Homework 11
HDP KNU+ Fall 2022

---

Hints are in the back of this homework set.

---

We developed a general method that allowed us to upgrade concentration inequalities to *random matrices*. Using this method, we proved matrix Hoeffding inequality (Lecture 29, November 9) and stated matrix Bernstein inequality (Lecture 30, November 11). In this problem, we use this method to establish matrix Chernoff inequality, whose scalar version was proved in Lecture 7 (September 16).

## Problem 1 (Matrix Chernoff inequality)

(a) Let $\xi$ be a Bernoulli random variable with parameter $p$, and $A$ be a (fixed) $d \times d$ symmetric matrix whose largest eigenvalue satisfies $\lambda_1(A) \le 1$. Prove the following bound on (matrix) moment generating function of $\xi A$:

$$\mathbb{E} \exp(\lambda \xi A) \preceq \exp\left[(e^{\lambda} - 1)p\right] I_d \quad \text{for } \lambda > 0,$$

Here $I_d$ denotes the identity matrix, $\preceq$ refers to the Löwner order, and $\lambda_1(\cdot)$ denotes the largest eigenvalue of a matrix.

(b) Prove the following matrix Chernoff inequality. Let $\xi_i$ be independent Bernoulli random variables with parameters $p_i$. Let $A_i$ be (fixed) $d \times d$ symmetric matrices that satisfy $\lambda_1(A_i) \le 1$ for all $i$. Then, for any $t > \mu$, we have

$$\mathbb{P}\left\{ \lambda_1\left( \sum_{i=1}^n \xi_i A_i \right) \ge t \right\} \le d \cdot e^{-\mu} \left( \frac{e\mu}{t} \right)^t, \quad \text{where } \mu = \sum_{i=1}^n p_i.$$

---

Next, we will prove a few basic facts about the operator norm. They will be helpful in the further problems.

## Problem 2 (Computing the operator norm)

(a) Let $A$ be any matrix. Explain why the following identities hold:

$$\|A\|^2 = \lambda_1\left(AA^{\mathsf{T}}\right) = \lambda_1\left(A^{\mathsf{T}}A\right) = \left\|AA^{\mathsf{T}}\right\| = \left\|A^{\mathsf{T}}A\right\|.$$

(b) Deduce from (a) that the operator norm is invariant under transposition, i.e.

$$\|A^{\mathsf{T}}\| = \|A\|.$$

(c) Let $S$ be a $d_1 \times d_1$ symmetric matrix, and $T$ be a $d_2 \times d_2$ symmetric matrix. Prove that the $(d_1 \times d_1) \times (d_1 \times d_1)$ block diagonal matrix

$$Y = \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix} \quad \text{satisfies} \quad \|Y\| = \|S\| \vee \|T\|.$$

Here $\vee$ is a shorthand for maximum, i.e. $a \vee b = \max(a, b)$.

(d) Let $A$ be a $d_1 \times d_2$ matrix, and $B$ be a $d_2 \times d_1$ matrix. Prove that the $(d_1 + d_2) \times (d_1 + d_2)$ block matrix

$$Z = \begin{bmatrix} 0 & A \\ B & 0 \end{bmatrix} \quad \text{satisfies} \quad \|Z\| = \|A\| \vee \|B\| .$$

---

So far, we proved matrix concentration inequalities—Hoeffding, Bernstein, Chernoff–for *symmetric* matrices only. The symmetry was crucial in the proof: it is required Lieb's trace inequality. But a cute step, called *Girko's hermitization trick*, allows us to automatically extend all those results to nonsymmetric matrices. In fact, arbitrary rectangular $d_1 \times d_2$ matrices $B$ will be allowed.

Girko's hermitization consists of replacing $B$ by the $(d_1 + d_2) \times (d_1 + d_2)$ block matrix

$$H(B) := \begin{bmatrix} 0 & B \\ B^{\mathsf{T}} & 0 \end{bmatrix}.$$

Then $H(B)$ is obviously symmetric. Moreover, by Problem 2 parts (d) and (a), hermitization preserves the operator norm: $\|H(B)\| = \|B\|$.

Using hermetization trick, one can extend many results from symmetric to general matrices. Let's do this for the expected form of matrix Hoeffding inequality (Lecture 30, November 11, p.2).

Problem 3 (Matrix Hoeffding inequality for rectangular matrices)

Let $\varepsilon_i$ be independent symmetric Bernoulli random variables, i.e. each $\varepsilon_i$ takes values $\pm 1$ independently with probability $1/2$. Let $B_i$ be (fixed) $d_1 \times d_2$ matrices. Prove that

$$\mathbb{E}\Big\|\sum_{i=1}^n \varepsilon_i B_i\Big\| \le 3\sigma\sqrt{\log(d_1 + d_2)} \quad \text{where } \sigma^2 = \Big\|\sum_{i=1}^n B_i B_i^{\mathsf{T}}\Big\| \vee \Big\|\sum_{i=1}^n B_i^{\mathsf{T}} B_i\Big\| .$$

---

Unlike Frobenius norm, the operator norm of a matrix $A$ is very hard to express in terms of the entries $A_{ij}$. However, if we *randomize* the matrix $A$ by flipping the sign of every entry $A_{ij}$ independently, the operator norm becomes simpler! In this problem, we will show that it is equivalent (up to a logarithmic factor) to the maximal Euclidean norm of all rows $A_{i*} = (A_{i1}, \ldots, A_{id_2})$ and columns $A_{*j} = (A_{1j}, \ldots, A_{d_1 j})^{\mathsf{T}}$ of $A$. Formally, let us define the *"row-column norm"* norm of $A$ by

$$\|A\|_{\mathrm{rc}} := \max_{i=1,\ldots,d_1} \|A_{i*}\|_2 \vee \max_{i=1,\ldots,d_2} \|A_{*j}\|_2 .$$

Problem 4 (Matrix Hoeffding inequality for rectangular matrices)

(a) Prove that the operator norm bounds the row-column norm, i.e.

$$\|B\|_{\mathrm{rc}} \le \|B\|$$

for any matrix $B$.

(b) Let $A$ be a $d \times d$ matrix. Let $A^{\pm}$ denote the $d \times d$ random matrix with entries $\varepsilon_{ij} A_{ij}$, where $\varepsilon_{ij}$ are independent symmetric Bernoulli random variables. Show that

$$\|A\|_{\mathrm{rc}} \leq \mathbb{E}\left\|A^{\pm}\right\| \leq 3\sqrt{\log(2d)}\,\|A\|_{\mathrm{rc}}\,.$$

TURN OVER FOR HINTS

# Hints

## Hints for Problem 1.

(a) A scalar version of this bound was established in our proof of Chernoff inequality (Lecture 7, September 16). It should not be hard to adapt this argument to the matrix setting, if you note that the assumption on $A$ gives $\exp(\lambda A) \preceq \exp(\lambda) I_d$.

(b) Adapt our proof of scalar Chernoff inequality (Lecture 7, September 16) to the matrix-valued setting in the same way as we adapted Hoeffding inequality in Lecture 29, November 9. The only real difference is in the bound on the moment generating function $\mathbb{E}\exp(\lambda \xi_i A_i)$; use the bound from part (a).

## Hints for Problem 2.

(a) For the first two identities, use the definition of the operator norm (Lecture 22, October 21). For the next two identities, use the result of Homework 8, Problem 2(a).

(c) Note that the spectrum of $Y$ is the union of the spectra of $S$ and $T$. Then recall the result of Homework 8, Problem 2(a).

(d) Compute $ZZ^\mathsf{T}$; it should be a block-diagonal matrix. Use parts (c) and (a) to compute its norm and relate it to the norms of $A$ and $B$.

## Hints for Problem 3.

Apply the expected form of matrix Hoeffding inequality, i.e. $\mathbb{E}\|\sum_{i=1}^n \varepsilon_i A_i\| \le 3\sigma\sqrt{\log d}$, for the Hermiticized matrices, i.e. for $A_i := H(B_i)$. As we noted, hermitization does not change the norm, so $\|\sum_{i=1}^n \varepsilon_i A_i\| = \|\sum_{i=1}^n \varepsilon_i B_i\|$. To compute $\sigma^2 = \left\|\sum_{i=1}^n \varepsilon_i A_i^2\right\|$, compute $A_i^2$, sum up, and use Problem 2(c).

## Hints for Problem 4.

(a) By definition of the operator norm, $\|A\| \ge \|Ae_i\|_2$ for all $i$, where $e_1, \dots, e_d$ is the canonical basis of $\mathbb{R}^d$. Note that $Ae_i$ is the $i$-th column of $A$. Next, use the same argument for the transposed matrix $A^\mathsf{T}$, whose norm is the same as that of $A$ by Problem 2(b).

(b) The lower bound follows from part (a). For the upper bound, decompose the matrix $A^\pm$ into a sum of independent random matrices, each of which contains a single entry of $A^\pm$ (the rest being zero); then apply the result of Problem 3. A similar computation was made in Lecture 31, November 14.

If this hint is still sounds too vague, here are more details. The matrix $e_i e_j^\mathsf{T}$ has all zero entries except the $(i,j)$-th entry that equals 1. Thus, we can decompose any matrix as $X = \sum_{ij} X_{ij} e_i e_j^\mathsf{T}$. In particular, $A^\pm = \sum_{ij} \varepsilon_{ij} A_{ij} e_i e_j^\mathsf{T}$. This allows us to apply the result of Problem 3 with matrices $B_{ij} := A_{ij} e_i e_j^\mathsf{T}$ instead of $B_i$. To compute $\sigma^2$, we need to compute both $\sum_{ij} B_{ij} B_{ij}^\mathsf{T}$ and $\sum_{ij} B_{ij}^\mathsf{T} B_{ij}$. Do this; each of these two matrices will be diagonal. The eigenvalues $\lambda_i$ of diagonal matrices are the diagonal entries; then use Homework 8, Problem 2(a).