A theorem on the estimation and prediction properties of the Lasso

Comments and corrections to r.samworth@statslab.cam.ac.uk

Consider the linear model $Y = \beta_0 \mathbf{1}_n + X\beta + \epsilon$, where the columns of the deterministic design matrix $X = (x_1, \ldots, x_p) \in \mathbb{R}^{n \times p}$ are centred and have $\|x_j\|_2^2 = n$ for $j = 1, \ldots, p$, and where $\epsilon \sim N_n(0, \sigma^2 I)$. Recall that the Lasso estimator of β with regularisation parameter $\lambda > 0$ is $\hat{\beta}^L_\lambda$, where $(\hat{\beta}_0, \hat{\beta}^L_\lambda)$ minimises

$$Q_1(\beta_0, \beta) = \frac{1}{2n}\|Y - \beta_0 \mathbf{1}_n - X\beta\|_2^2 + \lambda\|\beta\|_1$$

over $(\beta_0, \beta) \in \mathbb{R} \times \mathbb{R}^p$. Let $S = \big\{j \in \{1, \ldots, p\} : \beta_j \neq 0\big\}$, let $N = \{1, \ldots, p\} \setminus S$, and let $s = |S|$. Recall that for an arbitrary $A \subseteq \{1, \ldots, p\}$ and $b \in \mathbb{R}^p$, we write b_A for the vector in $\mathbb{R}^{|A|}$ obtained by extracting the components of b that are in A. We will assume the following *compatibility condition*:

(A1) There exists $\phi_0 > 0$ such that for all $b \in \mathbb{R}^p$ with $\|b_N\|_1 \leq 3\|b_S\|_1$, we have

$$\|b_S\|_1^2 \leq \frac{s\|Xb\|_2^2}{n\phi_0^2}.$$

Theorem 1. *Assume **(A1)** and let $\lambda = A\sigma\sqrt{\frac{\log p}{n}}$ for some $A > 0$. Then with probability at least $1 - p^{-(A^2/8-1)}$, we have*

$$\frac{1}{n}\|X(\hat{\beta}^L_\lambda - \beta)\|_2^2 + \lambda\|\hat{\beta}^L_\lambda - \beta\|_1 \leq \frac{16\lambda^2 s}{\phi_0^2} = \frac{16A^2}{\phi_0^2}\frac{\sigma^2 s \log p}{n}.$$

Proof. We write $\bar{\epsilon} = n^{-1}\sum_{i=1}^n \epsilon_i$, and note that $\hat{\beta}_0 = n^{-1}\sum_{i=1}^n Y_i = \beta_0 + \bar{\epsilon}$. For convenience drop the λ subscript in $\hat{\beta}^L_\lambda$. By definition of $(\hat{\beta}_0, \hat{\beta}^L)$, we have $Q(\hat{\beta}_0, \hat{\beta}^L) \leq Q(\hat{\beta}_0, \beta)$, so

$$\frac{1}{2n}\|X\beta + \epsilon - \bar{\epsilon}\mathbf{1}_n - X\hat{\beta}^L\|_2^2 + \lambda\|\hat{\beta}^L\|_1 \leq \frac{1}{2n}\|\epsilon - \bar{\epsilon}\mathbf{1}_n\|_2^2 + \lambda\|\beta\|_1.$$

Thus

$$\frac{1}{n}\|X(\hat{\beta}^L - \beta)\|_2^2 + 2\lambda\|\hat{\beta}^L\|_1 \leq \frac{2}{n}\epsilon^T X(\hat{\beta}^L - \beta) + 2\lambda\|\beta_S\|_1.$$

Define the event

$$\Omega_0 = \bigg\{\frac{2}{n}\|X^T\epsilon\|_\infty \leq \lambda\bigg\}.$$

It can be shown (see example sheet) that $\mathbb{P}(\Omega_0) \geq 1 - p^{-(A^2/8-1)}$, so henceforth we work on Ω_0. Now

$$\begin{aligned}\frac{1}{n}\|X(\hat{\beta}^L - \beta)\|_2^2 + \lambda\|\hat{\beta}_N^L\|_1 &= \frac{1}{n}\|X(\hat{\beta}^L - \beta)\|_2^2 + 2\lambda\|\hat{\beta}^L\|_1 - 2\lambda\|\hat{\beta}_S\|_1 - \lambda\|\hat{\beta}_N\|_1 \\ &\leq \frac{2}{n}\epsilon^T X(\hat{\beta}^L - \beta) - \lambda\|\hat{\beta}_N^L\|_1 + 2\lambda\|\beta_S\|_1 - 2\lambda\|\hat{\beta}_S^L\|_1 \\ &\leq \lambda\|\hat{\beta}^L - \beta\|_1 - \lambda\|\hat{\beta}_N^L\|_1 + 2\lambda\|\hat{\beta}_S^L - \beta_S\|_1 \\ &= 3\lambda\|\hat{\beta}_S^L - \beta_S\|_1.\end{aligned}$$

But

$$\|\hat{\beta}_N^L - \beta_N\|_1 = \|\hat{\beta}_N^L\|_1 \leq \frac{1}{\lambda}\left\{\frac{1}{n}\|X(\hat{\beta}^L - \beta)\|_2^2 + \lambda\|\hat{\beta}_N^L\|_1\right\} \leq 3\|\hat{\beta}_S^L - \beta_S\|_1.$$

Hence, using the compatibility condition,

$$\begin{aligned}\frac{1}{n}\|X(\hat{\beta}^L - \beta)\|_2^2 + \lambda\|\hat{\beta}^L - \beta\|_1 \leq 4\lambda\|\hat{\beta}_S^L - \beta_S\|_1 &\leq \frac{4\lambda\sqrt{s}\|X(\hat{\beta}^L - \beta)\|_2}{\sqrt{n}\phi_0} \\ &\leq \frac{16\lambda^2 s}{\phi_0^2} = \frac{16A^2}{\phi_0^2}\frac{\sigma^2 s\log p}{n},\end{aligned}$$

as required. □