HOMEWORK 12
HIGH-DIMENSIONAL PROBABILITY FOR DATA SCIENCE, FALL 2023

Hints are in the back of this homework set.

As in the previous homework sets, $C, C_1, C_2, \dots$ and $c, c_1, c_2, \dots$ denote positive absolute constants of your choice.

In the following problem we show that whenever two unit vectors $u, v \in \mathbb{R}^d$ are close to each other, then so are $P_u = uu^\mathsf{T}$ and $P_v = vv^\mathsf{T}$, which are the orthogonal projections onto the lines spanned by these two vectors. We will also prove the converse: if the projections P_u and P_v are close, then the vectors themselves must be close up to a sign, i.e. $u \approx \pm v$. From this we derive a useful version of Davis-Kahan perturbation theorem for the top eigenvectors.

1. Perturbation of eigenvectors

(a) Let x and y be arbitrary vectors in $\mathbb{R}^d$. Compute the operator and Frobenius norms of the matrix xy^T. (Show your work!)

(b) Let u and v be unit vectors in $\mathbb{R}^d$. Prove that there exists a sign $s \in \{-1, 1\}$ such that

$$\frac{1}{2}\|u - sv\|_2 \le \|uu^\mathsf{T} - vv^\mathsf{T}\| \le 2\|u - v\|_2 \,,$$

where the norm in the middle is the operator norm.

(c) Deduce the following version of a Davis-Kahan theorem (see Lecture 20 for a general statement). Let A and B be $d \times d$ symmetric matrices. Then there exist a sign $s \in \{-1, 1\}$ such that

$$\|v_1(A) - sv_1(B)\|_2 \le \frac{2\|A - B\|}{\lambda_1(A) - \lambda_2(A)}$$

where $\lambda_k(A)$ denote the eigenvalues of A in the non-increasing order, and $v_k(A)$ denote the corresponding unit eigenvectors, and similarly for B.

Consider a probability distribution in $\mathbb{R}^d$ with mean zero and covariance matrix

$$\Sigma = I + \beta uu^\mathsf{T} \tag{1}$$

where $\beta \in (0, 1)$ is a fixed number and $u \in \mathbb{R}^d$ is a fixed unit vector. This is a *spike model*, a basic mathematical model of data with structure. The data sampled from this distribution looks have variance 1 in all directions except one: in the direction of u the variance is larger, making a "spike". The direction of u is interpreted as a "signal" which may contain some information about the data, while the other directions are noise. The magnitude of β measures the strength of the spike, known as the signal-to-noise ratio (SNR).

In this problem, we will accurately estimate the “signal” u from a sample of $n = O(d)$ points $X_1, \dots, X_n$. All we have to do is compute $v_1(\Sigma_n)$, the eigenvector corresponding to the largest eigenvalue of the sample covariance matrix $\Sigma_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\mathsf{T}$. Obviously, u can only be recovered up to a sign, and we will show that $u \approx \pm v_1(\Sigma_n)$ for some choice of the sign.

2. Learning a spike model

(a) Check that the two largest eigenvalues of the spike covariance matrix Σ in (1) are $\lambda_1(\Sigma) = 1 + \beta$ and $\lambda_2(\Sigma) = 1$. Check that the top eigenvector is $v_1(\Sigma) = u$.

(b) Assume that $X_1, \dots, X_n$ are i.i.d. subgaussian random vectors in $\mathbb{R}^d$ with mean zero, covariance matrix Σ as in (1), and with with $\|X_i\|_{\psi_2} \le 10$. Let $v = v_1(\Sigma_n)$ be the top eigenvector of the sample covariance matrix. Show that if $n \ge C_2 d/\beta^2$, then

$$\min_{s \in \{-1,1\}} \|u - sv\|_2 \le 0.01$$

with probability at least $1 - 2e^{-d}$.

In Lecture 21, we developed *matrix calculus* that allowed us to work with $d \times d$ symmetric matrices as if they were numbers. In particular, we introduced the (partial) semidefinite order[1] and defined functions of matrices.[2]

Most facts about numbers transfer to matrices, but some do not. Part (a) of this problem gives a matrix version of the following fact about scalars:

$$|x| \le t \iff -t \le x \le t \quad \text{for } x \in \mathbb{R}.$$

Part (d) unexpectedly shows that the scalar fact

$$f : \mathbb{R} \to \mathbb{R} \text{ increasing function}, \;\; x \le y \Longrightarrow f(x) \le f(y)$$

does not generalize to matrices.

3. Matrix calculus

In this problem, X and Y denote $d \times d$ symmetric real matrices.

(a) Prove that $\|X\| \le t$ if and only if $-tI \preceq X \preceq tI$.

(b) Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions. Prove that $f(x) \le g(x)$ for all $x \in \mathbb{R}$ satisfying $|x| \le K$, then $f(X) \preceq g(X)$ for all X satisfying $\|X\| \le K$.

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing function and X, Y be commuting matrices. Prove that $X \preceq Y$ implies $f(X) \preceq f(Y)$.

(d) Show by example that property (d) may fail for non-commuting matrices.

---

[1] To recall this definition, we write $X \succeq 0$ if X is a symmetric positive semidefinite matrix. We say that $X \succeq Y$, or equivalently, $Y \preceq X$, if $X - Y \succeq 0$.

[2] To recall this definition, let $f : \mathbb{R} \to \mathbb{R}$ and let $A = \sum_{i=1}^d \lambda_i u_i u_i^\mathsf{T}$ be a spectral decomposition of A. We define $f(A) = \sum_{i=1}^d f(\lambda_i) u_i u_i^\mathsf{T}$.

Hints

Hint for Problem 1.

(a) Both norms equal $\|x\|_2 \, \|y\|_2$.

(b) The second inequality is simpler. To prove it, add and subtract the cross term uv^T and use triangle inequality; this reduces the problem to computing the norms of $u(u-v)^\mathsf{T}$ and $(u-v)v^\mathsf{T}$; use the result of part (a) and the assumption that $\|u\|_2 = \|v\|_2 = 1$.

To prove the first inequality in (b), assume without loss of generality that $\langle u, v\rangle \geq 0$ (otherwise replace u by $-u$), let $R = uu^\mathsf{T} - vv^\mathsf{T}$ and $\varepsilon = \|R\|$. Then $\left|u^\mathsf{T} R u\right| \leq \varepsilon$ (why?); deduce from this that $|1 - \langle u, v\rangle| \leq \varepsilon$ and thus $\|u - u\langle u, v\rangle\|_2 \leq \varepsilon$. Moreover, $\|Rv\|_2 \leq \varepsilon$ (why?); this means that $\|u\langle u, v\rangle - v\|_2 \leq \varepsilon$. Add the two bounds using triangle inequality to conclude that $\|u - v\|_2 \leq \varepsilon$.

(c) Apply the general Davis-Kahan theorem for $k = 1$; then use the first inequality in part (b) to pass from the projections to the eigenvectors.

Hint for Problem 2.

(b) Use the covariance estimation result from HW 11, Problem 3(a). Combine it with the results of Problem 1(c) and 2(a) of the current homework.

Hint for Problem 3.

(a) Recall that the operator norm of X can be computed as the maximum of the quadratic form $u^\mathsf{T} X u$ over all unit vectors $u \in \mathbb{R}^d$.

(c) Commuting symmetric matrices are simultaneously diagonalizable by an orthogonal matrix, so we can write their spectral representations with the same eigenvectors.

(d) Find 2×2 matrices such that $0 \preceq X \preceq Y$ but $X^2 \not\preceq Y^2$.