

In this lecture note, we consider such a linear model  $y = \mathbf{X}\theta^* + w$ , and mainly study the consistency of the Lagrangian Lasso solution:

$$\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2n} \|y - \mathbf{X}\theta\|_2^2 + \lambda_n \|\theta\|_1 \right\}. \quad (4.1)$$

**Extension of restricted nullspace.** Define the set

$$\mathbb{C}_\alpha(S) := \{\Delta \in \mathbb{R} : \|\Delta_{S^c}\|_1 \leq \alpha \|\Delta_S\|_1\}.$$

This definition generalizes the set  $\mathbb{C}(S)$  used in our definition of the restricted nullspace property, which corresponds to the special case  $\alpha = 1$ .

**RE condition.** The matrix  $\mathbf{X}$  satisfies the restricted eigenvalue (RE) condition over  $S$  with parameters  $(k, \alpha)$  if

$$\frac{1}{n} \|\mathbf{X}\Delta\|_2^2 \geq k \|\Delta\|_2^2 \text{ for all } \Delta \in \mathbb{C}_\alpha(S). \quad (4.2)$$

**Assumption 4.1** In the case of hard sparse, we assume that

(A<sub>1</sub>) The vector  $\theta^*$  is supported on a subset  $S \subset \{1, \dots, d\}$  with  $|S| = s$ .

(A<sub>2</sub>) The design matrix  $\mathbf{X}$  satisfies the RE condition with parameter  $(k, 3)$ :

$$\frac{1}{n} \|\mathbf{X}\Delta\|_2^2 \geq k \|\Delta\|_2^2 \text{ for all } \Delta \in \mathbb{C}_3(S).$$

**Theorem 4.2 (Theorem 7.13 in Wainwright's book)** Under above assumption, for any solution  $\hat{\theta}$  of the Lagrangian Lasso (4.1) with  $\lambda_n \geq 2 \left\| \frac{\mathbf{X}^T w}{n} \right\|_\infty$ , we have

$$\|\hat{\theta} - \theta^*\|_2^2 \leq \frac{3}{k} \sqrt{s} \lambda_n.$$

**Proof:** Our first step is to show that, under the condition  $\lambda_n \geq 2 \left\| \frac{\mathbf{X}^T w}{n} \right\|_\infty$ , the error vector  $\hat{\Delta} = \hat{\theta} - \theta^*$  belongs to  $\mathbb{C}_3(S)$ .

Define the Lagrangian  $L(\theta; \lambda_n) = \frac{1}{2n} \|y - \mathbf{X}\theta\|_2^2 + \lambda_n \|\theta\|_1$ . Since  $\hat{\theta}$  is optimal, we have

$$L(\hat{\theta}; \lambda_n) \leq L(\theta^*; \lambda_n).$$

That is,

$$\begin{aligned} & \frac{1}{2n} \|y - \mathbf{X}\hat{\theta}\|_2^2 + \lambda_n \|\hat{\theta}\|_1 \leq \frac{1}{2n} \|y - \mathbf{X}\theta^*\|_2^2 + \lambda_n \|\theta^*\|_1 \\ \iff & \frac{1}{2n} \|\mathbf{X}\theta^* + w - \mathbf{X}\hat{\theta}\|_2^2 + \lambda_n \|\hat{\theta}\|_1 \leq \frac{1}{2n} \|w\|_2^2 + \lambda_n \|\theta^*\|_1 \\ \iff & \frac{1}{2n} (\mathbf{X}\hat{\Delta} - w)^T (\mathbf{X}\hat{\Delta} - w) + \lambda_n \|\hat{\theta}\|_1 \leq \frac{1}{2n} w^T w + \lambda_n \|\theta^*\|_1. \end{aligned}$$

Rearranging it, we have

$$0 \leq \frac{1}{2n} \|\mathbf{X}\hat{\Delta}\|_2^2 \leq \frac{w^T \mathbf{X}\hat{\Delta}}{n} + \lambda_n \left\{ \|\theta^*\|_1 - \|\hat{\theta}\|_1 \right\}.$$

Now since  $\theta^*$  is  $S$ -sparse, we can write

$$\|\theta^*\|_1 - \|\hat{\theta}\|_1 = \|\theta_S^*\|_1 - \left\| \theta_S^* + \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_S \right\|_1.$$

Thus,

$$\begin{aligned} 0 \leq \frac{1}{n} \|\mathbf{X}\hat{\Delta}\|_2^2 &\leq 2 \frac{w^T \mathbf{X}\hat{\Delta}}{n} + 2\lambda_n \left\{ \|\theta_S^*\|_1 - \left\| \theta_S^* + \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_{S^c} \right\|_1 \right\} \\ &\stackrel{(i)}{\leq} 2 \left\| \mathbf{X}^T w / n \right\|_\infty \|\hat{\Delta}\|_1 + 2\lambda_n \left\{ \left\| \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_{S^c} \right\|_1 \right\} \\ &\stackrel{(ii)}{\leq} \lambda_n \left\{ 3 \left\| \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_{S^c} \right\|_1 \right\}, \end{aligned} \quad (4.3)$$

where step (i) follows from a combination of Hölder's inequality and the triangle inequality, whereas step (ii) follows from the choice of  $\lambda_n$ .

Inequality (4.3) shows that  $\hat{\Delta} \in \mathbb{C}_3(S)$ , so that the RE condition may be applied. Doing so, we obtain

$$\kappa \|\hat{\Delta}\|_2^2 \leq 3\lambda_n \sqrt{s} \|\hat{\Delta}\|_2,$$

which implies the consistency. ■

**Example 7.14 in Wainwright's book.** Consider the classical linear Gaussian model, where  $w \in \mathbb{R}^n$  has i.i.d.  $\mathcal{N}(0, \sigma^2)$  entries.

Consider  $\mathbf{X} \in \mathbb{R}^{n \times d}$  is fixed. Suppose that  $\mathbf{X}$  satisfies the RE condition, and that it is C-column normalized, i.e.,

$$\max_{j=1, \dots, d} \frac{\|\mathbf{X}(\cdot, j)\|_2}{\sqrt{n}} \leq C.$$

Thus, the random variable  $\left\| \frac{\mathbf{X}^T w}{n} \right\|_\infty$  corresponds to the absolute maximum of  $d$  zero-mean Gaussian variables, each with variance at most  $\frac{C^2 \sigma^2}{n}$ , since

$$\text{var}\left(\frac{\mathbf{X}(\cdot, j)^T w}{n}\right) = \frac{\sigma^2}{n^2} \sum_{i=1}^n \mathbf{X}(i, j)^2 = \sigma^2 \frac{\|\mathbf{X}(\cdot, j)\|_2^2}{n^2} \leq \frac{C^2 \sigma^2}{n}.$$

The standard Gaussian tail bounds states that, for any  $i \in \{1, \dots, d\}$ ,

$$P\left(\left|\frac{\mathbf{X}(\cdot, j)^T w}{n}\right| \geq t\right) \leq 2 \exp\left\{-\frac{nt^2}{2C^2 \sigma^2}\right\} \text{ for all } t > 0.$$

Thus, for all  $\delta > 0$ ,

$$\begin{aligned} P\left(\left\|\frac{\mathbf{X}^T w}{n}\right\|_{\infty} \geq C\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)\right) &\leq \sum_{j=1}^d P\left(\frac{\mathbf{X}(\cdot, j)^T w}{n} \geq C\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)\right) \\ &\leq 2d \exp\left\{-\frac{nC^2\sigma^2\left(\sqrt{\frac{2\log d}{n}} + \delta\right)^2}{2C^2\sigma^2}\right\} \\ &\leq 2e^{-n\delta^2/2}. \end{aligned}$$

If we set  $\lambda_n = 2C\sigma\left(\frac{2\log d}{n} + \delta\right)$ , then the theorem implies that

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{6C\sigma}{k} \sqrt{s} \left\{ \frac{2\log d}{n} + \delta \right\}.$$