HDP
(25) Sparse Recovery

@ 滕佳烨

前情回顾

M* bound: For the intersect of a set T and a random space, its diam can be very small w.h.p.

Escaping Theorem: with a large m, the intersection can be empty w.h.p.

Recovery Problem

Signal $x \in R^n$
(noisy measurement) $\mathrm{Y} = Ax + w \in R^m$ $(m \ll n)$
Goal: recover x from (A, Y)

Different from regression, the goal is to recover x instead of estimating coefficient (but similarly).
Besides, we can usually choose A to meet some property (instead, the distribution of x in fixed in linear regression)

Note that when $m \ll n$, there can be many solutions (ill-posed).
Therefore, we need to add some prior information on signal x.

Noiseless Recovery Problem

Signal $x \in R^n$, (noiseless measurement) Y $= Ax \cancel{+ w} \in R^m$ $(m \ll n)$
Prior information on x: $x \in T$

All the solutions forms a kernel space of A $(x \in x' + Ker(A))$
Kernel space (with a shift)? M^* bound?

Theorem 10.2.1. *Suppose the rows A_i of A are independent, isotropic and sub-gaussian random vectors. Then any solution $\hat{x}$ of the program (10.4) satisfies*

$$\mathbb{E}\,\|\hat{x} - x\|_2 \leq \frac{CK^2 w(T)}{\sqrt{m}},$$

where $K = \max_i \|A_i\|_{\psi_2}$.

$$\boxed{\text{find } x' : \; y = Ax', \quad x' \in T. \qquad (10.4)}$$

Every point that satisfies Ax=Y is close to the true signal x.

Noisy Recovery Problem

Signal $x \in R^n$, (noisy measurement) $Y = Ax + w \in R^m$ $(m \ll n)$
Prior information on x: $x \in T$

Exercise 10.2.4 (Noisy measurements). Extend the recovery result (Theorem 10.2.1) for the noisy model $y = Ax + w$ we considered in (10.1). Namely, show that

$$\mathbb{E}\,\|\widehat{x} - x\|_2 \leq \frac{CK^2 w(T) + \|w\|_2}{\sqrt{m}}.$$

$$\boxed{\text{find } x' : \; y = Ax', \quad x' \in T. \qquad (10.4)}$$

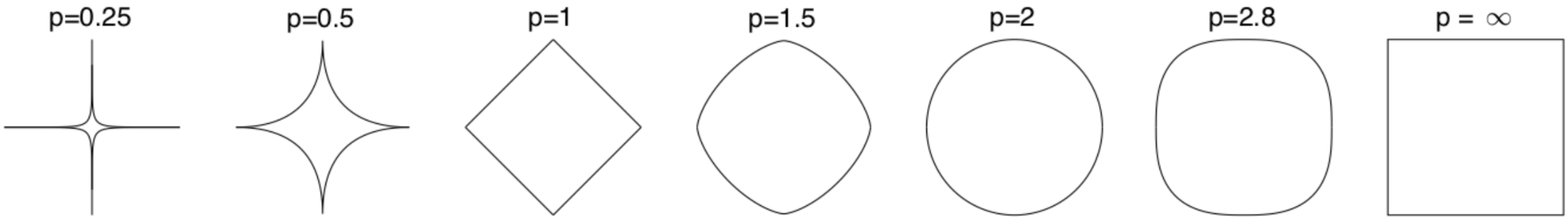
Hint: although x' does not satisfy $Ax' = Ax$, it is close to x^0 (any point with $Ax^0 = Ax$, choose a proper one) since $A\left(x' - x^0\right) = w$

The prior T: Sparsity (noiseless)

$$T = \{x \in \mathbb{R}^n : \ \|x\|_0 \leq s\}\,.$$

The problem is well posed when $m > 2\|x\|_0$
However, it is hard to calculate since T is non-convex.
Therefore, we want to relax T to its convex version.

$$\{x \in \mathbb{R}^n : \ \|x\|_0 \leq s, \ \|x\|_2 \leq 1\} \subset \sqrt{s} B_1^n\,.$$



The prior T: Sparsity (noiseless)

$$\{x \in \mathbb{R}^n : \ \|x\|_0 \le s, \ \|x\|_2 \le 1\} \subset \sqrt{s} B_1^n.$$

Corollary 10.3.4 (Sparse recovery: guarantees). *Assume the unknown s-sparse signal $x \in \mathbb{R}^n$ satisfies $\|x\|_2 \le 1$. Then x can be approximately recovered from the random measurement vector $y = Ax$ by a solution $\widehat{x}$ of the program (10.6). The recovery error satisfies*

$$\mathbb{E}\,\|\widehat{x} - x\|_2 \le CK^2 \sqrt{\frac{s \log n}{m}}.$$

Note that the approximation is almost tight!
(the gaussian width is optimal up to a constant).

Low-rank Matrix Recovery

Signal $x \in R^{d*d}$, (noiseless measurement) $Y_i = A_i \cdot X \in R$ $(m \ll n)$
Prior information on X: $Rank(X) \le r$ [c.f. $\|x\|_0$ in vector]

Nuclear norm (c.f. $\|x\|_1$) $$\|X\|_* := \|s(X)\|_1 = \sum_{i=1}^{d} s_i(X) = \operatorname{tr}\left(\sqrt{X^\intercal X}\right)$$

$$\left\{X \in \mathbb{R}^{d\times d} : \ \operatorname{rank}(X) \le r, \ \|X\|_F \le 1\right\} \subset \sqrt{r}B_*.$$

Exercise 10.4.5 (Low-rank matrix recovery: guarantees). Suppose the random matrices A_i are independent and have all independent, sub-gaussian entries.[1] Assume the unknown $d \times d$ matrix X with rank r satisfies $\|X\|_F \le 1$. Show that X can be approximately recovered from the random measurements y_i by a solution $\widehat{X}$ of the program (10.11). Show that the recovery error satisfies

$$\mathbb{E}\,\|\widehat{X} - X\|_F \le CK^2\sqrt{\frac{rd}{m}}.$$

Hint: use the Gaussian width of nuclear norm ball.

Exact sparse recovery (noiseless)

Theorem 10.5.1 (Exact sparse recovery)**.** *Suppose the rows A_i of A are independent, isotropic and sub-gaussian random vectors, and let $K := \max_i \|A_i\|_{\psi_2}$. Then the following happens with probability at least $1 - 2\exp(-cm/K^4)$.*

Assume an unknown signal $x \in \mathbb{R}^n$ is s-sparse and the number of measurements m satisfies

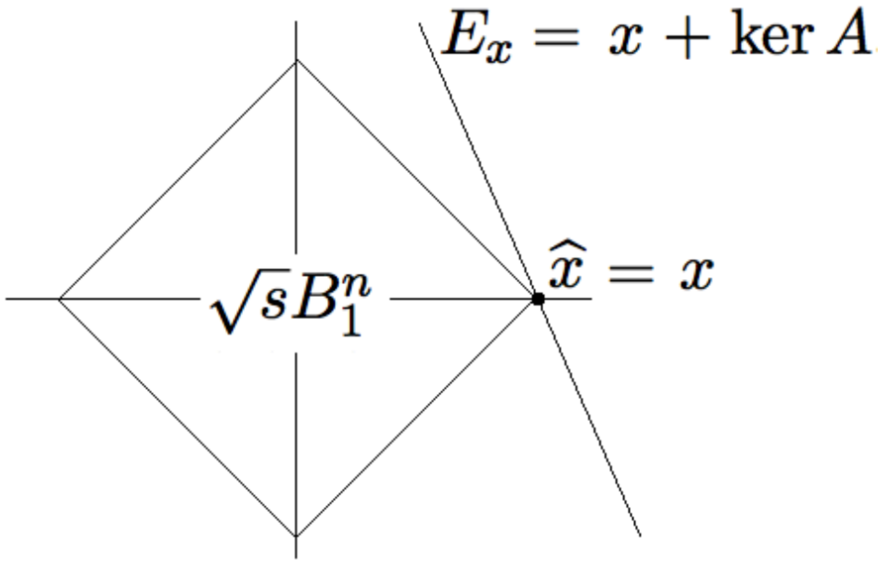
$$m \geq CK^4 s \log n.$$

Then a solution $\widehat{x}$ of the program (10.12) is exact, i.e.

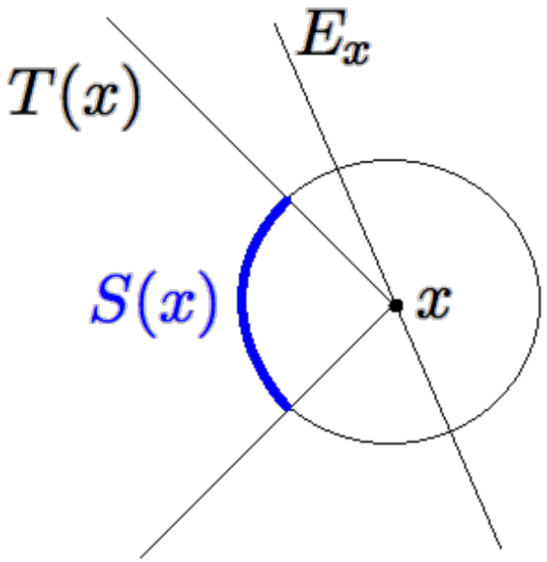
$$\widehat{x} = x.$$

When m (#samples) is large, the signal can be recovered perfectly.

Exact sparse recovery (noiseless, intuition)



(a) Exact sparse recovery happens when the random subspace E_x is tangent to the ℓ_1 ball at the point x.



(b) The tangency occurs iff E_x is disjoint from the spherical part $S(x)$ of the tangent cone $T(x)$ of the ℓ_1 ball at point x.

Hint: prove that the residual h $= x - \hat{x}$ satisfying $\|h\|_1 \leq \sqrt{2s}\,\|h\|_2$

Exact sparse recovery (noiseless, general RIP condition)

Definition 10.5.8 (RIP). An $m \times n$ matrix A satisfies the *restricted isometry property* (RIP) with parameters α, β and s if the inequality

$$\alpha \|v\|_2 \leq \|Av\|_2 \leq \beta \|v\|_2$$

holds for all vectors $v \in \mathbb{R}^n$ such that[2] $\|v\|_0 \leq s$.

Intuition: the matrix A is approximate isometric, closely related to eigenvalues.

Theorem 10.5.10 (RIP implies exact recovery). *Suppose an $m \times n$ matrix A satisfies RIP with some parameters α, β and $(1+\lambda)s$, where $\lambda > (\beta/\alpha)^2$. Then every s-sparse vector $x \in \mathbb{R}^n$ can be recovered exactly by solving the program (10.12), i.e. the solution satisfies*

$$\widehat{x} = x.$$

Exact recovery (noiseless, general RIP condition)

Random matrix meets the RIP condition.

Theorem 10.5.11 (Random matrices satisfy RIP). *Consider an $m \times n$ matrix A whose rows A_i of A are independent, isotropic and sub-gaussian random vectors, and let $K := \max_i \|A_i\|_{\psi_2}$. Assume that*

$$m \geq CK^4 s \log(en/s).$$

Then, with probability at least $1 - 2\exp(-cm/K^4)$, the random matrix A satisfies RIP with parameters $\alpha = 0.9\sqrt{m}$, $\beta = 1.1\sqrt{m}$ and s.

LASSO

$$\text{minimize } \|y - Ax'\|_2 \text{ s.t. } \|x'\|_1 \le R.$$

Theorem 10.6.1 (Performance of Lasso). *Suppose the rows A_i of A are independent, isotropic and sub-gaussian random vectors, and let $K := \max_i \|A_i\|_{\psi_2}$. Then the following happens with probability at least $1 - 2\exp(-s \log n)$.*

Assume an unknown signal $x \in \mathbb{R}^n$ is s-sparse and the number of measurements m satisfies

$$m \ge CK^4 s \log n. \tag{10.23}$$

Then a solution $\widehat{x}$ of the program (10.22) with $R := \|x\|_1$ is accurate, namely

$$\|\widehat{x} - x\|_2 \le C\sigma \sqrt{\frac{s \log n}{m}},$$

where $\sigma = \|w\|_2 / \sqrt{m}$.

Remark 10.6.3 (Exact recovery). In the noiseless model $y = Ax$ we have $w = 0$ and thus Lasso recovers x exactly, i.e.

$$\widehat{x} = x.$$

Take-away Messages (noiseless setting)

Vector recovery problem:

- $\ell_0 + \ell_2 \approx \ell_1$
- Error bound $\sqrt{s \log n / m}$ [M* bound]
- Exact recovery $m \sim s \log n$ [Escaping Theorem]

Matrix recovery problem:

- $\| \cdot \|_0 + \| \cdot \|_F \approx \| \cdot \|_*$
- RIP condition (random matrix)
- Error bound $\sqrt{rd/m}$

LASSO (noisy):

- Error bound $\sigma\sqrt{s \log n / m}$

@ 滕佳烨

Thanks!