9.1 Lasso

Consider a regression framework where $Y$ is an $n \times 1$ vector, $X$ is a $n \times d$ matrix, $\theta^*$ is a $d \times 1$ vector, and $\epsilon$ is a $n \times 1$ vector. Further assume that

$$Y = X\theta^* + \epsilon$$

with $\epsilon \in \text{SG}_n(\sigma^2)$. In the LASSO, we use estimate $\hat{\theta}$ to estimate $\theta$, where $\hat{\theta}$ is the solution to

$$\hat{\theta} \in \arg\min_{\theta \in \mathbb{R}^d} \left( \frac{1}{2n} ||Y - X\theta||^2 + \lambda_n ||\theta||_1 \right). \quad (9.1)$$

Equation (1) defines a convex optimization problem that produces sparse solutions depending on $\lambda_n$. The parameter $\lambda_n$ is chosen by the user. It can be thought of as $\lambda(n, d, \sigma)$ because the choice will depend on those values.

Equation (1) has solutions for both $d \leq n$ and $d > n$. There can be multiple optimal solution $\hat{\theta}$, but the maximizing value $X\hat{\theta}$ is unique. For a discussion of the uniqueness of solutions to the Lasso problem, see [1].

The basic inequality [2] is a useful inequality for proving results pertaining to the Lasso. It is given below as Lemma 1.1. It is used to prove Theorem 1.2.

**Lemma 9.1.** In the Lasso set-up, if $\theta^*$ is the true parameter value and $\hat{\theta}$ is the lasso solution, then

$$\frac{1}{2n} \left| \frac{1}{2n} ||X(\hat{\theta} - \theta^*)||^2 \right|^2 \leq \epsilon^T X(\hat{\theta} - \theta^*) + \lambda_n(||\theta^*||_1 - ||\hat{\theta}||_1).$$

**Proof.**

$$\frac{1}{2n} \left( ||\epsilon||^2 + ||X(\hat{\theta} - \theta^*)||^2 - 2\epsilon^T X(\hat{\theta} - \theta^*) \right) + \lambda_n ||\hat{\theta}||_1 = \frac{1}{2n} \left( ||\epsilon - X(\hat{\theta} - \theta^*)||^2 + \lambda_n ||\hat{\theta}||_1 \right) = \frac{1}{2n} ||Y - X\hat{\theta}||^2 + \lambda_n ||\hat{\theta}||_1 \leq \frac{1}{2n} ||Y - X\theta^*||^2 + \lambda_n ||\theta^*||_1$$

$$= \frac{1}{2n} ||\epsilon||^2 + \lambda_n ||\theta^*||_1$$
Theorem 9.2. If
\[
\lambda_n \geq \| \frac{X^T \epsilon}{n} \|_\infty = \max_{j=1, \ldots, d} \left| \frac{X_j^T \epsilon}{n} \right|
\]
then any Lasso solution satisfies
\[
\left\| \frac{X(\hat{\theta} - \theta^*)}{n} \right\|^2 \leq 4 \| \theta^* \|_1 \lambda_n.
\]

Proof. Lemma 1.1 provides
\[
\frac{1}{2n} \left\| \frac{X(\hat{\theta} - \theta^*)}{n} \right\|^2 \leq \epsilon^T \frac{X(\hat{\theta} - \theta^*)}{n} + \lambda_n(\| \theta^* \|_1 - \| \hat{\theta} \|_1) \text{ by Basic inequality (Lemma 1.1)}
\]
\[
\leq \frac{1}{n} \| X^T \epsilon \|_\infty \left\| \theta - \theta^* \right\|_1 + \lambda_n(\| \theta^* \|_1 - \| \hat{\theta} \|_1) \text{ by Holder Inequality}
\]
\[
\leq \frac{1}{n} \| X^T \epsilon \|_\infty \left( \| \theta^* \|_1 + \| \hat{\theta} \|_1 \right) + \lambda_n(\| \theta^* \|_1 - \| \hat{\theta} \|_1) \text{ by Triangle Inequality}
\]
\[
\leq \left( \frac{1}{n} \| X^T \epsilon \|_\infty - \lambda_n \right) \| \hat{\theta} \|_1 + \left( \frac{1}{n} \| X^T \epsilon \|_\infty + \lambda_n \right) \| \hat{\theta} \|_1
\]
\[
\leq 2\lambda_n \| \theta^* \|_1.
\]

What is a good choice for $\lambda_n$?
Recall that $\epsilon \in \text{SG}(\sigma^2)$ and assume $\max_{j=1, \ldots, d} \| X_j \| \leq C\sqrt{n}$ for some $C > 0$. Then for $t > 0$,
\[
P \left( \left\| \frac{X^T \epsilon}{n} \right\|_\infty \geq t \right) \leq P \left( \max_j |X_j^T \epsilon| \geq tn \right)
\]
\[
\leq \sum_j P \left( |X_j^T \epsilon| \geq tn \right)
\]
\[
\leq \sum_j P \left( \frac{|X_j^T \epsilon|}{\| X_j \|} \geq \frac{tn}{\| X_j \|} \right)
\]
\[
\leq 2d \exp \left( \frac{-t^2n}{2\sigma^2 \max |X_j|^2} \right) \text{ by Subgaussianity}
\]
\[
\leq 2d \exp \left( \frac{-t^2n}{2\sigma^2 C^2} \right) \text{ because } \| X_j \|^2 < C^2 n
\]
\[
\leq \delta
\]
if we choose
\[
t = \lambda_n = \sqrt{\frac{2\sigma^2 C^2}{n} \left( \log \left( \frac{1}{\delta} \right) + \log(2d) \right)}.
\]

Consider $\delta = 1/n$. Then with probability $1 - 1/n$,
\[
\left\| \frac{X(\hat{\theta} - \theta^*)}{n} \right\|^2 \leq \| \theta^* \|_1 \sqrt{\frac{2\sigma^2 C (\log(n) + \log(2d))}{n}}
\]
If \( d \leq n \) and \( \lambda_{\min}(X^TX/n) \geq C_{\min} > 0 \). Then you can also get a bound for

\[
\left\| \hat{\theta} - \theta^* \right\|_2 \leq \frac{\left\| \theta^* \right\|}{C_{\min} \lambda_n}
\]

### 9.2 Getting Fast Rates for Lasso

In order to get “fast” rates for the lasso, there needs to be additional assumptions on \( X \). These assumptions also provide consistency of estimation of \( \theta \).

A very useful condition is the restricted eigenvalue condition. In order to define the condition, we need to establish some notation. For \( S \subset \{1, 2, \ldots, d\} \) and \( \alpha > 0 \), define

\[
C(\alpha, S) = \{ \Delta \in \mathbb{R}^d : \| \Delta_S \|_1 \leq \alpha \| \Delta_S \|_1 \}.
\]

**Definition 9.3.** \( X \) satisfies the restricted eigenvalue (RE(\( \alpha, \kappa \))) condition over \( S = \{1, \ldots, d\} \neq \emptyset \) if

\[
\frac{1}{n} \| X \Delta \|_2^2 \geq \kappa \| \Delta \|_2^2 \text{ for all } \Delta \in C(\alpha, S).
\]

For intuition, think of \( \Delta = \hat{\theta} - \theta^* \). We want \( \| X \Delta \|_2^2 / n \) to be small. Note that if it is, this does necessarily mean that \( \| \Delta \|_2^2 \) is small. Especially if

\[
\Delta \rightarrow \frac{\| X \Delta \|_2^2}{n}
\]

is flat around \( \hat{\theta} - \theta^* \). To prevent this, we need the function (9.2) to be very curved. This is true if

\[
\frac{\| X \Delta \|_2^2}{n} \geq \kappa \| \Delta \|_2^2 \text{ for all } \Delta \in \mathbb{R}^d.
\]

Unfortunately, this implies that \( \lambda_{\min}(X^TX) \geq C_{\min} > 0 \) if \( d > n \), which is not possible. Instead, we consider the case where function (9.2) is curved only along certain directions.

These directions are \( C(\alpha, S) \) where \( S \) is defined by the support of \( \theta^* \). That is, \( s = \{J : \theta^*_J \neq 0\} \).

**Theorem 9.4.** Assume that

- the support of \( \theta^* \) is \( S \) where \( |S| = s > 0 \).
- \( X \) satisfies \( \text{RE}(3, \kappa) \) where \( \kappa > 0 \) with respect to \( S \).
- \( \lambda_n \geq 2 \|e^TX\| / n \).

Then any Lasso solution \( \hat{\theta} \) satisfies

\[
\frac{1}{n} \| X \left( \hat{\theta} - \theta^* \right) \|_2^2 \leq 9\lambda_n^2 s \kappa,
\]

and

\[
\left\| \hat{\theta} - \theta^* \right\| \leq \frac{3}{\kappa} \sqrt{s} \lambda_n.
\]
Proof. First, we show that given our choice $\lambda_n$, $\hat{\Delta} = (\hat{\theta} - \theta^*) \in C(3, S)$. By the Basic inequality,

$$0 \leq \frac{1}{2n} \|X\hat{\Delta}\|^2 \leq \frac{\epsilon^T X \Delta}{n} + \lambda_n \|\theta^*\|_1 - ||\hat{\theta}||_1.$$  

Since $\theta^*$ is $S$-sparse, we know

$$\|\theta^*\|_1 - ||\hat{\theta}||_1 = \|\theta^*_S\|_1 - \|\theta^*_S + \hat{\Delta}_S\|_1 - ||\hat{\theta}_S||_1$$  

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

Plugging this into the Basic inequality yields:

$$0 \leq \frac{1}{2n} \|X\hat{\Delta}\|^2 \leq 2\|X^T \epsilon\|_\infty \|\hat{\Delta}\|_1 + 2\lambda_n \left(\|\theta^*_S\|_1 - \|\theta^*_S + \hat{\Delta}_S\|_1 - ||\hat{\Delta}_S||_1\right)$$  

$$\leq 2\|X^T \epsilon\|_\infty \|\hat{\Delta}\|_1 + 2\lambda_n \left(\|\hat{\Delta}_S\|_1 - ||\hat{\Delta}_S||_1\right)$$  

By triangle inequality

$$\leq \lambda_n ||\hat{\Delta}_S||_1 + \lambda n \|\hat{\Delta}_S\|_1 + 2\lambda_n \|\hat{\Delta}_S\|_1 - 2\lambda_n ||\hat{\Delta}_S||_1$$  

$$= \lambda_n \left(3||\hat{\Delta}_S||_1 - \|\hat{\Delta}_S\|_1\right)$$  

$$\Rightarrow \hat{\Delta} \in C(3, S).$$  

Note that the fourth line used the fact

$$\lambda_n \geq \frac{2\|X^T \epsilon\|_\infty}{n}.$$  

References
