Simultaneous Analysis of LASSO and Datnzig Selector

Peter J. Bickel, Ya'Acov Ritov and Alexandre B. Tsybakov

Reviewer: Neil Dhingra

Outline

1. **LASSO and Dantzig Selector Definitions (Section 2)**
2. **Restricted Positive Definiteness (Section 3)**
3. Linear Regression Model (Section 7)
4. **Sufficient Conditions for Restricted Positive Definiteness (Section 4)**
5. Nonparametric regression model (Section 2)
6. Equivalence Results (Section 5)
7. Oracle inequalities for Prediction Risk (Section 6)

Assumptions and Notation

- W_i are independent processes with $\mathcal{N}(0, \sigma^2)$
- The regularization parameter r,

$$r = A\sigma\sqrt{\frac{\log M}{n}}$$

- $\mathcal{M}(\beta)$ is number of nonzero elements of β
- $|x|_p$ is the p-norm of x
- δ_I is the sub vector of δ corresponding to the set of indices I

LASSO and Dantzig Selector (Section 2)

Linear Regression model

$$y = X\beta + w$$

LASSO and Dantzig Selector (Section 2)

Linear Regression model

$$y = X\beta + w$$

LASSO (normalized columns of X)

$$\hat{\beta}_L = \underset{\beta}{\text{argmin}} \ \frac{1}{n}\|y - X\beta\|_2^2 + 2r|\beta|_1 \tag{1}$$

LASSO and Dantzig Selector (Section 2)

Linear Regression model

$$y = X\beta + w$$

LASSO (normalized columns of X)

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \ \frac{1}{n}\|y - X\beta\|_2^2 + 2r|\beta|_1 \tag{1}$$

Necessary condition for optimality of (1) is,

$$\left|\frac{1}{n}X^T(y - X\beta)\right|_\infty \leq r \tag{2}$$

LASSO and Dantzig Selector (Section 2)

Linear Regression model

$$y = X\beta + w$$

LASSO (normalized columns of X)

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \ \frac{1}{n}\|y - X\beta\|_2^2 + 2r|\beta|_1 \tag{1}$$

Necessary condition for optimality of (1) is,

$$\left| \frac{1}{n} X^T (y - X\beta) \right|_\infty \le r \tag{2}$$

Dantzig Selector

$$\begin{aligned} \hat{\beta}_D = \underset{\beta}{\operatorname{argmin}} & \quad |\beta|_1 \\ \text{subject to} & \quad (2) \end{aligned} \tag{3}$$

Restricted positive definiteness (Section 3)

When J_0 is the support of β^* and $\delta := \hat{\beta} - \beta^*$

$$|\delta_{J_0^c}|_1 \leq c_0 |\delta_{J_0}|_1$$

with high probability that $c_0 = 1$ for Dantzig and $c_0 = 3$ for LASSO

Restricted positive definiteness (Section 3)

When J_0 is the support of β^* and $\delta := \hat{\beta} - \beta^*$

$$|\delta_{J_0^c}|_1 \leq c_0 |\delta_{J_0}|_1$$

with high probability that $c_0 = 1$ for Dantzig and $c_0 = 3$ for LASSO

$$\kappa(s, c_0) := \min_{J_0 \in J} \quad \min_{\delta \in \delta(J_0)} \frac{|X\delta|_2}{\sqrt{n}|\delta_{J_0}|_2}$$

where

$$\begin{aligned} J &= \{J_0 : J_0 \subseteq \{1, \ldots, M\}, |J_0| \leq s\} \\ \delta(J_0) &= \{\delta : \delta \neq 0, |\delta_{J_0^c}|_1 \leq c_0 |\delta_{J_0}|_1\} \end{aligned}$$

Assumption $RE(s, c_0)$ is a condition that $\kappa(s, c_0) > 0$

Theorem 7.1 (Section 7)

Let $\mathrm{diag}(X^T X) = \mathbb{1}$, $A > \sqrt{2}$, and $RE(s, 1)$, then with probability $1 - M^{1-A^2/2}$

$$|\hat{\beta}_D - \beta^*| \leq \frac{8A}{\kappa^2(s,1)} \sigma s \sqrt{\frac{\log M}{n}}$$

$$|X(\hat{\beta}_D - \beta^*)|_2^2 \leq \frac{16A^2}{\kappa^2(s,1)} \sigma^2 s \log M$$

Additionally, for $p \in [1, 2]$, with $RE(s, m, 1)$

$$|\hat{\beta}_D - \beta^*|_p^p \leq 2^{p-1} 8 \left\{ 1 + \sqrt{\frac{s}{m}} \right\}^{2(p-1)} s \left(\frac{A\sigma}{\kappa^2(s,m,1)} \sqrt{\frac{\log M}{n}} \right)^p$$

Theorem 7.2 (Section 7)

Let $\operatorname{diag}(X^T X) = \mathbb{1}$, $A > 2\sqrt{2}$, and $RE(s, 1)$, then with probability $1 - M^{1-A^2/8}$

$$
\begin{aligned}
|\hat{\beta}_L - \beta^*| &\leq \frac{16A}{\kappa^2(s,3)} \sigma s \sqrt{\frac{\log M}{n}} \\
|X(\hat{\beta}_L - \beta^*)|_2^2 &\leq \frac{16A^2}{\kappa^2(s,3)} \sigma^2 s \log M
\end{aligned}
$$

Additionally, for $p \in [1, 2]$, with $RE(s, m, 1)$

$$
|\hat{\beta}_D - \beta^*|_p^p \leq 16 \left\{ 1 + 3\sqrt{\frac{s}{m}} \right\}^{2(p-1)} s \left(\frac{A\sigma}{\kappa^2(s,m,3)} \sqrt{\frac{\log M}{n}} \right)^p
$$

$$
\mathcal{M}(\hat{\beta}_L) \leq \frac{64\phi_{\max}}{\kappa^2(s,3)} s
$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

$$\phi_{\min}(u) := \min \frac{|X\delta|_2}{|\delta|_2}$$
$$\phi_{\max}(u) := \max \frac{|X\delta|_2}{|\delta|_2}$$

$$\mathcal{M}(\delta) \leq u$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

$$\begin{aligned} \phi_{\min}(u) &:= \min \frac{|X\delta|_2}{|\delta|_2} \\ \phi_{\max}(u) &:= \max \frac{|X\delta|_2}{|\delta|_2} \end{aligned}$$

$$\mathcal{M}(\delta) \leq u$$

Notion of how 'orthogonal' the columns of X are

$$\theta_{m_1,m_2} = \max \frac{c_1 X_{I_1}^T X_{I_2} c_2}{n|c_1|_2|c_2|_2}$$

$$\begin{aligned} I_1 \cap I_2 &= \emptyset \\ |I_i| &\leq m_i \\ c_i &\in \mathbb{R}^{I_i} \backslash \{0\} \end{aligned}$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

Assumption 4

$$\phi_{\min}(s) > 2c_0\theta_{1,1}s$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

Assumption 4

$$\phi_{\min}(s) > 2c_0\theta_{1,1}s$$

Assumption 5

When $\text{diag}(X^TX) = \mathbb{1}$ (i.e. normalized columns),

$$\theta_{1,1} < \frac{1}{(1+2c_0)s}$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

Assumption 3

$$\phi_{\min}(s) > 2c_0\theta_{s,1}\sqrt{s}$$

Sufficient Conditions for $RE(s, c_0)$ (Section 4)

Assumption 1

$$\phi_{\min}(2s) > c_0 \, \theta_{s,2s}$$

Assumption 2

$$m\phi_{\min}(s+m) > c_0^2 s \phi_{\max}(m)$$

General Nonparametric Regression (Section 2)

$$Y_i = f(Z_i) + W_i$$

where $f : \mathcal{Z} \to \mathbb{R}$. Dictionary $\mathcal{F}_M := \{f_1, \ldots, f_m\}$

$$\|g\|_n := \sqrt{\frac{1}{n}\sum_{i=1}^{n} g^2(Z_i)}$$

$$\begin{aligned} f_{\max} &:= \max_{1 \le j \le M} \|f_j\|_n \\ f_{\min} &:= \min_{1 \le j \le M} \|f_j\|_n \end{aligned}$$

General Nonparametric Regression (Section 2)

$$Y_i = f(Z_i) + W_i$$

where $f : \mathcal{Z} \to \mathbb{R}$. Dictionary $\mathcal{F}_M := \{f_1, \ldots, f_m\}$

$$\|g\|_n := \sqrt{\frac{1}{n}\sum_{i=1}^{n} g^2(Z_i)}$$

$$\begin{aligned} f_{\max} &:= \max_{1 \le j \le M} \|f_j\|_n \\ f_{\min} &:= \min_{1 \le j \le M} \|f_j\|_n \end{aligned}$$

In Linear Regression,

$$f_\beta = X\beta, \qquad f_i = i\text{th column of } X$$

General Nonparametric Regression (Section 2)

LASSO

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \; \frac{1}{n}\|y - X\beta\|_2^2 + 2r\sum_{j=1}^{M}\|f_j\|_n|\beta_j|_1$$

General Nonparametric Regression (Section 2)

LASSO

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \; \frac{1}{n}\|y - X\beta\|_2^2 + 2r\sum_{j=1}^{M}\|f_j\|_n|\beta_j|_1$$

Dantzig set changes similarly to reflect weights

$$\left|\frac{1}{n}\mathrm{diag}(\|f_i\|_n^{-1}) \;\; X^T(y - X\beta)\right|_\infty \le r$$

Theorem 5.1 (Section 5)

With $RE(s,1)$, $A > 2\sqrt{2}$

$$\left| \|\hat{f}_D - f\|_n^2 - \|\hat{f}_L - f\|_n^2 \right| \le 16A^2 \frac{\mathcal{M}(\hat{\beta}_L)\sigma^2}{n} \frac{f_{\max}^2}{\kappa^2(s,1)} \log M$$

Theorem 5.2 (Section 5)

With $RE(s,5)$, $A > 2\sqrt{2}$

$$\|\hat{f}_L - f\|_n^2 \le 10\|\hat{f}_D - f\|_n^2 + 81A^2 \frac{\mathcal{M}(\hat{\beta}_D)\sigma^2}{n} \frac{\log M}{\kappa^2(s,5)}$$

Theorem 6.1 (Section 6)

With $RE(s, 3 + 4/\epsilon)$

$$\|\hat{f}_L - f\|_n^2 \leq (1+\epsilon) \inf_{\mathcal{M}(\beta) \leq s} \left\{ \|f_\beta - f\|_n^2 + \frac{C(\epsilon) f_{\max}^2 A^2 \sigma^2}{\kappa^2(s, 3 + 4/\epsilon)} \frac{\beta \log M}{n} \right\}$$

Proposition 6.3 (Section 6)

With $RE(s\max\{C_1(\epsilon),1\}, 3+4/\epsilon)$

$$\|\hat{f}_D - f\|_n^2 \leq (1+\epsilon) \inf_{\mathcal{M}(\beta)\leq s} \left\{ \|f_\beta - f\|_n^2 + \frac{C_2(\epsilon) f_{\max}^2 A^2 \sigma^2}{\kappa_0^2 /} \frac{s\log M}{n} \right\}$$

where

$$\begin{aligned} C_1(\epsilon) &= 4[(1+\epsilon)C_0 + C(\epsilon)]\frac{\phi_{\max} f_{\max}^2}{\kappa^2 f_{\min}^2} \\ C_2(\epsilon) &= 16 C_1(\epsilon) + C(\epsilon) \\ \kappa_0 &= \kappa(\max\{C_1(\epsilon),1\},1)s, 3+4/\epsilon) \end{aligned}$$

Questions?