HDP
(15) Concentration of Quadratic forms

@ 滕佳烨

前情回顾:

f(X) → $\mathbb{E}$f(X) with f Lipschitz and X~$\mathbb{P}$

- sphere with Euclidean metric
- Gaussian RV with Euclidean metric
- Bernoulli RV with Hamming metric

→ Matrix Bernstein Inequality

→ Johnson-Lindenstrauss Lemma

f(X) Lipschitz is sufficient but not necessary

Consider a specific form (quadratic) $f(X) = X^T AX$

Honson-Wright Inequality

Theorem 6.2.1 (Hanson-Wright inequality). *Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent, mean zero, sub-gaussian coordinates. Let A be an $n \times n$ matrix. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\left\{|X^\top AX - \mathbb{E}\, X^\top AX| \geq t\right\} \leq 2\exp\left[-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right],$$

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

1. X is sub-Gaussian, X^2 is sub-Expenential
2. $X^\top AX$ is not Lipschitz w.r.t. X
3. Proof Hint: decoupling ($X^\top AX \to X^\top AX'$)

Decoupling ($X^\top AX \to X^\top AX'$)

1. Note that the bottleneck of analyzing $X^\top AX$ is the dependency between $a_{ij}x_ix_j$ and $a_{ik}x_ix_k$
2. By using the independent copy of X, we analyze $X^\top AX'$ instead of $X^\top AX$ by the following Lemma

Theorem 6.1.1 (Decoupling). *Let A be an $n \times n$, $\boxed{\textit{diagonal-free matrix}}$ (i.e. the diagonal entries of A equal zero). Let $X = (X_1, \dots, X_n)$ be a random vector with independent mean zero coordinates X_i. Then, for every convex function $F : \mathbb{R} \to \mathbb{R}$, one has*

$$\mathbb{E}\, F(X^\top AX) \le \mathbb{E}\, F(4X^\top AX') \tag{6.3}$$

where X' is an independent copy of X.

Set $F(x) = \exp \lambda x$ →MGF → (Markov) Tail Bound

Finish the proof

1. Diag entries: directly bound $a_{ii}x_i^2$
2. Off-diag entries:
   1. using decoupling and bound its MGF [$\mathbb{E}\exp \lambda X$, convex]
   2. Gaussian is worst-case of sub-Gaussian
   3. Calculate the bound using Gaussian RV

Lemma 6.2.3 (Comparison). *Consider independent, mean zero, sub-gaussian random vectors X, X' in $\mathbb{R}^n$ with $\|X\|_{\psi_2} \le K$ and $\|X'\|_{\psi_2} \le K$. Consider also independent random vectors $g, g' \sim N(0, I_n)$. Let A be an $n \times n$ matrix. Then*

$$\mathbb{E}\exp(\lambda X^\mathsf{T} A X') \le \mathbb{E}\exp(CK^2 \lambda g^\mathsf{T} A g')$$

for any $\lambda \in \mathbb{R}$.

Lemma 6.2.2 (MGF of Gaussian chaos). *Let $X, X' \sim N(0, I_n)$ be independent and let $A = (a_{ij})$ be an $n \times n$ matrix. Then*

$$\mathbb{E}\exp(\lambda X^\mathsf{T} A X') \le \exp(C\lambda^2 \|A\|_F^2)$$

for all λ satisfying $|\lambda| \le c/\|A\|$.

Honson-Wright Inequality (high dim)

Theorem 6.2.1 (Hanson-Wright inequality). *Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent, mean zero, sub-gaussian coordinates. Let A be an $n \times n$ matrix. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\left\{|X^\mathsf{T} A X - \mathbb{E}\, X^\mathsf{T} A X| \geq t\right\} \leq 2\exp\left[-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right],$$

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

Exercise 6.2.7 (Higher-dimensional Hanson-Wright inequality). Let $X_1, \ldots, X_n$ be independent, mean zero, sub-gaussian random vectors in $\mathbb{R}^d$. Let $A = (a_{ij})$ be an $n \times n$ matrix. Prove that for every $t \geq 0$, we have

$$\mathbb{P}\left\{\Big|\sum_{i,j:\, i\neq j}^{n} a_{ij}\,\langle X_i, X_j\rangle\Big| \geq t\right\} \leq 2\exp\left[-c\min\left(\frac{t^2}{K^4 d\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right]$$

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

Take-away Messages

1. Honson-Wright Inequality: quadratic form concentration
2. The bound is similar to sub-exponential
3. Decoupling: $X^\top AX \to X^\top AX'$
4. Honson-Wright Inequality can be used in deriving the concentration: $\|BX\|_2 \to \|B\|_F$, see (10)

@ 滕佳烨

Thanks!