MATH888: High-dimensional probability and statistics Fall 2021

Lecture 11 — October 1, 2021

Sebastien Roch, UW-Madison *Scribe: Vasilis Kontonis*

1 Overview

In the last lecture we presented the *decoupling* and *comparison* theorems which will be usefull in the proof of the main result of this lecture: the Hanson-Wright inequality. We have the following decoupling inequality. In the lecture we assumed that the diagonal elements of A are zero but the same argument actually proves a stronger version without this assumption, see Remark 6.13 in [1].

Lemma 1 (Decoupling, Theorem 6.1.1 in [1])**.** *Let A be an $n \times n$ matrix. 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*

$$\mathbb{E}\Big[F\Big(\sum_{i \neq j} a_{ij} X_i X_j\Big)\Big] \leq \mathbb{E}\Big[F(4X^T A X')\Big],$$

where $X^{'}$ is an independent copy of X.

The following comparison lemma shows that we can essentially replace the sub-Gaussian random variables X, X' in the MGF of the quadratic form $X^T A X'$ by Gaussians.

Lemma 2 (Comparison, Lemma 6.2.3 in [1])**.** *Let A be an $n \times n$ matrix and $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ 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*

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

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

2 Review of Matrix Norms and Singular Value Decomposition

We reviewed basic facts about matrix norms and the singular value decomposition. For details see the slides on `https://people.math.wisc.edu/~roch/mmids/`.

3 The Proof of the Hanson-Wright Inequality

In this lecture, we will prove the Hanson-Wright Inequality. We first restate its statement and then proceed to its proof.

Theorem 3 (Hanson-Wright)**.** *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 $n \times n$ matrix. Then, for every $t \geq 0$, we* *have*

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

where $K = \max_i ||X_i||_{\psi_2}$ *and* $c > 0$ *is some universal constant.*

Proof. In what follows we have that $X = (X_1, \dots, X_n)$ is a random vector with independent mean-zero coordinates and X' be an independent copy of X. We will use the decoupling and comparison lemmas stated above. Moreover, we will use the following estimate on the MGF of quadratic forms of independent Gaussian random variables.

Lemma 4 (Gaussian Chaos MGF)**.** *Let* $g, g' \sim N(0, I_n)$ *be independent and* A *be an* $n \times n$ *matrix. Then for any* $\lambda \in \mathbb{R}$, $|\lambda| \leq c/\|A\|_2$, *we have*

$$\mathbb{E}[\exp\left(\lambda g^T A g'\right)] \leq \exp\left(C\lambda^2\|A\|_F^2\right).$$

We can decompose X^TAX into a term involving the diagonal elements of A and one with the off-diagonal:

$$X^TAX = \sum_{i,j} a_{ij}X_iX_j = \sum_i a_{ii}X_i^2 + \sum_{i\neq j} a_{ij}X_iX_j.$$

Since the X_i's are independent, mean-zero random variables it holds that $\mathbb{E}[X^TAX] = \sum_i a_{ii}\mathbb{E}X_i^2$. Using the fact that for any two random variables X, Z it holds that $\mathbb{P}[X + Z \geq t] \leq \mathbb{P}[X \geq t/2] + \mathbb{P}[Z \geq t/2]$ we obtain the following upper bound on the tail probability of the quadratic form

$$\mathbb{P}\left[\left|X^TAX - \mathbb{E}(X^TAX)\right| \geq t\right] \leq \mathbb{P}\left[\sum_{i=1}^n a_{ii}(X_i^2 - \mathbb{E}X_i^2) \geq t/2\right] + \mathbb{P}\left[\sum_{i\neq j} a_{ij}X_iX_j \geq t/2\right].$$

Step 1: diagonal sum. X_i's are independent, sub-gaussian random variables, and therefore, the random variables $X_i^2 - \mathbb{E}X_i^2$ are centered (mean-zero), independent, and sub-exponential. Using the fact that centering a random variable can only reduce its sub-exponential norm we obtain

$$\left\|a_{ii}(X_i^2 - \mathbb{E}X_i^2)\right\|_{\Psi_1} \lesssim |a_{ii}|\|X_i^2\|_{\Psi_1} \lesssim |a_{ii}|\|X_i\|_{\Psi_2}^2 \lesssim |a_{ii}|K^2.$$

Thus we can apply Bernstein's Inequality.

Lemma 5 (Bernstein's inequality (Theorem 2.8.1 in [1]))**.** *Let* $X_1, \dots, X_N$ *be independent, mean zero, sub-exponential random variables. Then, for every* $t \geq 0$, *we have*

$$\mathbb{P}\left[\left|\sum_{i=1}^N X_i\right| \geq t\right] \leq 2\exp\left(-c\min\left\{\frac{t^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \frac{t}{\max_i \|X_i\|_{\psi_1}}\right\}\right),$$

where $c > 0$ *is an absolute constant.*

We can now bound the tail probability of the diagonal terms

$$\begin{aligned}\mathbb{P}\Big[\sum_{i=1}^n a_{ii}(X_i^2 - \mathbb{E}X_i^2) \geq t/2\Big] &\leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\sum_i a_{ii}^2}, \frac{t}{K^2\max_i |a_{ii}|}\right)\right) \\ &\leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right),\end{aligned}$$

where we used the fact that $\max_i |a_{ii}| \leq \|A\|_2$ and that $\sum_i a_{ii}^2 \leq \|A\|_F^2$. We move on to the off-diagonal terms. Let $S = \sum_{i \neq j} a_{ij} X_i X_j$. By Markov's Inequality, we have that for any $\lambda > 0$ it holds that

$$P[S \geq t/2] = P\left[e^{\lambda S} \geq e^{\lambda t/2}\right] \leq \exp(-\lambda t/2)\mathbb{E}[\exp(\lambda S)].$$

Given that $|\lambda| \leq c/\|A\|$, we can use the Decoupling, the Comparison, and the MGF of Gaussian Chaos lemmas to obtain

$$\mathbb{E}[\exp(\lambda S)] \leq \mathbb{E}[\exp(4\lambda X^\top A X')] \leq \mathbb{E}[\exp(C_1 K^2 \lambda g^\top A g')] \leq \exp\left(C_2 K^4 \lambda^2 \|A\|_F^2\right),$$

where C_1, C_2 are universal constants. Therefore, we obtain that $P[S \geq t/2] \leq \exp\left(-\lambda t/2 + C_2 K^4 \lambda^2 \|A\|_F^2\right)$. We can now pick the value of $\lambda \in (0, c/\|A\|_2)$, that maximizes the expression $-\lambda t/2 + C_2 K^4 \lambda^2 \|A\|_F^2$ and obtain that

$$P[S \geq t/2] \leq 2\exp\left(-c_1 \min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right),$$

for some other universal constant c_1. To complete the proof we put together the above estimates for the diagonal and off-diagonal tail probabilities and replace the corresponding universal constants by another one. □

As a corollary of the Hanson-Wright inequality we give the following theorem.

Theorem 6. (Concentration for Random Vectors, Thm 6.3.2 in [1]) *Let* $X = (X_1, X_2, ...X_n) \in \mathbb{R}^n$ *be a random vector with independent, mean-zero, sub-gaussian coordinates with* $K = \max \|X_i\|_{\Psi_2} < \infty$. *Let* B *be an* $n \times m$ *matrix then it holds* $\| \|BX\|_2 - \|B\|_F \|_{\Psi_2} \leq CK^2\|B\|$.

References

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