Corollary 6.1 in Bühlmann and van de Geer (2011)

Corollary 6.1

assume:

- $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$
- scaled columns $\hat{\sigma}_j^2 \equiv 1 \ \forall \ j$

For

$$\lambda = 4\hat{\sigma}\sqrt{\frac{t^2 + 2\log(p)}{n}}$$

where $\hat{\sigma}$ is an estimator for σ. Then, with probability at least $1 - \alpha$ where

$$\alpha = 2\exp(-t^2/2) + \mathbb{P}[\hat{\sigma} < \sigma]$$

we have that

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

Implications

Corollary 6.1 implies:

$$\|X(\hat{\beta} - \beta^0)\|_2^2/n = O_P(\underbrace{\lambda}_{\asymp\sqrt{\log(p)/n}} \|\beta^0\|_1) = O_P(\sqrt{\log(p)/n}\|\beta^0\|_1)$$

even for very sparse case with $\|\beta^0\|_1 = O(1)$:
slow convergence rate of order $O_P(\sqrt{\log(p)/n})$

benchmark: OLS orcale on the variables from $S_0 = \{j;\ \beta_j^0 \neq 0\}$

$$\|X(\hat{\beta}_{\text{OLS}-\text{oracle}} - \beta^0)\|_2^2/n = O_P(s_0/n),\ \ s_0 = |S_0|$$

we will later derive for the Lasso, under additional assumptions on X: fast convergence rate

$$\|X(\hat{\beta} - \beta^0)\|_2^2/n = O_P(\log(p)\frac{s_0}{n})$$

for slow rate: no assumptions on X (could have perfectly correlated columns)

Extensions

the proof technique decouples into a deterministic and probablistic part (the set $\mathcal{T}$)

the deterministic part remains the same for other probabilistic structures (other analysis for $\mathbb{P}[\mathcal{T}]$) such as:

- heteroscedastic errors with
  $\mathbb{E}[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_i^2 \not\equiv$ const.
- dependent observations $\rightsquigarrow$ for fixed design, dependent errors
- non-Gaussian errors
  sub-Gaussian distribution
  second moments plus bounded X: see Example 14.3 in Bühlmann and van de Geer (2011)
- random design: assume that ε is independent of X
  $\rightsquigarrow$ condition on X: invoke the results for fixed design and integrate out

heteroscedastic errors

$\varepsilon \sim \mathcal{N}_n(0, D)$, where $D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$
assume that: $\sigma_i^2 \le \underbrace{\sigma^2}_{\text{some pos. const.}} < \infty$

Then, Coroallry 6.1 remains true with σ^2 as above

Proof:
exactly as before but exploiting that $V_j \sim \mathcal{N}(0, \tau_j^2)$ with $\tau_j \le 1$
and using that $\mathbb{P}[|V_j| > c] \le \mathbb{P}[\underbrace{|Z|}_{\sim|\mathcal{N}(0,1)|} > c]$

Exercise: work out the details.

errors from stationary distribution

$\varepsilon \sim \mathcal{N}_n(0, \Gamma)$, where $\Gamma_{i,j} = R(i-j) = R(j-i)$
assume that: $\sum_{k=-\infty}^{\infty} |R(k)| < \infty$ and $|X_i^{(j)}| \leq K_X < \infty$

Then, Corollary 6.1 remains true with $\sigma^2 = K_X^2 \sum_{k=-\infty}^{\infty} |R(k)|$

Proof:
Exercise. (A bit more tricky...)