# Homework 5

## This homework is due at 11 PM on February 16, 2024.

**Submission Format:** Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

1. **Matrix Norm Calculations**

   Let $A$ have SVD equal to

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{1}$$

   (a) Compute $\|A\|_F$, the Frobenius norm of $A$.

   (b) Compute $\|A\|_2$, the spectral norm of $A$.

**2. PCA and low-rank compression**

We have a data matrix $X = \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_n^\top \end{bmatrix}$ of size $n \times d$ containing $n$ data points[1], $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$, with $\vec{x}_i \in \mathbb{R}^d$. Note that $\vec{x}_i^\top$ is the $i$th row of $X$. Assume that the data matrix is centered, i.e. each column of $X$ is zero mean. In this problem, we will show equivalence between the following three problems:

($P_1$) Finding a line going through the origin that maximizes the variance of the scalar projections of the points on the line. Formally $P_1$ solves the problem:

$$\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmax}} \quad \vec{u}^\top C \vec{u} \tag{2}$$

with $C = \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^\top$ denoting the covariance matrix associated with the centered data.

($P_2$) Finding a line going through the origin that minimizes the sum of squares of the $\ell^2$ distances from the points to their vector projections. Formally $P_2$ solves the minimization problem:

$$\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{v_i \in \mathbb{R}} \|\vec{x}_i - v_i \vec{u}\|_2^2 . \tag{3}$$

Note that the vector projection of $\vec{x}$ on $\vec{u}$ is given by $v^\star \vec{u}$, where

$$v^\star = \underset{v \in \mathbb{R}}{\operatorname{argmin}} \|\vec{x} - v\vec{u}\|_2^2 , \tag{4}$$

and we will show that $v^\star = \langle \vec{x},\ \vec{u} \rangle$ in part (a).

($P_3$) Finding a rank-one approximation to the data matrix. Formally $P_3$ solves the minimization problem:

$$\underset{Y : \operatorname{rank}(Y) \le 1}{\operatorname{argmin}} \quad \|X - Y\|_F . \tag{5}$$

Note that loosely speaking, two problems are said to be "equivalent" if the solution of one can be "easily" translated to the solution of the other. Some form of "easy" translations include adding/subtracting a constant or some quantity depending on the data points.

Note the significance of these results. $P_1$ is finding the first principal component of $X$, the direction that maximizes variance of scalar projections. $P_2$ says that this direction also minimizes the distances between the points to their vector projections along this direction. If we view the distances as errors in approximating the points by their projections along a line, then the error is minimized by choosing the line in the same direction as the first principal component. Finally $P_3$ tells us that finding a rank one matrix to best approximate the data matrix (in terms of error computed using Frobenius norm) is equivalent to finding the first principal component as well!

(a) Consider the line $\mathcal{L} = \{\vec{x}_0 + a\vec{u} : a \in \mathbb{R}\}$, with $\vec{x}_0 \in \mathbb{R}^d, \vec{u}^\top \vec{u} = 1$. Recall that the vector projection of a point $\vec{x} \in \mathbb{R}^d$ on to the line $\mathcal{L}$ is given by $\vec{z} = \vec{x}_0 + a^\star \vec{u}$, where $a^\star$ is given by:

$$a^\star = \underset{a}{\operatorname{argmin}} \|\vec{x}_0 + a\vec{u} - \vec{x}\|_2 . \tag{6}$$

[1] Data matrices are sometimes represented as above, and sometimes as the transpose of the matrix here. Make sure you always check this, and recall that based on the definition of the data matrix, the definition of the covariance matrix also changes.

 

Show that $a^{\star} = (\vec{x} - \vec{x}_0)^{\top}\vec{u}$. Use this to show that the square of the distance between $x$ and its vector projection on $\mathcal{L}$ is given by:

$$\|\vec{x} - \vec{z}\|_2^2 = \|\vec{x} - \vec{x}_0\|_2^2 - ((\vec{x} - \vec{x}_0)^{\top}\vec{u})^2. \tag{7}$$

(b) Show that $P_2$ is equivalent to $P_1$.

*HINT: Start with $P_2$ and using the result from part (a) show that it is equivalent to $P_1$.*

(c) Show that every matrix $Y \in \mathbb{R}^{n \times d}$ with rank at most 1, can be expressed as $Y = \vec{v}\vec{u}^{\top}$ for some $\vec{v} \in \mathbb{R}^n$, $\vec{u} \in \mathbb{R}^d$ and $\|\vec{u}\|_2 = 1$.

(d) Show that $P_3$ is equivalent to $P_2$.

*HINT: Use the result from part (c) to show that $P_3$ is equivalent to:*

$$\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^{\top}\vec{u} = 1, \vec{v} \in \mathbb{R}^n}{\operatorname{argmin}} \left\|X - \vec{v}\vec{u}^{\top}\right\|_F^2 \tag{8}$$

*Prove that this is equivalent to $P_2$.*

 

**3. Operator Norms**

For a matrix $A \in \mathbb{R}^{m \times n}$, the *induced norm* or *operator norm* $\|A\|_p$ is defined as

$$\|A\|_p \doteq \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p}. \tag{9}$$

In this problem, we provide a characterization of the induced norm for certain values of $p$. Let $a_{ij}$ denote the $(i, j)$-th entry of $A$. Prove the following:

(a) $\|A\|_2 = \sigma_{\max}\{A\}$, the maximum singular value of $A$. *HINT: Consider connecting $\|A\|_2^2$ to a particular Rayleigh coefficient.*

(b) $\|A\|_1$ is the maximum absolute column sum of $A$,

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|. \tag{10}$$

*HINT: Write $A\vec{x}$ as a linear combination of the columns of $A$ to obtain $\|A\vec{x}\|_1 = \|\sum_{i=1}^{n} x_i \cdot \vec{a}_i\|_1$, where $\vec{a}_i$ denotes the $i$-th column of $A$. Then apply triangle inequality to terms within the sum.*

(c) **(OPTIONAL)** $\|A\|_\infty$ is the maximum absolute row sum of $A$,

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|. \tag{11}$$

*HINT: First write $\|A\vec{x}\|_\infty = \max_i \left|\sum_{j=1}^{n} a_{ij} x_j\right|$. Then apply triangle inequality and use the fact that $|x_j| \leq \max_i |x_i|, \ \forall j$.*

**4. Gradients, Jacobians, and Hessians**

The *gradient* of a scalar-valued function $g\colon \mathbb{R}^n \to \mathbb{R}$ is the column vector of length $n$, denoted as $\nabla g$, containing the derivatives of components of $g$ with respect to the input variables:

$$(\nabla g(\vec{x}))_i = \frac{\partial g}{\partial x_i}(\vec{x}),\ i = 1, \dots n. \tag{12}$$

The *Hessian* of a scalar-valued function $g\colon \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ matrix, denoted as $\nabla^2 g$, containing the second derivatives of components of $g$ with respect to the input variables:

$$(\nabla^2 g(\vec{x}))_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}(\vec{x}),\ \ i = 1, \dots, n,\ \ j = 1, \dots, n. \tag{13}$$

The *Jacobian* of a vector-valued function $\vec{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix, denoted as $D\vec{g}$, containing the derivatives of components of $\vec{g}$ with respect to the input variables:

$$(D\vec{g}(\vec{x}))_{ij} = \frac{\partial g_i}{\partial x_j}(\vec{x}),\ \ i = 1, \dots, m,\ \ j = 1, \dots, n. \tag{14}$$

For the remainder of the class, we will repeatedly have to take gradients, Hessians and Jacobians of functions we are trying to optimize. This exercise serves as a warm up for future problems.

For the first two parts, suppose $A \in \mathbb{R}^{n\times n}$ is a square matrix whose entries are denoted $a_{ij}$ and whose rows are denoted $\vec{a}_1^\top, \dots, \vec{a}_n^\top$, and $\vec{b} \in \mathbb{R}^n$ is a vector whose entries are denoted $b_i$.

(a) Compute the Jacobians for the following functions.

i. $\vec{g}(\vec{x}) = A\vec{x}$.

ii. $\vec{g}(\vec{x}) = f(\vec{x})\vec{x}$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable.

iii. $\vec{g}(\vec{x}) = f(A\vec{x} + \vec{b})\vec{x}$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable.

(b) Compute the gradients and Hessians for the following functions.

i. $g_1(\vec{x}) = \vec{x}^\top A\vec{x}$.

ii. $g_2(\vec{x}) = \|\vec{x}\|_2^2$.

iii. $g_3(\vec{x}) = g_2(A\vec{x} - \vec{b}) = \|A\vec{x} - \vec{b}\|_2^2$. (Use the chain rule and the Jacobians computed in part (a).)

iv. $g_4(\vec{x}) = \log(\sum_{i=1}^n e^{x_i})$.

v. $g_5(\vec{x}) = g_4(A\vec{x} - \vec{b}) = \log(\sum_{i=1}^n e^{\vec{a}_i^\top \vec{x} - b_i})$. (Use the chain rule and the Jacobians computed in part (a); you can use the gradient $\nabla g_4$ and Hessian $\nabla^2 g_4$ in your answer without having to rewrite it.)

vi. $g_6(\vec{x}) = e^{\|\vec{x}\|_2^2} = e^{g_2(\vec{x})}$. (Use the chain rule and the Jacobians computed in part (a).)

vii. $g_7(\vec{x}) = e^{\|A\vec{x} - \vec{b}\|_2^2} = g_6(A\vec{x} - \vec{b})$. (Use the chain rule and the Jacobians computed in part (a); you can use the gradient $\nabla g_6$ and Hessian $\nabla^2 g_6$ in your answer without having to rewrite it.)

Consider the case now where all vectors and matrices above are scalar; do your answers above make sense? (No need to answer this in your submission.)

(c) Plot/hand-draw the level sets of the following functions:

i. $g(x_1, x_2) = \dfrac{x_1^2}{4} + \dfrac{x_2^2}{9}$

ii. $g(x_1, x_2) = x_1 x_2$

Also point out the gradient directions in the level-set diagram. Additionally, compute the first and second order Taylor series approximation around the point $(1, 1)$ for each function and comment on how accurately they approximate the true function.

 

**5. Homework Process**

With whom did you work on this homework? List the names and SIDs of your group members.

*NOTE*: If you didn't work with anyone, you can put "none" as your answer.