MATH888: High-dimensional probability and statistics Fall 2021

Lecture 10 — September 29, 2021

Sebastien Roch, UW-Madison *Scribe: Liam Johnston*

1 Overview

In the last lecture we stated the Hanson-Wright inequality.

In this lecture we explore some useful tricks that will be helpful in proving the Hanson-Wright inequality.

Theorem 1 (Hanson-Wright inequality (Thm 6.2.1. in Vershynin))**.** *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ 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*

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

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

Recall the following lemma, which we apply later today,

Lemma 2 (Pulled from previous lecture)**.** *Let Y and Z be independent random variables with $\mathbb{E}Z = 0$. Then for every convex function F*

$$\mathbb{E}\big[F(y)\big] \leq \mathbb{E}\big[F(Y + Z)\big]$$

2 Useful Tricks for Proving Hanson-Wright Inequality

Setting: Let A be an $n \times n$ matrix and $\underset{\sim}{X} = (X_1, \dots, X_n)$ where coordinates are independent, zero-mean and sub-gaussian. Our objective is to prove a concentration inequality for $\underset{\sim}{X}^T A \underset{\sim}{X}$.

2.1 Idea 1: Decoupling

Theorem 3 (Decoupling (Thm 6.1.1. in Vershynin))**.** *Let A be an $n \times n$, diagonal-free matrix (i.e. the diagonal entries of A equal zero). Let $\underset{\sim}{X} = (X_1, \dots, X_n)$ be a random vector with independent mean zero coordinates X_i. Then, for every convex function $F : \mathbb{R} \to \mathbb{R}$, one has*

$$\mathbb{E}F(\underset{\sim}{X}^T A \underset{\sim}{X}) \le \mathbb{E}F(4\underset{\sim}{X}^T A \underset{\sim}{X}')$$

where $\underset{\sim}{X}'$ is an independent copy of $\underset{\sim}{X}$.

Proof. **Note:** For any subset of indices $I \subseteq [n]$

$$\sum_{(i,j) \in I \times I^C} a_{ij} X_i X_j \overset{\mathcal{D}}{=} \sum_{(i,j) \in I \times I^C} a_{ij} X_i X_j'$$

because coordinates are independent.

Let

$$\delta_i = \begin{cases} 0 & w.p.\ 0.5 \\ 1 & w.p.\ 0.5 \end{cases} \quad \forall i \in [n] \text{ independently}$$

and $I = \{i : \delta_i = 1\}$.

Note:

$$\begin{aligned} \mathbb{E}\Big[\sum_{(i,j)\in I\times I^C} a_{ij}X_iX_j \big| \underset{\sim}{X}\Big] =& \mathbb{E}\Big[\sum_{i\neq j} \delta_i(1-\delta_j)a_{ij}X_iX_j \big| \underset{\sim}{X}\Big] \\ =& \sum_{i \neq j} a_{ij}X_iX_j \mathbb{E} \underbrace{\big[\delta_i(1-\delta_j)\big]}_{i\neq j \implies \delta_i \text{ independent of } \delta_j} \\ =& \sum_{i\neq j} \frac{a_{ij}X_iX_j}{4} \end{aligned}$$

We now turn our attention to

$$\begin{aligned} \mathbb{E}\Big[F(4\sum_{(i,j)\in I\times I^C} a_{ij}X_iX_j)\Big] =& \mathbb{E}\Big[\mathbb{E}\big[4\sum_{(i,j)\in I\times I^C} a_{ij}X_iX_j \big| \underset{\sim}{X}\big]\Big] && \text{(tower property)} \\ \le& \mathbb{E}\Big[F\big(\mathbb{E}\big[4\sum_{(i,j)\in I\times I^C} a_{ij}X_iX_j \big| \underset{\sim}{X}\big]\big)\Big] && \text{(By Jensen's)} \\ =& \mathbb{E}\Big[F\big(\underset{\sim}{X}^T A \underset{\sim}{X}\big)\Big] && (\star) \end{aligned}$$

Note: There exists a *deterministic* set $J \subseteq [n]$ such that $(\star)$ holds with $I = J$.

To complete the sum, write

$$\underset{\sim}{X}^T A \underset{\sim}{X} = \underbrace{\sum_{(i,j)\in J\times J^C} a_{ij} X_i X_j'}_{=y} + \underbrace{\sum_{(i,j)\in J\times J} a_{ij} X_i X_j'}_{=z_1} + \underbrace{\sum_{(i,j)\in J^C\times [n]} a_{ij} X_i X_j'}_{=z_2}$$

Let $g = \sigma\big((X_i)_{i\in J}, (X_i')_{i\in J^C}$. Then $y \in g$ (i.e. y is g-measurable) and

$$\mathbb{E}[z_1|g] = 0 = \mathbb{E}[z_2|g]$$

Applying the lemma from the previous lecture (see top of notes), we have

$$F(4y) \leq \mathbb{E}\big[F(4y + 4z_1 + 4z_2)|g\big]$$

Taking the expectation of both sides yields the result. □

2.2 Idea 2: Comparison

Note: We use this to bound the moment generating function.

Theorem 4 (Comparison (Thm 6.2.3 in Vershynin))**.** *Assuming the general setting described above with the addition of $K = \max_i ||X_i||_{\psi_2} < +\infty$. Let $g, g' \sim N(0, I_n)$ independent. Then*

$$\mathbb{E}\Bigg[\exp\big(\lambda \underset{\sim}{X}^T A \underset{\sim}{X}'\big)\Bigg] \leq \mathbb{E}\Bigg[\exp\big(c\lambda k^2 g^T A g'\big)\Bigg]$$

Proof. Write $Y = A\underset{\sim}{X}'$.

$$\begin{aligned}
\mathbb{E}\Bigg[\exp\big(\lambda \underset{\sim}{X}^T A \underset{\sim}{X}'\big)\big|\underset{\sim}{X}'\Bigg] &= \mathbb{E}\Bigg[\exp\big(\lambda \sum_{i=1}^n X_i Y_i\big)\big|\underset{\sim}{X}'\Bigg] \\
&= \prod_{i=1}^n \mathbb{E}\Bigg[\exp\big(\lambda X_i Y_i\big)\big|\underset{\sim}{X}'\Bigg] \\
&\leq \prod_{i=1}^n \exp\big(cK^2\lambda^2 Y_i^2\big) \\
&= \exp\big(cK^2\lambda^2 ||A\underset{\sim}{X}'||_2^2\big)
\end{aligned}$$

Now replacing X by g and using the moment generating function of normal random variables we get

$$\mathbb{E}\big[\exp\big(\mu g^T A X'\big)|X'\big] = \exp\big(\frac{1}{2}\mu^2 ||AX'||_2^2\big)$$

and selecting $\mu = \sqrt{2c}K\lambda$ we get the desired result

$$\mathbb{E}\big[\exp\big(\lambda X^T A X'\big)\big] \leq \mathbb{E}\big[\exp\big(\sqrt{2c}K\lambda g^T A X'\big)\big]$$

□

References

[1] R. Vershynin, *High dimensional probability. An introduction with applications in Data Science*, Cambridge University Press, 2019.