Outline of the notes

- Proof of the risk bound for the hard thresholding
- Proof of Lemma 1
- Proof of the fast rate for the Dantzig selector
- Proof of the slow rate for the Dantzig selector
- Poof of the fast rate for the Lasso under the RE property

1 Proof of the risk bound for the hard thresholding

$$Z = \beta^* + \frac{\sigma^*}{\sqrt{n}} \varepsilon \qquad \text{with} \quad \varepsilon \sim \mathcal{N}_p(0, I)$$

$$\hat{\beta}_j = Z_j \, \mathbb{1}\left(|Z_j| > \frac{\lambda \sigma^*}{\sqrt{n}}\right)$$

1st case: $j \in J = \{ i : \beta^*_i \neq 0 \}$

$$\begin{aligned}
(\hat{\beta}_j - \beta^*_j)^2 &= (Z_j - \beta^*_j)^2 \mathbb{1}\left(|Z_j| > \frac{\lambda\sigma^*}{\sqrt{n}}\right) + \beta_j^{*2} \mathbb{1}\left(|Z_j| \leq \frac{\lambda\sigma^*}{\sqrt{n}}\right) \\
&= \frac{\sigma^{*2}}{n} \varepsilon_j^2 \mathbb{1}\left(|Z_j| > \frac{\lambda\sigma^*}{\sqrt{n}}\right) + \left(Z_j - \frac{\sigma^*}{\sqrt{n}}\varepsilon_j\right)^2 \mathbb{1}\left(|Z_j| \leq \frac{\lambda\sigma^*}{\sqrt{n}}\right) \\
&\leq \frac{\sigma^{*2}}{n} \varepsilon_j^2 \mathbb{1}\left(|Z_j| > \frac{\lambda\sigma^*}{\sqrt{n}}\right) + \left(2Z_j^2 + \frac{2\sigma^{*2}}{n}\varepsilon_j^2\right) \mathbb{1}\left(|Z_j| \leq \frac{\lambda\sigma^*}{\sqrt{n}}\right) \\
&\leq \frac{2\sigma^{*2}}{n} \varepsilon_j^2 + 2Z_j^2 \mathbb{1}\left(|Z_j| \leq \frac{\lambda\sigma^*}{\sqrt{n}}\right) \\
&\leq \frac{2\sigma^{*2}}{n} \left(\varepsilon_j^2 + \lambda^2\right)
\end{aligned}$$

Taking the expectation on both sides of this inequality, we get

$$\begin{aligned}
\mathbb{E}\left[(\hat{\beta}_j - \beta^*_j)^2\right] &\leq \frac{2\sigma^{*2}}{n}(1 + \lambda^2) \leq \frac{2\sigma^{*2}}{n}\left(\frac{\lambda^2}{2} + \lambda^2\right) \\
&\leq \frac{3\sigma^2}{n} \lambda^2
\end{aligned}$$

2nd case : $j \notin J \Leftrightarrow \beta_j^* = 0$

$$(\hat\beta_j - \beta_j^*)^2 = \hat\beta_j^2 = Z_j^2 \mathbb{1}\left(|Z_j| > \frac{\lambda\sigma^*}{\sqrt{n}}\right)$$
$$= \frac{\sigma^{*2}}{n} \varepsilon_j^2 \mathbb{1}(|\varepsilon_j| > \lambda)$$

Taking the expectation, we get

$$\mathbb{E}\left[(\hat\beta_j - \beta_j^*)^2\right] = \frac{\sigma^{*2}}{n} \times \frac{2}{\sqrt{2\pi}} \int_\lambda^{+\infty} t^2 e^{-t^2/2} dt$$
$$\leqslant \frac{\sigma^{*2}}{n} \int_\lambda^{+\infty} t^2 e^{-t^2/2} dt$$
$$\leqslant \frac{\sigma^{*2}}{n} \int_\lambda^{+\infty} 2(t^2-1) e^{-t^2/2} dt$$
$$= \frac{2\sigma^{*2}}{n} \left[-t e^{-t^2/2}\right]_\lambda^{+\infty} = \frac{2\lambda\sigma^{*2}}{n} e^{-\lambda^2/2}$$

Combining these two cases, we get

$$\mathbb{E}\left[\|\hat\beta - \beta^*\|_2^2\right] = \sum_{j\in J} \mathbb{E}\left[(\hat\beta_j - \beta_j^*)^2\right] + \sum_{j \in J^c} \mathbb{E}\left[(\hat\beta_j - \beta_j^*)^2\right]$$
$$\leqslant \frac{3|J|\sigma^{*2}}{n}\lambda^2 + \frac{2|J^c|\sigma^{*2}}{n}\lambda e^{-\lambda^2/2}.$$

To complete the proof, it suffices to notice that $|J^c| \leqslant p$.

Proof of Lemma 1

1 Let $\mathcal{B} = \left\{ \|X^T(Y - X\beta^*)\|_\infty \leqslant \lambda\sigma^* \right\}$
$= \left\{ |X_j^T \xi| \leqslant \lambda \ \forall j \in \{1,\dots,p\} \right\}$

where X_j denotes the j^{th} column of the matrix X.

One easily checks that since $\xi \sim \mathcal{N}_n(0, I_n)$ and $\|X_j\|_2^2 = n$, the random variables $\frac{1}{\sqrt{n}} X_j^T \xi$ are Gaussian $\mathcal{N}(0,1)$. Therefore, using the union bound, we find

$$\mathbb{P}(\mathcal{B}^c) \leqslant \sum_{j=1}^{p} \mathbb{P}\left(\frac{1}{\sqrt{n}}\,|X_j^T\xi| > \frac{\lambda}{\sqrt{n}}\right)$$

$$\leqslant 2p \times \exp\left\{-\frac{\lambda^2}{2n}\right\} \qquad \text{(we use here that } 1-\Phi(x) \leqslant e^{-x^2/2}\text{)}$$

$$= \alpha \quad \text{if } \lambda = \sqrt{2n\log(2p/\alpha)}$$

(2) We want to prove that

$$\frac{1}{n}|\langle Xu, Xv\rangle| \leqslant \delta_{2s}\,\|u\|_2\cdot\|v\|_2 .$$

This is equivalent to proving

$$\frac{1}{n}\left|\left\langle X\frac{u}{\|u\|}, X\frac{v}{\|v\|}\right\rangle\right| \leqslant \delta_{2s}$$

or, equivalently, that

$$(1) \qquad \frac{1}{n}|\langle Xu, Xv\rangle| \leqslant \delta_{2s} \qquad \forall u,v \text{ s.t. } \|u\|_2 = \|v\|_2 = 1 \quad u \perp v$$
$$\|u\|_0 \leqslant s \quad \|v\|_2 \leqslant s.$$

To prove (1), we use the paralelogramme rule:

$$\frac{1}{n}\langle Xu, Xv\rangle = \frac{1}{4n}\left(\|X(u+v)\|_2^2 - \|X(u-v)\|_2^2\right)$$

$$\overset{RIP(2s)}{\leqslant} \frac{n}{4n}\left((1+\delta_{2s})\underbrace{\|u+v\|_2^2}_{=2} - (1-\delta_{2s})\underbrace{\|u-v\|_2^2}_{=2}\right)$$

$$= \delta_{2s}$$

■

(3) We have, since $\beta_j^* = 0$ for every $j \in J^c$,

$$\|h_{J^c}\|_1 = \sum_{j\in J^c}|\hat\beta_j - \beta_j^*| = \sum_{j\in J^c}|\hat\beta_j| = \|\hat\beta\|_1 - \|\hat\beta_J\|_1$$

Now, on the event $\mathcal{B}$, both $\hat\beta$ and β^* satisfy the constraints of the DS, therefore, $\|\hat\beta\|_1 \leqslant \|\beta^*\|_1$.

This leads to

$$\|h_{J^c}\|_1 \leqslant \|\beta^*\|_1 - \|\hat\beta_J\|_1 = \|\beta_J^*\|_1 - \|\hat\beta_J\|_1 \leqslant \|(\beta^* - \hat\beta)_J\|_1$$

This completes the proof. ■

④ According to the triangle inequality,

$$\| h_{(J\cup J_1)^c} \|_2 = \| \sum_{j\geq 2} h_{J_k} \|_2 \leq \sum_{j\geq 2} \| h_{J_k} \|_2$$

Let now fix a $k \geq 2$. By definition of J_k, we have

$$|h_j| \leq \min_{\ell \in J_{k-1}} |h_\ell| \qquad \forall j \in J_k.$$

Therefore,

$$\| h_{J_k} \|_2 = \sqrt{\sum_{j\in J_k} |h_j|^2} \leq \sqrt{s} \cdot \min_{\ell \in J_{k-1}} |h_\ell| \leq \frac{1}{\sqrt{s}} \sum_{\ell \in J_{k-1}} |h_\ell| .$$

Thus, $\| h_{J_k} \|_2 \leq \frac{1}{\sqrt{s}} \| h_{J_{k-1}} \|_1$

and summing up these inequalities for all $k \geq 2$, we get

$$\sum_{k\geq 2} \| h_{J_k} \|_2 \leq \sum_{k\geq 1} \| h_{J_{k-1}} \|_1 / \sqrt{s} \leq \frac{1}{\sqrt{s}} \| h_{J^c} \|_1 .$$

In view of the Cauchy-Schwarz inequality, and claim ③,

$$\frac{1}{\sqrt{s}} \| h_{J^c} \|_1 \leq \frac{1}{\sqrt{s}} \| h_J \|_1 \leq \| h_J \|_2$$

This completes the proof of the Lemma ∎

PROOF OF FAST RATES FOR DS

Assume that we are on the event $\mathcal{B}$ and that $h = \hat{\beta}^{DS} - \beta^*$.
It is clear that

$$\| \hat{\beta}^{DS} - \beta^* \|_2^2 = \| h_{J\cup J_1} + h_{(J\cup J_1)^c} \|_2^2 \leq \| h_{J\cup J_1} \|_2^2 + \| h_{(J\cup J_1)^c} \|_2^2$$

According to the 4th claim of the lemma, we have

$$\| \hat{\beta}^{DS} - \beta^* \|_2^2 \leq \| h_{J\cup J_1} \|_2^2 + \| h_J \|_2^2 \leq 2 \| h_{J\cup J_1} \|_2^2 \qquad \text{(D.1)}$$

Since $h_{J\cup J_1}$ is $2s$-sparse and X satisfies RIP($2s$),

$$\| h_{J\cup J_1} \|_2^2 \leq (1 - \delta_{2s})^{-1} \cdot \frac{1}{n} \| X h_{J\cup J_1} \|_2 \qquad \text{(D.2)}$$

The last norm can be boundes as follows:

$$\|Xh_{J\cup J_1}\|_2 = \langle Xh_{J\cup J_1}, Xh_{J\cup J_1}\rangle$$

$$= \langle Xh_{J\cup J_1}, Xh\rangle - \langle Xh_{J\cup J_1}, Xh_{(J\cup J_1)^c}\rangle$$

$$\leq \|h_{J\cup J_1}\|_1 \cdot \|X^T X h\|_\infty + \sum_{k\geq 2} |\langle Xh_J, Xh_{J_k}\rangle| + |\langle Xh_{J_1}, Xh_{J_k}\rangle|$$

$$\leq \sqrt{2s}\, \|h_{J\cup J_1}\|_2 \cdot \left(\|X^T(Y - X\hat{\beta}^{DS})\|_\infty + \|X^T(Y - X\beta^*)\|_\infty\right)$$

Using ②

$$+ \sum_{k\geq 2} \delta_{2s}\, n \left(\|h_J\|_2 \cdot \|h_{J_k}\|_2 + \|h_{J_1}\|_2 \cdot \|h_{J_k}\|_2\right)$$

Using ①

$$\leq \sqrt{2s}\, \|h_{J\cup J_1}\|_2 \cdot \left(\sqrt{2n\log(2p/\alpha)}\,\sigma^* + \sqrt{2n\log(2p/\alpha)}\,\sigma^*\right)$$

Using ④

$$+ n\delta_{2s}\left(\|h_J\|_2 + \|h_{J_1}\|_2\right)\|h_J\|_2$$

$$\leq 4\sigma^* \sqrt{sn\log(2p/\alpha)}\, \|h_{J\cup J_1}\|_2 + \sqrt{2}\, n\delta_{2s} \|h_{J\cup J_1}\|_2^2$$

Combining this inequality with (D.2), we get

$$\|h_{J\cup J_1}\|_2 \leq \frac{4\sigma^* \sqrt{s\log(2p/\alpha)}}{(1-\delta_{2s})\sqrt{n}} + \frac{\sqrt{2}\,\delta_{2s}}{1-\delta_{2s}} \|h_{J\cup J_1}\|_2$$

which yields

$$\|h_{J\cup J_1}\|_2 \leq \frac{4\sigma^* \sqrt{s\log(2p/\alpha)}}{(1-\delta_{2s}-\sqrt{2}\,\delta_{2s})\sqrt{n}} \qquad (D.3)$$

Combining (D.3) and (D.1), we get the desired result.

SLOW RATES FOR THE DS

Let us prove that the DS is consistent w.r.t. the prediction loss without any assumption on the matrix X. Indeed,

$$\begin{aligned}
\frac{1}{n}\|X(\hat\beta^{DS}-\beta^*)\|_2^2 &= \frac{1}{n} h^T X^T X h \\
&\leq \frac{1}{n}\|h\|_1 \cdot \|X^T X h\|_\infty \\
&\leq \frac{1}{n}\|h\|_1 \left(\|X^T(Y-X\hat\beta^{DS})\|_\infty + \|X^T(Y-X\beta^*)\|_\infty\right) \\
&\leq 2\sqrt{\frac{\log(2p/\alpha)}{n}}\,\sigma^* \cdot \|\hat\beta^{DS}-\beta^*\|_1
\end{aligned}$$

This last norm can be bounded in 2 ways. First, since on B β^* satisfies the constraints of the DS, we have $\|\hat\beta^{DS}\|_1 \leq \|\beta^*\|_1$. Therefore, using the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\frac{1}{n}\|X(\hat\beta^{DS}-\beta^*)\|_2^2 &\leq 4\sqrt{\frac{\log(2p/\alpha)}{n}}\,\sigma^*\|\beta^*\|_1 \\
&\leq 4\sqrt{\frac{s\log(2p/\alpha)}{n}}\,\sigma^*\|\beta^*\|_2 .
\end{aligned}$$

One can also use the claim 3 of the lemma, to bound $\|(\hat\beta^{DS}-\beta^*)\|_1 = \|h\|_1$ from above by $2\|h_J\|_1 \leq 2\sqrt{s}\|h_J\|_2$ ∎

FAST RATES FOR THE LASSO UNDER RE

Without loss of generality, we assume that $\sigma^*=1$. Since $\hat\beta = \hat\beta^{Lasso}$ minimizes the penalized log-likelihood, we have

$$\underbrace{\|X\hat\beta - Y\|_2^2}_{X(\hat\beta-\beta^*)-\xi} + 2\lambda\|\hat\beta\|_1 \leq \underbrace{\|X\beta^*-Y\|_2^2}_{\xi} + 2\lambda\|\beta^*\|_1$$

$$\|X(\hat\beta-\beta^*)\|_2^2 \leq 2\lambda\left(\|\beta^*\|_1 - \|\hat\beta\|_1\right) + 2\xi^T X(\hat\beta-\beta^*).$$

Using the duality of the l_1 and l_∞ norms, we get

$$\|X(\hat{\beta}-\beta^*)\|_2^2 \leq 2\lambda\left(\|\beta^*\|_1 - \|\hat{\beta}\|_1\right) + 2\|X^T\xi\|_\infty \cdot \|\hat{\beta}-\beta^*\|_1.$$

Repeating the arguments of the proof of the 1^{st} claim of Lemma, we get that on the event $\mathcal{B}$ (with prob. $\geq 1-\alpha$) $\|X^T\xi\|_\infty \leq \lambda/2$.

Therefore,

$$\|X(\hat{\beta}-\beta^*)\|_2^2 \leq 2\lambda\left(\|\beta^*\|_1 - \|\hat{\beta}\|_1\right) + \lambda\|\hat{\beta}-\beta^*\|_1. \qquad (L.1)$$

Since the left-hand side above is ≥ 0, we get

$$\|\hat{\beta}-\beta^*\|_1 \leq 2\|\beta^*\|_1 - 2\|\hat{\beta}\|_1$$

Using the fact that $\beta^*_{J^c} = 0$, this yields

$$\begin{aligned}\|(\hat{\beta}-\beta^*)_J\|_1 + \|\hat{\beta}_{J^c}\|_1 &\leq 2\|\beta^*_J\|_1 - 2\|\hat{\beta}_J\|_1 - 2\|\hat{\beta}_{J^c}\|_1 \\ &\leq 2\|(\beta^*-\hat{\beta})_J\|_1 - 2\|\hat{\beta}_{J^c}\|_1 \quad \text{(triangle ineq.)}\end{aligned}$$

Thus, $\|\hat{\beta}_{J^c}\|_1 \leq 3\|(\hat{\beta}-\beta^*)_J\|_1 \iff \|(\hat{\beta}-\beta^*)_{J^c}\|_1 \leq 3\|(\hat{\beta}-\beta^*)_J\|_1$.

This means that we can apply to $\hat{\beta}-\beta^*$ the $\widetilde{RE}(s,3)$:

$$\|\hat{\beta}-\beta^*\|_2^2 \leq \frac{1}{n\varkappa^2}\|X(\hat{\beta}-\beta^*)\|_2^2 \qquad (L.2)$$

Using once again (L.1), we get

$$\begin{aligned}\|X(\hat{\beta}-\beta^*)\|_2^2 &\leq 2\lambda\left(\|\beta^*\|_1 - \|\hat{\beta}\|_1 + \|\hat{\beta}-\beta^*\|_1\right) \\ &= 2\lambda\left(\|\beta^*_J\|_1 - \|\hat{\beta}_J\|_1 + \|(\hat{\beta}-\beta^*)_J\|_1\right) \\ &\leq 4\lambda\|(\hat{\beta}-\beta^*)_J\|_1 \quad \text{(triangle inequality)} \\ &\leq 4\lambda\sqrt{s}\,\|(\hat{\beta}-\beta^*)_J\|_2 \quad \text{(Cauchy-Schwarz)} \\ &\leq \frac{4\lambda}{\varkappa}\sqrt{\frac{s}{n}}\,\|X(\hat{\beta}-\beta^*)\|_2 \quad \text{(by (L.2))}\end{aligned}$$

This implies that $\|X(\hat{\beta}-\beta^*)\|_2 \leq \frac{4\lambda}{\varkappa}\sqrt{\frac{s}{n}}$. Replacing this in (L.2), we get the desired result. $\blacksquare$