EECS 127/227AT Optimization Models in Engineering UC Berkeley Fall 2023

Homework 1

This homework is due at 11 PM on September 6, 2023.

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. **Diagonalization and Singular Value Decomposition**

   Let matrix $A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

   (a) Compute the eigenvalues and associated eigenvectors of A.

   (b) Express A as $P\Lambda P^{-1}$, where Λ is a diagonal matrix and $PP^{-1} = I$. State P, Λ, and P^{-1} explicitly.

   (c) Compute $\lim_{k\to\infty} A^k$.

   (d) Give the singular values σ_1 and σ_2 of A.

2. Least Squares

(a) Let $A \in \mathbb{R}^{m\times n}$ and $\vec{y} \in \mathbb{R}^m$ be given. Suppose A has full column rank. Show that the minimizer $\vec{x}^\star$ of the least-squares problem:

$$\min_{\vec{x}\in\mathbb{R}^n} \|A\vec{x} - \vec{y}\|_2^2$$

is given by $\vec{x}^\star = (A^\top A)^{-1}A^\top\vec{y}$.

(b) The Michaelis-Menten model for enzyme kinetics relates the rate y of an enzymatic reaction to the concentration x of a substrate, as follows:

$$y = \frac{\beta_1 x}{\beta_2 + x}, \tag{1}$$

for constants $\beta_1, \beta_2 > 0$. This model will be used throughout the remaining sub-parts of this problem.

Show that the model can be expressed as a linear relation between the values $1/y = y^{-1}$ and $1/x = x^{-1}$. Specifically, give an equation of the form $y^{-1} = w_1 + w_2x^{-1}$, specifying the values of w_1 and w_2 in terms of β_1 and β_2.

(c) In general, reaction parameters β_1 and β_2 (and, thus, w_1 and w_2) are not known a priori and must be fitted from data — for example, using least squares. Suppose you collect m measurements (x_i, y_i), $i = 1, \ldots, m$ over the course of a reaction. Formulate the least squares problem

$$\vec{w}^\star = \operatorname*{argmin}_{\vec{w}} \|X\vec{w} - \vec{y}\|_2^2\,, \tag{2}$$

where $\vec{w}^\star = \begin{bmatrix} w_1^\star & w_2^\star \end{bmatrix}^\top$, and you must specify $X \in \mathbb{R}^{m\times 2}$ and $\vec{y} \in \mathbb{R}^m$. Specifically, your solution should include explicit expressions for X and $\vec{y}$ as a function of (x_i, y_i) values and a final expression for $\vec{w}^\star$ in terms of X and $\vec{y}$, which should contain only matrix multiplications, transposes, and inverses.

Assume without loss of generality that $x_1 \neq x_2$.

(d) Assume that we have used the above procedure to calculate values for $w_1^\star$ and $w_2^\star$, and we now want to estimate $\widehat{\vec{\beta}} = \begin{bmatrix} \widehat{\beta}_1 & \widehat{\beta}_2 \end{bmatrix}^\top$. Write an expression for $\widehat{\vec{\beta}}$ in terms of $w_1^\star$ and $w_2^\star$.

NOTE: This problem was taken (with some edits) from the textbook *Optimization Models* by Calafiore and El Ghaoui.

3. Vector Spaces and Rank

The *rank* of a $m \times n$ matrix A, $\text{rank}(A)$, is the dimension of its *range*, also called span, and denoted $\mathcal{R}(A) := \{A\vec{x} \, : \, \vec{x} \in \mathbb{R}^n\}$.

(a) Assume that $A \in \mathbb{R}^{m \times n}$ takes the form $A = \vec{u}\vec{v}^\top$, with $\vec{u} \in \mathbb{R}^m$, $\vec{v} \in \mathbb{R}^n$, and $\vec{u}, \vec{v} \neq \vec{0}$. (Note that a matrix of this form is known as a *dyad*.) Find the rank of A.

HINT: Consider the quantity $A\vec{x}$ for arbitrary $\vec{x}$, i.e., what happens when you multiply any vector by matrix A.

(b) Show that for arbitrary $A, B \in \mathbb{R}^{m \times n}$,

$$\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B), \tag{3}$$

i.e., the rank of the sum of two matrices is less than or equal to the sum of their ranks.

HINT: First, show that $\mathcal{R}(A + B) \subseteq \mathcal{R}(A) + \mathcal{R}(B)$, meaning that any vector in the range of $A + B$ can be expressed as the sum of two vectors, each in the range of A and B, respectively. Remember that for any matrix A, $\mathcal{R}(A)$ is a subspace, and for any two subspaces S_1 and S_2, $\dim(S_1 + S_2) \leq \dim(S_1) + \dim(S_2)$.[1] *(Note that the sum of vector spaces $S_1 + S_2$ is not equivalent to $S_1 \cup S_2$, but is defined as $S_1 + S_2 := \{\vec{s_1} + \vec{s_2} | \vec{s_1} \in S_1, \vec{s_2} \in S_2\}$.)*

(c) Consider an $m \times n$ matrix A that takes the form $A = UV^\top$, with $U \in \mathbb{R}^{m \times k}$, $V \in \mathbb{R}^{n \times k}$. Show that the rank of A is less or equal than k. *HINT: Use parts (a) and (b), and remember that this decomposition can also be written as the* dyadic expansion

$$A = UV^\top = \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_k \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_k^\top \end{bmatrix} = \sum_{i=1}^{k} \vec{u}_i \vec{v}_i^\top, \tag{4}$$

for $U = \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_k \end{bmatrix}$ *and* $V = \begin{bmatrix} \vec{v}_1 & \dots & \vec{v}_k \end{bmatrix}$.

[1] This fact can be proved by taking a basis of S_1 and extending it to a basis of S_2 (during which we can only add *at most* $\dim(S_2)$ basis vectors). This extended basis must now also be a basis of $S_1 + S_2$. Thus, $\dim(S_1 + S_2) \leq \dim(S_1) + \dim(S_2)$.

4. Gram Schmidt

Any set of n linearly independent vectors in $\mathbb{R}^n$ could be used as a basis for $\mathbb{R}^n$. However, certain bases could be more suitable for certain operations than others. For example, an orthonormal basis could facilitate solving linear equations.

(a) Given a matrix $A \in \mathbb{R}^{n\times n}$, it could be represented as a multiplication of two matrices

$$A = QR,$$

where Q is an orthonormal in $\mathbb{R}^n$ and R is an upper-triangular matrix. For the matrix A, describe how Gram-Schmidt process could be used to find the Q and R matrices, and apply this to

$$A = \begin{bmatrix} 3 & -3 & 1 \\ 4 & -4 & -7 \\ 0 & 3 & 3 \end{bmatrix}$$

to find an orthogonal matrix Q and an upper-triangular matrix R.

(b) Given an invertible matrix $A \in \mathbb{R}^{n\times n}$ and an observation vector $b \in \mathbb{R}^n$, the solution to the equality

$$Ax = b$$

is given as $x = A^{-1}b$. For the matrix $A = QR$ from part (a), assume that we want to solve

$$Ax = \begin{bmatrix} 8 \\ -6 \\ 3 \end{bmatrix}.$$

By using the fact that Q is an orthonormal matrix, find v such that

$$Rx = v.$$

Then, given the upper-triangular matrix R in part (a) and v, find the elements of x sequentially.

(c) Given an invertible matrix $B \in \mathbb{R}^{n\times n}$ and an observation vector $c \in \mathbb{R}^n$, find the computational cost of finding the solution z to the equation $Bz = c$ by using the QR decomposition of B. Assume that Q and R matrices are available, and adding, multiplying, and dividing scalars take one unit of “computation".

As an example, computing the inner product $a^\top b$ is said to be $\mathcal{O}(n)$, since we have n scalar multiplication for each $a_i b_i$. Similarly, matrix vector multiplication is $\mathcal{O}(n^2)$, since matrix vector multiplication can be viewed as computing n inner products. The computational cost for inverting a matrix in $\mathbb{R}^n$ is $\mathcal{O}(n^3)$, and consequently, the cost grows rapidly as the set of equations grows in size. This is why the expression $A^{-1}b$ is usually not computed by directly inverting the matrix A. Instead, the QR decomposition of A is exploited to decrease the computational cost.

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.