MATH888: High-dimensional probability and statistics Fall 2021

Lecture 11 — October 1, 2021

Sebastien Roch, UW-Madison *Scribe: Changyu Gao*

1 Overview

In the last lecture we presented the *decoupling* and *comparison* theorems which prepare us for the proof of the Hanson-Wright Inequality.

In this lecture, we will prove the Hanson-Wright Inequality.

2 Hanson-Wright Proof and Applications

2.0 Review of Matrix Norms and Singular Value Decomposition

See *Topic 2* on `https://people.math.wisc.edu/~roch/mmids/` for details.

Recall that the Frobenius norm of an $n \times m$ matrix $A \in \mathbb{R}^{n\times m}$ is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2}.$$

Define $\mathbb{S}^{m-1} = \{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}\| = 1\}$. The induced 2-norm of a matrix $A \in \mathbb{R}^{n\times m}$ is

$$\|A\|_2 = \max_{\mathbf{0}\neq\mathbf{x}\in\mathbb{R}^m} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{x}\in\mathbb{S}^{m-1}} \|A\mathbf{x}\|.$$

The singular value decomposition (SVD) of a matrix $A \in \mathbb{R}^{n\times m}$ is a matrix factorization

$$A = U\Sigma V^T = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^T,$$

where the columns of $U \in \mathbb{R}^{n\times r}$ and those of $V \in \mathbb{R}^{m\times r}$ are orthonormal, and $\Sigma \in \mathbb{R}^{r\times r}$ is a diagonal matrix. Here the $\mathbf{u}_j$'s are the columns of U and are referred to as left singular vectors. Similarly the $\mathbf{v}_j$'s are the columns of V and are referred to as right singular vectors. The σ_j 's, which are non-negative and in decreasing order

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

are the diagonal elements of Σ and are referred to as singular values.

2.1 Hanson-Wright Proof and Applications

Setting (*): let $X = (X_1, \dots, X_n)$ be a random vector with independent mean-zero coordinates and X' be an independent copy of X. Let A be an $n \times n$ matrix.

Theorem 1. (Hanson-Wright Inequality) *Suppose* (*) *holds and further* $K = \max \|X_i\|_{\Psi_2} < \infty$, $\forall i$. *Then for every* $t \geq 0$*, we have*

$$\mathbb{P}(\left|X^T AX - \mathbb{E}(X^T AX)\right| \geq t) \leq 2 \exp\left(-c \min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right).$$

Proof. We will use three lemmas:

L1 (Decoupling) Suppose (*) holds and further the diagonal entries of A are all zero. For any convex function F, we have

$$\mathbb{E}(F(X^T AX)) \leq \mathbb{E}(F(4X^T AX')).$$

Remark of L1 (Remark 6.1.3. in [V]) A slightly stronger version of decoupling inequality holds, in which A needs not be diagonal-free. For any square matrix $A = (a_{ij})$, we have

$$\mathbb{E}F\left(\sum_{i,j:i\neq j} a_{ij}X_iX_j\right) \leq \mathbb{E}F\left(4\sum_{i,j} a_{ij}X_iX_j'\right).$$

L2 (Comparison, Lemma 6.2.3 in [V]) Suppose (*) holds with $\|X\|_{\psi_2} \leq K$ and $\|X'\|_{\psi_2} \leq K$. Consider also independent random vectors $g, g' \sim N(0, I_n)$. Then for any $\lambda \in \mathbb{R}$,

$$\mathbb{E}(F(4X^T AX')) \leq \mathbb{E}(\exp\left(cK^2 g^T Ag'\right)).$$

L3 (Gaussian chaos, Lemma 6.2.2 in [V]) 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

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

We write

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

It follows from independence and the mean-zero assumption that

$$\mathbb{E}(X^T AX) = \sum_i a_{ii}\mathbb{E}X_i^2.$$

The problem reduces to estimating the diagonal and off-diagonal sums:

$$\mathbb{P}(\left|X^T AX - \mathbb{E}(X^T AX)\right| \geq t) \leq \underbrace{\mathbb{P}\left(\sum_{i=1}^n a_{ii}(X_i^2 - \mathbb{E}X_i^2) \geq t/2\right)}_{p_1} + \underbrace{\mathbb{P}\left(\sum_{i\neq j} a_{ij}X_iX_j \geq t/2\right)}_{p_2}.$$

(Diagonal sum) Note that by centering

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

Applying Bernstein's Inequality,

$$p_1 \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\sum_i a_{ii}^2}, \frac{t}{K^2\max_i|a_{ii}|}\right)\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right).$$

(Off diagonal sum) Let $S = \sum_{i\neq j} a_{ij}X_iX_j$. By Chebyshev's Inequality, for any $\lambda > 0$

$$p_2 = P(S \geq t/2) = P\left(e^{\lambda S} \geq e^{\lambda t/2}\right) \leq \exp(-\lambda t/2)\mathbb{E}(\exp(\lambda S)).$$

Now,

$$\begin{aligned}\mathbb{E}\exp(\lambda S) &\leq \mathbb{E}\exp\left(4\lambda X^\top AX'\right) && \text{(by L1)}\\ &\leq \mathbb{E}\exp\left(C_1\lambda g^\top Ag'\right) && \text{(by L2)}\\ &\leq \exp\left(C\lambda^2\|A\|_F^2\right) && \text{(by L3)},\end{aligned}$$

provided that $|\lambda| \leq c/\|A\|$.

Hence,

$$p_2 \leq \exp\left(-\lambda t/2 + C\lambda^2\|A\|_F^2\right).$$

Optimize over $0 < \lambda \leq c/\|A\|$, we conclude that

$$p_2 \leq 2\exp\left(-c'\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right).$$

We complete the proof by putting together the diagonal part p_1 and the off-diagonal part p_2.

□

We have the following theorem as a consequence of the Hanson-Wright Inequality:

Theorem 2. (Concentration for random vectors, Thm 6.3.2 in [V]) *Suppose* (*) *holds and further* $K = \max\|X_i\|_{\Psi_2} < \infty$, $\forall i$. *Let* B *be an* $n \times m$ *matrix. Then*

$$\|\|BX\|_2 - \|B\|_F\|_{\Psi_2} \leq CK^2\|B\|.$$

Proof. See Vershynin's textbook. □

References

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