Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 20

Yun Yang

- High-dimensional linear regression

Exact recovery and restricted nullspace

Theorem

For any fixed subset S, the following two properties are equivalent:

1. *For any $\theta^* \in \mathbb{R}^d$ with support S, the basis pursuit linear program has unique solution $\theta = \theta^*$;*
2. *The matrix X satisfies the restricted nullspace property with respect to S.*

Sufficient conditions for restricted nullspace

The earliest sufficient conditions were based on the incoherence parameter of the design matrix:

$$\delta_{PI}(X) = \max_{j \neq k} \left| \frac{\langle X_j, X_k \rangle}{n} \right|.$$

Property

If the pairwise incoherence satisfies the bound

$$\delta_{PI}(X) \leq \frac{1}{3s},$$

then the restricted nullspace property holds for all subsets S of cardinality at most s.

This condition holds with high probability for sub-Gaussian random matrices with i.i.d. elements as long as $n = \Omega(s^2 \log d)$.

Restricted isometry property (RIP)

Definition

For each $s = 1, \ldots, d$, the restricted isometry constant of $X \in \mathbb{R}^{n \times d}$ of order s is the smallest quantity $\delta_s(X) > 0$ such that

$$\left\| \frac{X_S^T X_S}{n} - I_s \right\|_{\text{op}} \leq \delta_S(X) \quad \text{for all subsets } S \text{ of size at most } s.$$

- Connection to the incoherence parameter: If $X/\sqrt{n}$ has unit-norm columns, then $\delta_{PI}(X) = \delta_2(X)$.
- In general, we have for $s \geq 2$,

$$\delta_{PI}(X) \leq \delta_s(X) \leq s\, \delta_{PI}(X).$$

RIP and restricted nullspace

Property

If the RIP constant of order $2s$ satisfies $\delta_{2s} < 1/3$, then the *uniform restricted nullspace property* holds for any subset S of cardinality $|S| \leq s$.

- The RIP constants for sub-Gaussian random matrices with i.i.d. elements are well-controlled as long as $n = \Omega(s \log(d/s))$.
- Neither the pairwise incoherence condition nor the RIP condition are necessary conditions.
- Counter-example: $\Sigma = (1-\mu)\, I_d + \mu\, \mathbf{1}\mathbf{1}^T$ for $\mu \in (0, 1)$.

Estimation in noisy settings

- Focusing on the observation model

$$y = X\theta^* + w,$$

  where $w \in \mathbb{R}^n$ is the noise vector.

- As a natural extension of the basis pursuit linear program, we consider the *Lasso program*

$$\widehat{\theta} \in \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|y - X\theta\|_2^2 + \lambda_n \|\theta\|_1 \right\}.$$

  Here $\lambda_n > 0$ is a regularization parameter.

- Two different constrained forms that are equivalent to the Lasso.

Restricted eigenvalue condition

- No longer expect to achieve perfect recovery.
- Need a condition slightly stronger than the restricted nullspace property.
- For a constant $\alpha > 1$, define

$$\mathcal{C}_\alpha(S) = \{\Delta \in \mathbb{R}^d : \|\Delta_{S^c}\|_1 \leq \alpha \|\Delta_S\|_1\}.$$

Definition

The matrix X satisfies the *restricted eigenvalue* (RE) condition over S with parameters (κ, α) if

$$\frac{1}{n}\|X\Delta\|_2^2 \geq \kappa \|\Delta\|_2^2 \quad \text{for all } \Delta \in \mathcal{C}_\alpha(S).$$

Bounds on ℓ_2-error

Conditions:

(A1) θ^* is supported on S with $|S| = s$

(A2) X satisfies the restricted eigenvalue condition over S with parameters $(\kappa, 3)$.

Theorem

Under conditions (A1) and (A2), if $\lambda_n \geq 2\|\frac{X^T w}{n}\|_\infty$, then any Lasso solution satisfies

$$\|\widehat{\theta} - \theta^*\|_2 \leq \frac{3}{\kappa}\sqrt{s}\lambda_n, \quad \textit{and}$$

$$\|\widehat{\theta} - \theta^*\|_1 \leq 4\sqrt{s}\,\|\widehat{\theta} - \theta^*\|_2.$$

Example: Classical linear Gaussian model

Consider deterministic design regression

$$y = X\theta + w, \quad w \sim \mathcal{N}(0, \sigma^2 I_n),$$

where $X \in \mathbb{R}^{n \times d}$ is fixed.

Suppose X satisfies the RE condition, and is C-column normalized, meaning

$$\max_j \frac{\|X_j\|_2}{\sqrt{n}} \leq C.$$

By standard Gaussian tail bounds, we have

$$\mathbb{P}\Big[\Big\|\frac{X^T w}{n}\Big\|_\infty \geq C\sigma\Big(\sqrt{\frac{2\log d}{n}} + \delta\Big)\Big] \leq 2\, e^{-\frac{n\delta^2}{2}}, \quad \text{for all } \delta > 0.$$

Example: Classical linear Gaussian model

Property

If $\lambda_n = 2C\,\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)$, then any Lasso solution satisfies

$$\|\widehat{\theta} - \theta^*\|_2 \leq \frac{6C\,\sigma}{\kappa}\,\sqrt{s}\left(\sqrt{\frac{2\log d}{n}} + \delta\right)$$

with probability at least $1 - 2\,e^{-\frac{n\delta^2}{2}}$.

Example: Compressed sensing

The design matrix X can be chosen by the user, and one standard choice is the standard Gaussian matrix with i.i.d. $\mathcal{N}(0,1)$ entries.

Suppose the noise vector $w \in \mathbb{R}^n$ is deterministic, with bounded entries $\|w\|_\infty \leq \sigma$.

Then each variable $X_j^T w/\sqrt{n}$ is zero-mean sub-Gaussian with parameter σ^2. We have a similar finite sample error bound as in the previous example.

Proof outline

- Denote the objective function by

$$L(\theta;\, \lambda_n) = \frac{1}{2n}\|y - X\theta\|_2^2 + \lambda_n\|\theta\|_1.$$

- Since $\widehat{\theta}$ minimizes $L(\theta;\, \lambda_n)$, we have

$$L(\widehat{\theta};\, \lambda_n) \le L(\theta^*;\, \lambda_n).$$

- Let $\widehat{\Delta} = \widehat{\theta} - \theta^*$. Re-arranging yields the basic inequality

$$0 \le \frac{1}{2n}\|X\widehat{\Delta}\|_2^2 \le \frac{w^T X\widehat{\Delta}}{n} + \lambda_n\big(\|\theta^*\|_1 - \|\widehat{\theta}\|_1\big).$$

- If $\lambda_n \ge 2\|\frac{X^T w}{n}\|_\infty$, then this leads to $\widehat{\Delta} \in \mathcal{C}_3(S)$, and

$$\kappa\,\|\widehat{\Delta}\|_2^2 \le 3\lambda_n\sqrt{s}\|\widehat{\Delta}\|_2.$$