HDP
(26) Dvoretzky-Milman’s Theorem

@ 滕佳烨

前情回顾：sparse recovery

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 $\sigma\sqrt{s \log n / m}$

前情回顾：Matrix Deviation Inequality

Can we replace the norm with general norm?

Theorem 9.1.1 (Matrix deviation inequality). *Let A be an $m \times n$ matrix whose rows A_i are independent, isotropic and sub-gaussian random vectors in $\mathbb{R}^n$. Then for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \boxed{\|Ax\|_2 - \sqrt{m}\|x\|_2} \right| \le CK^2 \gamma(T).$$

Here $\gamma(T)$ is the Gaussian complexity introduced in Section 7.6.2, and $K = \max_i \|A_i\|_{\psi_2}$.

General Matrix Deviation Inequality (with general norm):

Positive-homogeneous: $f(\alpha x) = \alpha f(x)$
Sub-additive: $f(x + y) \leq f(x) + f(y)$

Example: norm, $f(x) = x \cdot y$, $f(x) = \sup_{y \in S} x \cdot y$

Theorem 11.1.5 (General matrix deviation inequality). *Let A be an $m \times n$ Gaussian random matrix with $\boxed{\text{i.i.d. } N(0,1) \text{ entries.}}$ Let $f : \mathbb{R}^m \to \mathbb{R}$ be a positive-homogeneous and subadditive function, and let $b \in \mathbb{R}$ be such that*

$$\boxed{f(x) \leq b\|x\|_2} \text{ for all } x \in \mathbb{R}^n. \tag{11.3}$$

Then for any subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} \left| \boxed{f(Ax) - \mathbb{E} f(Ax)} \right| \leq Cb\gamma(T).$$

Here $\gamma(T)$ is the Gaussian complexity introduced in Section 7.6.2.

General Matrix Deviation Inequality (with general norm):

Theorem 11.1.5 (General matrix deviation inequality). *Let A be an $m \times n$ Gaussian random matrix with i.i.d. $N(0,1)$ entries. Let $f : \mathbb{R}^m \to \mathbb{R}$ be a positive-homogeneous and subadditive function, and let $b \in \mathbb{R}$ be such that*

$$\boxed{f(x) \le b\|x\|_2} \quad \text{for all } x \in \mathbb{R}^n. \tag{11.3}$$

Then for any subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} \left| \boxed{f(Ax) - \mathbb{E}\, f(Ax)} \right| \le Cb\gamma(T).$$

Here $\gamma(T)$ is the Gaussian complexity introduced in Section 7.6.2.

Remark 11.1.9. It is an open question if Theorem 11.1.5 holds for general sub-gaussian matrices A.

Johnson-Lindenstrauss Lemma (general norm)

Exercise 11.2.2 (Johnson-Lindenstrauss Lemma for ℓ_1 norm). ☕☕ Specialize the previous exercise to the ℓ_1 norm. Thus, let $\mathcal{X}$ be a set of N points in $\mathbb{R}^n$, let A be an $m \times n$ Gaussian matrix with i.i.d. $N(0,1)$ entries, and let $\varepsilon \in (0,1)$.

Suppose that

$$m \geq C(\varepsilon) \log N.$$

Show that with high probability the matrix $Q := \sqrt{\pi/2} \cdot m^{-1} A$ satisfies

$$(1-\varepsilon)\|x-y\|_2 \leq \|Qx - Qy\|_1 \leq (1+\varepsilon)\|x-y\|_2 \quad \text{for all } x, y \in \mathcal{X}.$$

Chevet's inequality (general norm)

Theorem 11.2.4 (General Chevet's inequality). *Let A be an $m \times n$ Gaussian random matrix with i.i.d. $N(0,1)$ entries. Let $T \subset \mathbb{R}^n$ and $S \subset \mathbb{R}^m$ be arbitrary bounded sets. Then*

$$\mathbb{E} \sup_{x \in T} \left| \sup_{y \in S} \langle Ax, y \rangle - w(S)\|x\|_2 \right| \leq C\gamma(T)\,\mathrm{rad}(S).$$

$$\mathbb{E} \sup_{x \in T,\, y \in S} \langle Ax, y \rangle \leq CK\left[w(T)\,\mathrm{rad}(S) + w(S)\,\mathrm{rad}(T)\right]$$

Dvoretzky-Milman's Theorem

Theorem 11.3.3 (Dvoretzky-Milman's theorem: Gaussian form). *Let A be an $m \times n$ Gaussian random matrix with i.i.d. $N(0,1)$ entries, $T \subset \mathbb{R}^n$ be a bounded set, and let $\varepsilon \in (0,1)$. Suppose*

$$m \le c\varepsilon^2 d(T)$$

where $d(T)$ is the stable dimension of T introduced in Section 7.6. Then with probability at least 0.99, we have

$$(1-\varepsilon)B \subset \mathrm{conv}(AT) \subset (1+\varepsilon)B$$

where B is a Euclidean ball with radius $w(T)$.

1. Random Gaussian projection of cubes onto subspace m~n is close to round balls.
2. Convex hull of Gaussian cloud is approximately Euclidean ball $\sqrt{\log n}$.

Dvoretzky-Milman's Theorem (Comparison)

1. Phase transition:

$$\operatorname{diam}(PT) \le \begin{cases} C\sqrt{\frac{m}{n}}\operatorname{diam}(T), & \text{if } m \ge d(T) \\ Cw_s(T), & \text{if } m \le d(T). \end{cases}$$

Part A: JL Lemma

Proposition 9.3.2 (Additive Johnson-Lindenstrauss Lemma)**.** *Consider a set $\mathcal{X} \subset \mathbb{R}^n$. Let A be an $m \times n$ matrix whose rows A_i are independent, isotropic and sub-gaussian random vectors in $\mathbb{R}^n$. Then, with high probability (say,* 0.99*), the scaled matrix*

$$Q := \frac{1}{\sqrt{m}} A$$

satisfies

$$\|x - y\|_2 - \delta \le \|Qx - Qy\|_2 \le \|x - y\|_2 + \delta \quad \textit{for all } x, y \in \mathcal{X}$$

where

$$\delta = \frac{CK^2 w(\mathcal{X})}{\sqrt{m}}$$

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

Dvoretzky-Milman's Theorem (Comparison)

1. Phase transition:

$$\operatorname{diam}(PT) \leq \begin{cases} C\sqrt{\dfrac{m}{n}}\,\operatorname{diam}(T), & \text{if } m \geq d(T) \\ Cw_s(T), & \text{if } m \leq d(T). \end{cases}$$

Part B: DM Theorem

Exercise 11.3.9 (Random projection in the Grassmanian)**.** ☕☕☕ Prove a version of Dvoretzky-Milman's theorem for the projection P onto a random m-dimensional subspace in $\mathbb{R}^n$. Under the same assumptions, the conclusion should be that

$$(1-\varepsilon)B \subset \operatorname{conv}(PT) \subset (1+\varepsilon)B$$

Take-away Messages

General Matrix Deviation Inequality: extend the norm to the general norm.

DM Theorem: random project to a ball

@ 滕佳烨

Thanks!