# Restricted Strong Convexity of LASSO

2022年2月11日 9:13

Our next task is to prove

**Theorem:** Suppose that the rows of $X$ are iid $\mathcal{N}(0, I_d)$.

Then, with probability $\geq 1 - c_1 \exp(-c_2 n)$,

where $c_1, c_2 > 0$ are constants,

$$\begin{array}{cc} & d \\ n & \left[\; - \; x_j^T \; - \; \right] \\ & X \end{array}$$

$$(**) \qquad \frac{\|X\Delta\|_2}{\sqrt{n}} \geq \frac{1}{4}\|\Delta\|_2 - 9\sqrt{\frac{\log d}{n}}\,\|\Delta\|_1 \qquad \forall \Delta \in \mathbb{R}^d$$

$\|\Delta\|_2 \leq \|\Delta\|_1 \leq \sqrt{d}\,\|\Delta\|_2$

The proof of the theorem has 3 main steps:

**Step 1:** Observe that $(**)$ is trivial if $\Delta = 0$. So we may assume $\|\Delta\|_2 = 1$ by homogeneity. For any $r > 0$, define

$$V(r) = \{ \Delta \in \mathbb{R}^d : \|\Delta\|_2 = 1, \ \|\Delta\|_1 \leq r \}$$

**Proposition 1:**

$$\mathbb{E}\left[ \inf_{\Delta \in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}} \right] \geq 3\left[ \frac{1}{4} - \sqrt{\frac{\log d}{n}}\, r \right]$$

whenever $V(r) \neq \emptyset$.

**Step 2:** We show that the random variable

$$Q(r, X) = \inf_{\Delta \in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}}$$

is concentrated around its mean:

**Proposition 2:** Let $r > 0$ be s.t. $V(r) \neq \emptyset$. Define $t(r) = \frac{1}{4} + 3\sqrt{\frac{\log d}{n}}\, r$. Then,

$$\Pr\left[ \left| Q(r, X) - \mathbb{E}[Q(r, X)] \right| \geq \frac{1}{2} t(r) \right] \leq 2\exp\left(-n t^2(r)/8\right)$$

**Step 3:** The results from Propositions 1 and 2 show that with probability at least $1-2\exp(-nt^2(r)/8)$.

$$Q(r,X) = \inf_{\Delta\in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}} \geq \mathbb{E}[Q(r,X)] - \frac{1}{2}t(r)$$
$$\geq 1 - \frac{3}{2}t(r) = \frac{5}{8} - \frac{9}{2}\sqrt{\frac{\log d}{n}}\, r.$$

This holds for a fixed $r$. We need this to hold for all values of $r$ of interest, but union bound does not quite work here.

Probability Tools Used:

(i) Comparison inequality for Gaussian processes

(ii) Concentration inequality for Lipschitz functions of Gaussian RVs

(iii) "peeling argument" from empirical process theory

Proof of Proposition 1: Recall $X_{ij} \sim \text{iid } \mathcal{N}(0,1)$, $V(r) = \{\Delta : \|\Delta\|_2 = 1, \|\Delta\|_1 \leq r\}$.

We are interested in

$$\tilde{Q}(r,X) = \inf_{\Delta\in V(r)} \|X\Delta\|_2 = \inf_{\Delta\in V(r)} \sup_{\underbrace{u\in S^{n-1}}_{\text{unit sphere}}} \underbrace{u^T X\Delta}_{Y_{u,\Delta}}$$

Note that for each $(u,\Delta) \in S^{n-1}\times V(r)$,

$$Y_{u,\Delta} = u^T X\Delta = \sum_{i,j} u_i \Delta_j X_{ij} \text{ is a mean-zero}$$

Gaussian RV.

To lower bound $\mathbb{E}[\tilde{Q}(r,X)]$, we can upper bound

$$\mathbb{E}[-\tilde{Q}(r,X)] = \mathbb{E}\Big[\sup_{\Delta\in V(r)} \inf_{u\in S^{n-1}} \underbrace{u^T X\Delta}_{Y_{u,\Delta}}\Big]$$

A powerful idea is to construct another family of Gaussian RVs $\{Z_{u,\Delta}\}$ s.t. $\mathbb{E}\Big[\sup \inf Z_{u,\Delta}\Big]$ is (a) easy to

RVs $\{Z_{u,\Delta}\}$ s.t $\mathbb{E}\left[\sup_{\Delta \in V(r)} \inf_{u \in S^{n-1}} Z_{u,\Delta}\right]$ is (a) easy to compute and (b) is related to $\mathbb{E}\left[\sup_{\Delta \in V(r)} \inf_{u \in S^{n-1}} Y_{u,\Delta}\right]$

**Fact** (Gordon's comparison inequality)

Let $U, V$ be arbitrary index sets. Consider two families of zero-mean Gaussian RVs $\{Y_{u,v}\}$ and $\{Z_{u,v}\}$. Suppose that

$$\sigma(Y_{u,v} - Y_{u',v'}) \le \sigma(Z_{u,v} - Z_{u',v'}) \quad \forall \begin{array}{l} u, u' \in U \\ v, v' \in V \end{array}$$

and $\sigma(Y_{u,v} - Y_{u,v'}) = \sigma(Z_{u,v} - Z_{u,v'}) \quad \forall \begin{array}{l} u \in U, \\ v, v' \in V, \end{array}$

where $\sigma(\,)$ denotes the standard deviation function. Then,

$$\mathbb{E}\left[\sup_{u \in U} \inf_{v \in V} Y_{u,v}\right] \le \mathbb{E}\left[\sup_{u \in U} \inf_{v \in V} Z_{u,v}\right].$$

To apply the fact, let us first compute $\sigma(Y_{u,\Delta} - Y_{u',\Delta'})$ and see what properties are needed for $\{Z_{u,\Delta}\}$. By definition,

$$\sigma^2(Y_{u,\Delta} - Y_{u',\Delta'}) = \sigma^2(u^T X \Delta - (u')^T X (\Delta')) \quad \text{(by definition of } Y_{u,\Delta})$$

$$= \mathbb{E}\left[(u^T X \Delta - (u')^T X (\Delta'))^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} \sum_{j=1}^{d} X_{ij} (u_i \Delta_j - u_i' \Delta_j')\right)^2\right]$$

$$\left[\sum_{i,j} \sum_{k,l} \mathbb{E}\left[X_{ij} X_{kl}\right] (\cdot)(\cdot)\right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} (u_i \Delta_j - u_i' \Delta_j')^2$$

$(i,j) \neq (k,l)$ $\quad \mathbb{E}[X_{ij}] \cdot \mathbb{E}[X_{kl}]$ $= 0$

$(i,j) = (k,l)$ $\quad \mathbb{E}[X_{ij}^2] = 1$

$$= \| u\Delta^T - (u')(\Delta')^T \|_F^2$$

Observe that

$$\| u\Delta^T - (u')(\Delta')^T \|_F^2 = \| (u-u')\Delta^T + u'(\Delta - \Delta')^T \|_F^2$$

$$= \|\Delta\|_2^2 \, \|u - u'\|_2^2 + \|u'\|_2^2 \, \|\Delta - \Delta'\|_2^2$$

$$+2\,(u^Tu' - \|u'\|_2^2)(\|\Delta\|_2^2 - \Delta^T\Delta')$$

$$= \|u-u'\|_2^2 + \|\Delta-\Delta'\|_2^2 - 2\,\underline{(1-u^Tu')(1-\Delta^T\Delta')}$$

(since $\underbrace{\|u\|_2 = \|u'\|_2 = 1}_{\because\, u,u' \in S^{n-1}}$, $\underbrace{\|\Delta\|_2 = \|\Delta'\|_2 = 1}_{\because\, \Delta,\Delta' \in V(r)}$)

$$\le \underline{\|u-u'\|_2^2 + \|\Delta-\Delta'\|_2^2}$$

by Cauchy-Schwarz inequality; equality holds when $u=u'$ or $\Delta=\Delta'$.

This suggests that we should define

$$Z_{u,\Delta} = g_1^T u + g_2^T \Delta,$$

where $g_1 \sim \mathcal{N}(0, I_n)$, $g_2 \sim \mathcal{N}(0, I_d)$ are independent.

(Exercise) (i) $Z_{u,\Delta} \sim \mathcal{N}(0, \|u\|_2^2 + \|\Delta\|_2^2)$

(ii) $\sigma^2(Z_{u,\Delta} - Z_{u',\Delta'}) = \|u-u'\|_2^2 + \|\Delta-\Delta'\|_2^2$.

Hence, by Gordon's inequality with $U = V(r)$ and $V = S^{n-1}$, we have

$$\mathbb{E}\left[\sup_{\Delta\in V(r)} \inf_{u\in S^{n-1}} \underbrace{u^T X \Delta}_{Y_{u,\Delta}}\right] \le \mathbb{E}\left[\sup_{\Delta\in V(r)} \inf_{u\in S^{n-1}} \underbrace{(g_1^T u + g_2^T\Delta)}_{Z_{u,\Delta}}\right]$$

$$= \mathbb{E}\left[\underline{\inf_{u\in S^{n-1}} g_1^T u}\right] + \mathbb{E}\left[\sup_{\Delta\in V(r)} g_2^T\Delta\right]$$

$$\le \underbrace{-\mathbb{E}\left[\|g_1\|_2\right]}_{\text{Cauchy-Schwarz}} + \mathbb{E}\left[\sup_{\Delta\in V(r)} \overbrace{\|g_2\|_\infty \underbrace{\|\Delta\|_1}_{\le r}}^{\text{generalized Cauchy-Schwarz}}\right]$$

$\Delta\in V(r) \to \|\Delta\|_2 = 1$, $\|\Delta\|_1 \le r$

$$\le -\mathbb{E}\left[\|g_1\|_2\right] + r\cdot\mathbb{E}\left[\|g_2\|_\infty\right]$$