Theoretical Results of Lasso Solutions

Yajie Bao
Shanghai Jiao Tong University

January 25, 2020

1 Single variable Lasso

Consider single variable Lasso problem

$$\underset{\beta}{\operatorname{minimize}}\left\{\frac{1}{2N}\sum_{i=1}^{N}(y_i - z_i\beta)^2 + \lambda|\beta|\right\} \tag{1.1}$$

which is a convex problem. By optimal condition of convex optimization problem, set the subgradient be 0 then we have

$$\widehat{\beta} = \begin{cases} \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle - \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle \\ 0 & \text{if } \frac{1}{N}|\langle \mathbf{z}, \mathbf{y}\rangle| \leq \lambda \\ \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle + \lambda & \text{if } \frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle < -\lambda \end{cases}$$

which can be write as

$$\widehat{\beta} = \mathcal{S}_\lambda\left(\frac{1}{N}\langle \mathbf{z}, \mathbf{y}\rangle\right),$$

where $\mathcal{S}_\lambda(x) = \operatorname{sign}(x)(|x| - \lambda)_+$ is the *soft-thresholding operator*.

2 Uniqueness of Lasso solution

Now given a response vector $y \in \mathbb{R}^n$, a design matrix $X \in \mathbb{R}^{n\times p}$ and a tuning parameter $\lambda \geq 0$, the lasso estimate is

$$\hat{\beta} \in \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1. \tag{2.1}$$

Tibshirani et al. [2013] gave some basic propeties about the lasso solutions, which is called the uniqueness of fitted values.

Lemma 2.1 *For any y, X and $\lambda \geq 0$, the lasso problem (2.1) have the following properties:*

1. *Every lasso solution gives the same fitted value $X\hat{\beta}$*

2. If $\lambda > 0$, then every lasso solution has the same ℓ_1 norm $\|\hat{\beta}\|_1$.

Proof: Suppose there are two different solutions $\hat{\beta^{(1)}}$ and $\hat{\beta^{(2)}}$, and denote the minimum value by c^*. By the strong convexity of $(y-x)^2$ and convexity of ℓ_1 norm, for $0 < \alpha < 1$ we have

$$\frac{1}{2}\left\|y - X\left(\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}\right)\right\|_2^2 + \lambda\left\|\alpha\hat{\beta}^{(1)} + (1-\alpha)\hat{\beta}^{(2)}\right\|_1 < \alpha c^* + (1-\alpha)c^* = c^*,$$

which leads a contradiction. Then we have proved assertion 1, it follows assertion 2 immediately. ∎

By KTT conditions, we have the following results

$$X^T(y - X\hat{\beta}) = \lambda\gamma, \tag{2.2}$$

$$\gamma_i \in \begin{cases} \left\{\text{sign}\left(\hat{\beta}_i\right)\right\} & \text{if } \hat{\beta}_i \neq 0 \\ [-1, 1] & \text{if } \hat{\beta}_i = 0 \end{cases}, \quad \text{for } i = 1, \dots p$$

Here γ is the subgradient of $\|\beta\|_1$ evaluated at $\hat{\beta}$. First we define the equicorrelation set

$$S = \left\{i \in \{1, \dots p\} : \left|X_i^T(y - X\hat{\beta})\right| = \lambda\right\},$$

and by (2.2) and the uniqueness of fitted value, every lasso optimal solution has the same subgradient. **If there exist a optimal solution $\hat{\beta}$ such that its subgradient satisfies $|\gamma_i| < 1$ for any $i \notin S$**, then the lasso problem (2.1) has the unique equicorrelation set for fixed $\lambda > 0$.

There are some other equivalent forms of Lasso problem (2.1),

$$\min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2 \quad \text{subject to } \|\beta\|_1 \leq \lambda_s,$$

and

$$\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{subject to } \|y - X\beta\|_2^2 \leq \gamma_s.$$

Candes and Tao [2007] proposed Dantzig selector as

$$\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{subject to } \|X^T(y - X\beta)\|_\infty \leq \lambda_D. \tag{2.3}$$

Let $\tilde{\beta}$ be the optimal solution of Lasso program (2.1), by the zero-subgradient condition, it's easy to see that $\tilde{\beta}$ is a feasible solution of (2.3) when $\lambda_D = \lambda$.

3 Lasso dual

Given $y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}$, recall the Lasso problem

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1.$$

We can transform the primal to

$$\min_{\beta\in\mathbb{R}^p, z\in\mathbb{R}^n} \frac{1}{2}\|y-z\|_2^2 + \lambda\|\beta\|_1 \quad \text{subject to} \quad z = X\beta.$$

Then the dual function is

$$g(u) = \min_{\beta\in\mathbb{R}^p, z\in\mathbb{R}^n} \frac{1}{2}\|y-z\|_2^2 + \lambda\|\beta\|_1 + u^T(z - X\beta),$$

taking derivatives on z and $beta$, according to optimal condion we have

$$z = y - u$$
$$\lambda\partial\|\beta\|_1 = X^T u.$$

This yields the lasso dual problem

$$\max_{u\in\mathbb{R}^n} \frac{1}{2}\left(\|y\|_2^2 - \|y-u\|_2^2\right) \quad \text{subject to} \quad \left\|X^T u\right\|_\infty \leq \lambda,$$

or equivalently

$$\min_{u\in\mathbb{R}^n} \|y-u\|_2^2 \quad \text{subject to} \quad \left\|X^T u\right\|_\infty \leq \lambda.$$

Further, note that given the dual solution u, any lasso solution β satisfies

$$X\beta = y - u,$$

so **the lasso fit is just the dual residual**.

4 ℓ_2 bound

Next we will establish the ℓ_2 bound for $\left|\hat{\beta} - \beta^*\right|$. First we will rewrite the Lasso program as

$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^p} \frac{1}{2}\beta^T A\beta - b^T\beta + \lambda\|\beta\|_1, \tag{4.1}$$

where $A = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$ and $b = \frac{1}{n}\sum_{i=1}^n y_i X_i^T$.

Theorem 4.1 *Suppose that there exists some positive constants c_1 and c_2 such that*

$$\min_{\delta:\|\delta\|_1 < c_1\sqrt{s}\|\delta\|_2} \frac{\delta^T A\delta}{\|\delta\|_2^2} \geq c_2, \tag{4.2}$$

and the regularization parameter satisfies $\|A\beta^* - b\|_\infty \leq \frac{\lambda}{2}$. Then there exists some positive constant c such that

$$\|\hat{\beta} - \beta^*\|_2 \leq c\sqrt{s}\lambda, \tag{4.3}$$

where $k = |S|$.

Proof: First we need to verify the scale of regularization parameter, note that

$$\|A\beta^* - b\|_\infty = \left\| \frac{1}{n} \sum_{i=1}^{n} X_i w_i \right\|_\infty = \max_{1 \le j \le p} \left| \sum_{i=1}^{n} X_{ij} w_i \right|.$$

Then by exponential inequality, there exists some positive constant C such that

$$\lambda = C\sqrt{\frac{\log p}{n}},$$

and $\|A\beta^* - b\|_\infty \le \frac{\lambda}{2}$ holds with probability 1 for sufficiently large n. Next we will prove $\|\hat{\beta} - \beta^*\|_1 \le 4\sqrt{s}\|\hat{\beta} - \beta^*\|_2$. Using the fact that $\hat{\beta}$ is the optimal solution of the Lasso program, then we obtain

$$\frac{1}{2}\hat{\beta}^T A\beta - b^T\hat{\beta} + \lambda\|\hat{\beta}\|_1 \le \frac{1}{2}\beta^{*T} A\beta^* - b^T\beta^* + \lambda\|\beta^*\|_1, \tag{4.4}$$

which implies that

$$\begin{aligned}
\frac{1}{2}\hat{\beta}^T A\beta - b^T\hat{\beta} - \left(\frac{1}{2}\beta^{*T} A\beta^* - b^T\beta^*\right) &\le \lambda\left(\|\beta^*\|_1 - \|\hat{\beta}\|_1\right) \\
&= \lambda\left(\|\beta^*_S\|_1 - \|\hat{\beta}_S\|_1 - \|\hat{\beta}_{S^c}\|_1\right) \\
&\le \lambda\left(\|\beta^*_S - \hat{\beta}_S\|_1 - \|\beta^*_{S^c} - \hat{\beta}_{S^c}\|_1\right).
\end{aligned}$$

Note that,

$$\begin{aligned}
\frac{1}{2}\hat{\beta}^T A\beta - b^T\hat{\beta} - \left(\frac{1}{2}\beta^{*T} A\beta^* - b^T\beta^*\right) &\ge (A\beta^* - b)^T(\hat{\beta} - \beta^*) \\
&\ge -\left\|A\beta^* - b\right\|_\infty \|\hat{\beta} - \beta^*\|_1 \\
&\ge -\frac{\lambda}{2}\|\hat{\beta} - \beta^*\|_1 \\
&= -\frac{\lambda}{2}\left(\|\hat{\beta}_S - \beta^*_S\|_1 + \|\hat{\beta}_{S^c} - \beta^*_{S^c}\|_1\right)
\end{aligned}$$

and combining the inqualities above we have

$$\|\hat{\beta}_{S^c} - \beta^*_{S^c}\|_1 \le 3\|\hat{\beta}_S - \beta^*_S\|_1. \tag{4.5}$$

Thus we can obtain

$$\begin{aligned}
\|\hat{\beta} - \beta^*\|_1 &= \|\hat{\beta}_S - \beta^*_S\|_1 + \|\hat{\beta}_{S^c} - \beta^*_{S^c}\|_1 \\
&\le 4\sqrt{s}\|\hat{\beta}_S - \beta^*_S\|_2.
\end{aligned}$$

Then by assumption (4.1), we have

$$\left(\hat{\beta} - \beta^*\right)^T A\left(\hat{\beta} - \beta^*\right) \ge c_2\|\hat{\beta} - \beta^*\|_2.$$

In addition, the zero-subgradient conditon implies $\|A\hat{\beta} - b\|_\infty \leq \lambda$, then

$$\left(\hat{\beta} - \beta^*\right)^T A \left(\hat{\beta} - \beta^*\right) \leq \left(\hat{\beta} - \beta^*\right)^T \left(A\hat{\beta} - b - (A\beta^* - b)\right)$$
$$\leq \frac{3}{2}\lambda\|\hat{\beta} - \beta^*\|_1.$$

Thus $\|\hat{\beta} - \beta^*\|_2 \leq 6\sqrt{s}\lambda$. ■

5 Primal-Dual Witness Construction

For any vector $\beta \in \mathbb{R}^p$, we define its support $S(\beta) = \{i|\beta_i \neq 0\}$. Then we begin with a lemma in Wainwright [2009].

Lemma 5.1 *1. A vector is optimal if and only if there exists a subgradient vector $\hat{z} \in \partial|\hat{\beta}|_1$ satisfying*

$$\frac{1}{n}X^T X\left(\widehat{\beta} - \beta^*\right) - \frac{1}{n}X^T w + \lambda_n \widehat{z} = 0. \tag{5.1}$$

2. *Suppose the subgradient vector satisfies the strict dual feasibility condition $|\widehat{z_j}| < 1$ for any $j \notin S(\hat{\beta})$. Then any optimal solution $\tilde{\beta}$ of lasso program satisfies $\tilde{\beta}_j = 0$ for any $j \notin S(\hat{\beta})$.*

3. *Under the condition of part (2), if the $k \times k$ matrix $X^T_{S(\hat{\beta})}X_{S(\hat{\beta})}$ is invertible, then $\hat{\beta}$ is the unique optimal solution of the Lasso program.*

Proof: The assertion 1 and 2 follows KKT conditions and uniqueness of subgradient in section 2.

Under the condition of part (2), the lasso program can be restricted in subspace $\beta \in \mathbb{R}^S$. If $X^T_{S(\hat{\beta})}X_{S(\hat{\beta})}$ is invertible, then the lasso program is strictly convex, therefore its optimal solution is unique. ■

Let S be the support set of the ture vector β^*, and assume that $X_S^T X_S$ is invertible. Wainwright [2009] proposed an incredible method called *Primal-Dual Witness Construction*, which conscruted an lasso solution covering ture support set under strict dual feasibility condition. The steps of *Primal-Dual Witness Construction* are presented as following

1. First, we obtain $\check{\beta}_S \in \mathbb{R}^k$ by solving restricted Lasso problem

$$\check{\beta}_S = \arg\min_{\beta_S \in \mathbb{R}^k} \left\{\frac{1}{2n}\|y - X_S\beta_S\|_2^2 + \lambda_n\|\beta_S\|_1\right\}. \tag{5.2}$$

   And the solution is unqiue under the condition that $X_S^T X_S$ is invertible. We set $\check{\beta}_{S^c} = 0$.

2. Second, we choose an subgradient of the ℓ_1 norm of $\check{\beta}_S$ denoted by $\check{z}_S \in \mathbb{R}^k$.

3. Third, we obtain $\check{z}_{S^c}$ by solving the zero-subgradient condition (5.1). Then we check for strict dual feasibility, i.e. $|\check{z}_{S^c}| < 1$ for all $j \in S^c$.

4. Fourth, we check whether the sign consistency condition $\check{z}_S = \text{sign}\left(\beta_S^*\right)$ is satisfied.

Since the pair $(\check{\beta},\ \check{z})$ satisfies the zero-subgradient condition (5.1), it must be an optimal solution of the Lasso program. Then according to Lemma 5.1, it's easy to see that if steps 1-3 succeed with strict dual feasibility, then the Lasso program has a unqiue solution with $S(\hat{\beta}) \subseteq S\left(\beta^*\right)$; If steps 1-4 succeed with strict dual feasibility, then the Lasso program has a unqiue solution with the correct signed support. Next we will prove strict dual feasibility and sign consistency condition.

We can write zero-subgradient condition (5.1) as matrix form,

$$\frac{1}{n}\begin{bmatrix} X_S^T X_S & X_S^T X_{S^c} \\ X_{S^c}^T X_S & X_{S^c}^T X_{S^c} \end{bmatrix}\begin{bmatrix} \beta_S - \beta_S^* \\ 0 \end{bmatrix} - \frac{1}{n}\begin{bmatrix} X_S^T \\ X_{S^c}^T \end{bmatrix}\begin{bmatrix} w_S \\ w_{S^c} \end{bmatrix} + \lambda \begin{bmatrix} z_S \\ z_{S^c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{5.3}$$

the element of noise vector w is sub-gaussian variable with zero-mean and parameter σ^2. Using the assumed invertibility of $X_S^T X_S$ we can solve for $\check{\beta}_S - \beta_S^*$ as follows

$$\check{\beta}_S - \beta_S^* = \left(\frac{1}{n} X_S^T X_S\right)^{-1}\left[\frac{1}{n} X_S^T w - \lambda \check{z}_S\right]. \tag{5.4}$$

Moreover, we can solve $\check{z}_{S^c}$ in terms of $\check{\beta}_S - \beta_S^*$ as

$$\hat{z}_{S^c} = X_{S^c}^T\left[X_S\left(X_S^T X_S\right)^{-1}\check{z}_S + \Pi_{X_S^{\perp}}\left(\frac{w}{\lambda_n n}\right)\right], \tag{5.5}$$

where $\Pi_{X_S^{\perp}} = I_S - \left(X_S^T X_S\right)^{-1} X_S^T$ is orthogonal projection matrix. We define the random variable

$$\Delta_i := e_i^T\left(\frac{1}{n} X_S^T X_S\right)^{-1}\left[\frac{1}{n} X_S^T w - \lambda_n \,\text{sgn}\left(\beta_S^*\right)\right], \tag{5.6}$$

and note that, $\Delta_i = \check{\beta}_S - \beta_S^*$, when $\check{z}_S = \text{sgn}\left(\beta_S^*\right)$. Then the sign consistency condition is equivalent to the assertion that for all $i \in S$,

$$\text{sgn}\left(\beta_i^* + \Delta_i\right) = \text{sgn}(\beta_i^*) \tag{5.7}$$

holds. To prove the strict dual feasibility condtion and sign consistency condition, now we impose the following assumptions:

Assumption 5.2 *There exists some* ***incoherence parameter*** $\gamma \in (0, 1]$*, such that*

$$\left\|X_{S^c}^T X_S\left(X_S^T X_S\right)^{-1}\right\|_\infty \leq (1-\gamma). \tag{5.8}$$

And there exists some $C > 0$ *such that*

$$\lambda_{min}\left(\frac{1}{n} X_S^T X_S\right) \geq C_{min}. \tag{5.9}$$

Theorem 5.3 *Under Assumption 5.2, and the n-dimensional columns of the design matrix X satisfy that $n^{-1/2}\max_{j\in S^c}\|X_j\| \leq 1$. Suppose that the sequence of regularization parameter $\{\lambda_n\}$ satisfies*

$$\lambda_n > \frac{2}{\gamma}\sqrt{\frac{2\sigma^2\log p}{n}}. \tag{5.10}$$

Then there exists some constant $c_1 > 0$, the following properties hold with probability greater than $1 - 2\exp(-c_1 n\lambda_n^2) \to 1$.

1. *The Lasso program has unqiue solution $\hat{\beta}$ and its support set contained within the ture support set, i.e., $S(\widehat{\beta}) \subseteq S(\beta^*)$. And $\hat{\beta}$ satisfies the ℓ_∞ bound*

$$\left\|\widehat{\beta}_S - \beta_S^*\right\|_\infty \leq \lambda_n\left[\left\|\left(X_S^T X_S/n\right)^{-1}\right\|_\infty + \frac{4\sigma}{\sqrt{C_{\min}}}\right]. \tag{5.11}$$

2. *If the minimum value of the ture parameter β^* on its support is bounded below as $\beta_{min} > g(\lambda_n)$ where $g(\lambda_n) = \lambda_n\left[\left\|\left(X_S^T X_S/n\right)^{-1}\right\|_\infty + \frac{4\sigma}{\sqrt{C_{\min}}}\right]$, then $\hat{\beta}$ has the correct signed support.*

Proof: Using the fact $|\check{z}_S| \leq 1$, (5.5) and incoherence condition (5.8), we have

$$\check{z}_{S^c} \leq 1 - \gamma + X_{S^c}^T \Pi_{X_S^\perp}\left(\frac{w}{\lambda_n n}\right).$$

Let $\tilde{Z}_i = X_j^T \Pi_{X_S^\perp}\left(\frac{w}{\lambda_n n}\right),\ j \in S^c$, then

$$\max_{j\in S^c} \check{z}_j \leq 1 - \gamma + \max_{j\in S^c} \tilde{Z}_i.$$

By the propertity of sub-gaussian random variable and the fact that the spectral norm of Π_{X_S} is 1 and the condition $n^{-1/2}\max_{j\in S^c}\|X_j\| \leq 1$, the parameter of $\tilde{Z}_i$ is bounded by

$$\frac{\sigma^2}{\lambda_n^2 n^2}\left\|\Pi_{X_S^\perp}(X_j)\right\|_2^2 \leq \frac{\sigma^2}{\lambda_n^2 n}.$$

Consequently, by tail probability bound of sub-gaussian variable and the uniform bound, we have

$$\mathbb{P}\left[\max_{j\in S^c}\left|\widetilde{Z}_j\right| \geq t\right] \leq 2(p-k)\exp\left(-\frac{\lambda_n^2 n t^2}{2\sigma^2}\right).$$

Set $t = \frac{\gamma}{2}$, we obtain

$$\begin{aligned}\mathbb{P}\left[\max_{j\in S^c}|\check{z}_j| > 1 - \frac{\gamma}{2}\right] &\leq 2\exp\left\{-\frac{\lambda_n^2 n\gamma^2}{8\sigma^2} + \log(p-k)\right\}\\ &\leq 2\exp\left\{-\log p + \log(p-k)\right\},\end{aligned}$$

then there exists some constant c_1 such that

$$\mathbb{P}\left[\max_{j\in S^c}|\check{z}_j| > 1 - \frac{\gamma}{2}\right] \leq 2\exp\left(-c_1 n\lambda_n^2\right).$$

Thus with paobability greater than $1-2\exp\left(-c_1 n\lambda_n^2\right)$, $S(\widehat{\beta}) \subseteq S\left(\beta^*\right)$ holds. By triangle inequality,

$$\left\|\check{\beta}_S-\beta_S^*\right\|_\infty \leq \left\|\left(\frac{X_S^T X_S}{n}\right)^{-1} X_S^T \frac{w}{n}\right\|_\infty + \left\|\left(\frac{X_S^T X_S}{n}\right)^{-1}\right\|_\infty \lambda_n.$$

For each $i \in S$, we define

$$V_i := e_i^T \left(\frac{X_S^T X_S}{n}\right)^{-1} X_S^T \frac{w}{n},$$

obviously V_i is a zero-mean sub-gaussian variable with parameter at most

$$\frac{\sigma^2}{n}\left\|\left(\left(\frac{1}{n} X_S^T X_S\right)^{-1} \|\right)_2 \leq \frac{\sigma^2}{C_{\min} n}$$

Then by sub-gaussian tail bound and the union bound we have

$$\mathbb{P}\left[\max_{i=1,\ldots,k} |V_i| > t\right] \leq 2\exp\left(-\frac{t^2 C_{\min} n}{2\sigma^2} + \log k\right),$$

setting $t = 4\sigma\lambda_n/\sqrt{C_{\min}}$ and using the fact $8n\lambda_n^2 > \log p \geq \log k$, we obtain

$$\mathbb{P}\left[\max_{i=1,\ldots,k} \left|\bar{Z}_i\right| > 4\sigma\lambda_n/\sqrt{C_{\min}}\right] \leq 2\exp\left(-c_1 n\lambda_n^2\right).$$

Therefore, we have proved assertion 1 and assertion 2 follows immediately. ■

6 Some useful inequalities

Lemma 6.1 (Tail Bounds for χ^2 Variables) *Given a centralized χ^2 variable X with d degrees of freedom, then for all $t \in (0, \frac{1}{2})$, we have*

$$\mathbb{P}[X \geq d(1+t)] \leq \exp\left(-\frac{3}{16} dt^2\right), \tag{6.1}$$

$$\mathbb{P}[X \leq (1-t)d] \leq \exp\left(-\frac{1}{4} dt^2\right) \tag{6.2}$$

Lemma 6.2 (Concentration of spectral norms) *For $k \leq n$, let $U \in \mathbb{R}^{n\times k}$ be a random matrix from the standard Gaussian random ensemble (i.e., $U_{ij} \sim N(0,1)$, i.i.d). Then for all $t > 0$, we have*

$$\mathbb{P}\left[\left\|\frac{1}{n} U^T U - I_{k\times k}\right\|_2 \geq \delta(n,k,t)\right] \leq 2\exp\left(-nt^2/2\right), \tag{6.3}$$

where $\delta(n,k,t) := 2\left(\sqrt{\frac{k}{n}} + t\right) + \left(\sqrt{\frac{k}{n}} + t\right)^2$.

Lemma 6.3 *For $k \leq n$, let $X \in \mathbb{R}^{n\times k}$ have i.i.d rows $X_i \sim N(0, \Lambda)$.*

1. *If the covariance matrix* Λ *has maximum eigenvalue* $C_{max} < \infty$*, then for all* $t > 0$*, we have*

$$\mathbb{P}\left[\left\|\frac{1}{n}X^TX - \Lambda\right\|_2 \geq C_{\max}\delta(n,k,t)\right] \leq 2\exp\left(-nt^2/2\right). \tag{6.4}$$

2. *If the covariance matrix* Λ *has minimum eigenvalue* $C_{min} > 0$*, then for all* $t > 0$*, we have*

$$\mathbb{P}\left[\left\|\left(\frac{X^TX}{n}\right)^{-1} - \Lambda^{-1}\right\|_2 \geq \frac{\delta(n,k,t)}{C_{\min}}\right] \leq 2\exp\left(-nt^2/2\right). \tag{6.5}$$

References

E. Candes and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.*, 35(6):2313–2351, 12 2007.

R. J. Tibshirani et al. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55(5):2183–2202, 2009.