MATH888: High-dimensional probability and statistics Fall 2021

Lecture 10 — September 29, 2021

Sebastien Roch, UW-Madison *Scribe: Zaidan Wu*

1 Overview

In the last lecture we reviewed the properties of conditional expectations, proved a useful result in Lemma 1, lecture 9:

$$\mathbb{E}F(Y) \le \mathbb{E}F(Y+Z),$$

where function F is convex, Y and Z are independent, $\mathbb{E}Z = 0$. Furthermore, we introduced the symmetrization method with corresponding lemma, and discussed the main idea of proving Hanson-Wright inequality.

The proof of Hanson-Wright inequality relies on two steps, the decoupling step and the comparison step. In this lecture we will prove a helpful result for Hanson-Wright inequality at each step.

2 Main Section

Our aim is to proof Hanson-Wright inequality inequality, let's review the theorem.

Theorem 1. *(Theorem 6.2.1 in [1] Hanson-Wright inequality) Let* $X = (X_1, X_2, ...X_n) \in \mathbb{R}^n$ *be a random vector with independent, mean-zero, sub-gaussian coordinates. Let* A *be an* $n \times n$ *deterministic matrix. Then, for every* $t \ge 0$*, we have*

$$\mathbb{P}\{|X^T AX - \mathbb{E}X^T AX| \ge t\} \le 2exp[-c\ min(\frac{t^2}{K^4||A||_F^2}, \frac{t}{K^2||A||})]$$

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

2.1 Decoupling

Theorem 2 (Theoeorem 6.1.1 in [1])**.** *Let* A *be an* $n \times n$ *deterministic matrix with* 0 *diagonal, that is,* $a_{ii} = 0$*, for all* $i = 1, 2, \dots, n$*. Let* $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ *be a random vector with independent, mean-zero coordinates* X_i*. Then for every convex function* F

$$\mathbb{E}\Big(F(X^T AX)\Big) \le \mathbb{E}\Big(F(4X^T AX')\Big),$$

where X' *is an independent copy of* X*.*

Proof. The notation $[n]$ represents the set $\{1, 2, \ldots, n\}$. Note that for any subset of indices $I \subset [n]$

$$\sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j \overset{d}{=} \sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j^{'}, \tag{1}$$

where $I^c = [n] \setminus I$. Because equation (1) can be rewritten as $X_I^T A X_{I^c} \overset{d}{=} X_I^T A X_{I^c}'$, where the i-th coordinate of vector X_I is X_i if $i \in I$, otherwise, it is 0.

Let δ_i $i = 1, 2, \ldots, n$ be independent fair Bernoulli random variables and they are independent of X and $X^{'}$. Namely, $\mathbb{P}(\delta_i = 0) = \mathbb{P}(\delta_i = 1) = \frac{1}{2}$. Now, set $I = \{i : \delta_i = 1, i \in [n]\}$, the set I is a discrete random subset of $[n]$. Note

$$\begin{aligned}
\mathbb{E}\Big(\sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j | X\Big) &= \mathbb{E}\Big(\sum_{i\neq j} \delta_i (1-\delta_j) a_{ij} X_i X_j | X\Big) \quad (\delta_i(1-\delta_j) \neq 0 \text{ iff } (i,j) \in I \times I^c) \\
&= \sum_{i\neq j} a_{ij} X_i X_j \mathbb{E}\big(\delta_i(1-\delta_j)\big) \quad (\delta_i \text{ are independent of } X) \\
&= \sum_{i\neq j} \frac{1}{4} a_{ij} X_i X_j = \frac{1}{4} X^T A X. \quad (\mathbb{E}\delta_i = \frac{1}{2})
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\mathbb{E}\Big(F\big(4 \sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j^{'}\big)\Big) &= \mathbb{E}\Big(F\big(4 \sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j\big)\Big) \quad \text{(By equation (1))} \\
&= \mathbb{E}\Big[\mathbb{E}\Big(F\big(4 \sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j\big) | X\Big)\Big] \quad \text{(Tower property)} \\
&\geq \mathbb{E}\Big[F\Big(\mathbb{E}\big(4 \sum_{(i,j)\in I\times I^c} a_{ij} X_i X_j | X\big)\Big)\Big] \quad \text{(Jensen's inequality)} \\
&= \mathbb{E}\Big(F(X^T A X)\Big). \quad \text{(Tower property and previous result)}
\end{aligned}$$

Conditional on random set I on left hand side, since δ_i are independent of X and $X^{'}$, and I is discrete, the average (the expectation) satisfies the inequality implies there exist a deterministic set $J \subset [n]$ such that

$$\mathbb{E}\Big(F\big(4 \sum_{(i,j)\in J\times J^c} a_{ij} X_i X_j^{'}\big)\Big) \geq \mathbb{E}\Big(F(X^T A X)\Big). \tag{2}$$

To compute the sum

$$X^T A X^{'} = \sum_{(i,j)\in J\times J^c} a_{ij} X_i X_j^{'} + \sum_{(i,j)\in J\times J} a_{ij} X_i X_j^{'} + \sum_{(i,j)\in J^c\times [n]} a_{ij} X_i X_j^{'}.$$

Set $Y = \sum_{(i,j)\in J\times J^c} a_{ij} X_i X_j^{'}$, $Z_1 = \sum_{(i,j)\in J\times J} a_{ij} X_i X_j^{'}$, and $Z_2 = \sum_{(i,j)\in J^c\times [n]} a_{ij} X_i X_j^{'}$.

Let $\mathcal{G} = \sigma\Big((X_i)_{i\in J}, (X_i^{'})_{i\in J^c}\Big)$, then Y is $\mathcal{G}$-measurable, and

$$\mathbb{E}(Z_1|\mathcal{G}) = \mathbb{E}(Z_2|\mathcal{G}) = 0,$$

since for any $(i,j) \in J \times J$, $\mathbb{E}(a_{ij}X_iX_j^{'}|\mathcal{G}) = a_{ij}X_i\mathbb{E}(X_j^{'}) = 0$ by tower property and $X_j^{'}$ is independent of $\mathcal{G}$. For other cases, by similar argument, we can obtain the zero mean conclusion as well.

Using Jensen's inequality to conclude that

$$\mathbb{E}\Big(F(4Y + 4Z_1 + 4Z_2)|\mathcal{G}\Big) \geq F\Big(\mathbb{E}(4Y + 4Z_1 + 4Z_2|\mathcal{G})\Big) = F(4Y).$$

Taking expectation for both sides and we combine inequality (2), we get

$$\mathbb{E}\Big(F(4X^TAX^{'})\Big) = \mathbb{E}\Bigg[\mathbb{E}\Big(F(4Y + 4Z_1 + 4Z_2)|\mathcal{G}\Big)\Bigg] \geq \mathbb{E}\big(F(4Y)\big) \geq \mathbb{E}\Big(F(X^TAX)\Big). \quad \square$$

2.2 Comparison

Theorem 3 (Lemma 6.2.3 in [1])**.** *Let A be an $n \times n$ matrix and $X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent, sub-Gaussian, mean-zero coordinates X_i. Let independent vectors $g, g^{'} \sim N(0, I_n)$, and assume $K = \max_i \|X_i\|_{\Psi_2} < \infty$. Then there exists a constant C such that*

$$\mathbb{E}\exp(\lambda X^TAX^{'}) \leq \mathbb{E}\exp(CK^2\lambda g^TAg^{'}),$$

for any $\lambda \in \mathbb{R}$, where $X^{'}$ is independent copy of X and independent of g and $g^{'}$.

Proof. Write $Y = (Y_1, Y_2, \ldots, Y_n) = AX^{'}$, then by independent of X_i we get

$$\mathbb{E}\Big(\exp(\lambda X^TAX^{'})|X^{'}\Big) = \mathbb{E}\Big(\exp(\lambda\sum_{i=1}^{n}X_iY_i)|X^{'}\Big) = \prod_{i=1}^{n}\mathbb{E}\Big(\exp(\lambda X_iY_i)|X^{'}\Big). \tag{3}$$

$K = \max_i \|X_i\|_{\Psi_2} < \infty$ and equivalent property of sub-Gaussian implies that there exists constant c such that

$$\prod_{i=1}^{n}\mathbb{E}\Big(\exp(\lambda X_iY_i)|X^{'}\Big) \leq \prod_{i=1}^{n}\exp(cK^2\lambda^2Y_i^2) = \exp(cK^2\lambda^2\|AX^{'}\|_2^2).$$

In equation (3), replace X by g, and compute by generating function of normal random variables we get

$$\mathbb{E}\Big(\exp(\mu g^TAX^{'})|X^{'}\Big) = \exp(\frac{1}{2}\mu^2\|AX^{'}\|_2^2).$$

Choosing $\mu = \sqrt{2c}K\lambda$, we match two results

$$\mathbb{E}\exp(\lambda X^TAX^{'}) \leq \mathbb{E}\exp(\sqrt{2c}K\lambda g^TAX^{'}).$$

Comparing to our statement, we replaced X by g and payed a factor of $\sqrt{2c}K$. In addition, by symmetry, and arguing in a similar way, we can replace $X^{'}$ with $g^{'}$ and pay an extra factor of $\sqrt{2c}K$. Then the proof is complete and $C = 2c$. $\square$

References

[1] R.Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge university press, 2008.