq \|\hat{\beta}\|_1 + \|\beta^0\|_1} + \lambda\|\beta^0\|_1$$

→ on $\mathcal{T}$:

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

$$\leq (\lambda_0+\lambda)\|\beta^0\|_1 \overset{\lambda_0 \leq \lambda/2}{\leq} \frac{3}{2}\lambda\|\beta^0\|_1$$

Probabilistic part

Lemma 6.2 in book:

Assume $\hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(j)})^2 \equiv 1$ $\boxed{\text{errors Gaussian}}$

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

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

Proof. Consider

$$V_j = \varepsilon^T X^{(j)} / \sqrt{n\sigma^2} \quad , \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$$

$\longrightarrow V_j \sim \mathcal{N}(0, 1) \quad \forall\, j = 1 \to p$

BUT dependent

$$\mathbb{P}\left[\max_{j=1\to p} |V_j| > \sqrt{t^2 + 2\log(p)}\right]$$

$$\underset{\text{union bound}}{\le} \sum_{j=1}^{p} \mathbb{P}\left[\underbrace{|V_j|}_{\sim \mathcal{N}(0,1)} > \sqrt{t^2 + 2\log(p)}\right]$$

union bound

Bonferroni bound

$$\underset{\text{tail bound for } \mathcal{N}(0,1)}{\le} p \cdot 2\exp\left(-\frac{t^2 + 2\log(p)}{2}\right) = 2\exp\left(-t^2/2\right)$$

tail bound for $\mathcal{N}(0,1)$

$$\mathbb{P}[\mathcal{T}] = \mathbb{P}\left[\max_j 2|\varepsilon^T X^{(j)}|/n \leq \lambda_0\right]$$

$$= \mathbb{P}\left[\max_j \underbrace{|\varepsilon^T X^{(j)}|/\sqrt{n\sigma^2}}_{V_j} \leq \frac{\lambda_0}{2} \cdot \sqrt{n}/\sigma\right]$$

$$\geq 1 - 2\exp(-t^2/2)$$