EE 127/227A Midterm Review Problems

EE 127/227AT: Optimization Models in Engineering
Instructor: Profs. Gireeja Ranade, Venkat Anantharam
Authors: Chinmay Maheshwari, Chih-Yuan Chiu, Kshitij Kulkarni, Tim Li

October 5, 2023

Problem ("Convexity", Spring 2023 Midterm)

(a) Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Fix $a, b \in \mathbb{R}$. Prove that for any $x \in [a, b]$:

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b).$$

This problem is from [Boyd and Vandenberghe, Problem 1 a].

Solution: Define $\alpha = \frac{b-x}{b-a}$. Note that $\alpha \in [0, 1]$ and it holds that

$$1 - \alpha = \frac{x-a}{b-a}.$$

Next, note that

$$a\alpha + b(1-\alpha) = a \cdot \frac{b-x}{b-a} + b \cdot \frac{x-a}{b-a} = x.$$

Finally, using the definition of convexity of a function $f(\cdot)$ it holds that

$$\begin{aligned} f(x) = f(a\alpha + b(1-\alpha)) &\leq \alpha f(a) + (1-\alpha) f(b) \\ &= \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b). \end{aligned}$$

(b) Let n be a positive integer. The *probability simplex* on $\mathbb{R}^n$, denoted $\mathcal{P}_n$, is the set

$$\mathcal{P}_n = \left\{ \vec{x} \in \mathbb{R}^n | x_i \geq 0 \ \forall i, \ \sum_{i=1}^{n} x_i = 1 \right\} \qquad \text{where} \qquad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{0.1}$$

Is $\mathcal{P}_n$ convex? If yes, prove it. If no, justify your answer using an example.

Solution: Yes, $\mathcal{P}_n$ is convex. Let $\vec{x}, \vec{y} \in \mathcal{P}_n$ let $\theta \in [0, 1]$, and define $\vec{z} = \theta\vec{x} + (1-\theta)\vec{y}$. We show that $\vec{z} \in \mathcal{P}_n$. Indeed,

$$z_i = \underbrace{\theta}_{\geq 0} \underbrace{x_i}_{\geq 0} + \underbrace{(1-\theta)}_{\geq 0} \underbrace{y_i}_{\geq 0}$$

$$
\begin{aligned}
&\geq 0. \\
\sum_{i=1}^{n} z_i &= \sum_{i=1}^{n} (\theta x_i + (1-\theta) y_i) \\
&= \sum_{i=1}^{n} \theta x_i + \sum_{i=1}^{n} (1-\theta) y_i \\
&= \theta \underbrace{\sum_{i=1}^{n} x_i}_{=1} + (1-\theta) \underbrace{\sum_{i=1}^{n} y_i}_{=1} \\
&= \theta + (1-\theta) = 1.
\end{aligned}
$$

Thus $\vec{z} \in \mathcal{P}_n$ so $\mathcal{P}_n$ is convex.

Problem ("Shift Matrix", Spring 2023 Midterm)

Let $V \in \mathbb{R}^{n\times n}$ be a square orthonormal matrix, i.e., its columns are orthogonal and have norm 1:

$$V = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n-1} & \vec{v}_n \\ \downarrow & \downarrow & \cdots & \downarrow & \downarrow \end{bmatrix}. \tag{0.2}$$

Now, we define the shifted matrix $W \in \mathbb{R}^{n\times n}$, which is composed of the columns of V shifted to the left by 1 index and padded by a zero vector:

$$W = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow & \uparrow \\ \vec{v}_2 & \vec{v}_3 & \cdots & \vec{v}_n & \vec{0} \\ \downarrow & \downarrow & \cdots & \downarrow & \downarrow \end{bmatrix}. \tag{0.3}$$

(a) What is $\mathsf{rank}(V)$? What about $\mathsf{rank}(W)$? *You do not need to justify your answers.*

Solution: V is orthogonal, so it has full column rank. Therefore $\mathsf{rank}(V) = n$. Since $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a set of n linearly independent vectors, $\{\vec{v}_2, \ldots, \vec{v}_n\}$ is a set of $n-1$ linearly independent vectors. $\vec{0}$ is linearly dependent to all other vectors, so there are just $n-1$ linearly independent columns of W. Therefore $\mathsf{rank}(W) = n - 1$.

(b) Find a basis for the null space of $V - W$ and compute $\mathsf{rank}(V - W)$. *Show your work.*

Solution: Suppose $\vec{x} \in \mathcal{N}(V - W)$. Then,

$$\begin{aligned} \vec{0} &= (V - W)\vec{x} \\ \implies \vec{0} &= \left[\sum_{i=1}^{n-1} x_i(\vec{v}_i - \vec{v}_{i+1})\right] + x_n\vec{v}_n \\ \implies \vec{0} &= x_1\vec{v}_1 + \left[\sum_{i=1}^{n-1}(x_{i+1} - x_i)\vec{v}_{i+1}\right] \end{aligned}$$

Since v_i are all linearly independent, this implies that $x_n = \ldots = x_1 = 0$, which means that the null space is trivial. By Rank-Nullity Theorem, this means that $\mathsf{rank}(V - W) = n - \dim \mathcal{N}(V - W) = n$

Problem ("Singular value decomposition", Spring 2019 Midterm 1)

(13 points) The compact form of the singular value decomposition of a matrix $A \in \mathbb{R}^{3\times 3}$ is given as

$$A = \begin{bmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} \\ \frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}$$

(a) (2 points) What is the rank of A? Justify.

(b) (3 points) What is the dimension of the column space (range) of A? Write a basis for the column space (range) of A.

(c) (4 points) What is the dimension of the null space of $A^\top$? Write a basis for the null space of $A^\top$.

(d) (4 points) Let $\mathcal{B}_2$ denote the unit-norm ball in ℓ_2 norm: $\mathcal{B}_2 := \{\vec{z} \in \mathbb{R}^3 : \|\vec{z}\|_2 \leq 1\}$. Compute the minimum value of $\vec{x}^\top A\vec{y}$, where $\vec{x}$ and $\vec{y}$ are two vectors in $\mathcal{B}_2$; that is, find $\min_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y}$.

Solution: For convenience, define $U_r \in \mathbb{R}^{3\times 2}$, $\Sigma_r \in \mathbb{R}^{2\times 2}$, and $V_r \in \mathbb{R}^{2\times 3}$ by:

$$U_r := \begin{bmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} \\ \frac{1}{3} & 0 \end{bmatrix},$$

$$\Sigma_r := \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix},$$

$$V_r := \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \end{bmatrix}.$$

Then $A = U_r \Sigma_r V_r^\top$.

(a) A has two singular values, so $\text{rank}(A) = 2$.

(b) From (a), $\dim(R(A)) = 2$. Since $A = U_r \Sigma_r V_r^\top$ is a compact SVD of A, a basis for $R(A)$ is given by the set of columns of U_r, i.e.,:

$$\left\{ \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \right\}.$$

(c) By the Fundamental Theorem of Linear Algebra, $N(A^\top) = R(A)^\perp$. Since $R(A) \subset \mathbb{R}^3$, with $\dim(R(A)) = 2$, we have:

$$\dim(N(A^\top)) = \dim(R(A)^\perp) = \dim(\mathbb{R}^3) - \dim(R(A)) = 3 - 2 = 1$$

Let $\vec{v} := [v_1 \; v_2 \; v_3]^\top \in \mathbb{R}^3$ be given such that $\|\vec{v}\|_2 = 1$ and $N(A^\top) = \text{span}(\{\vec{v}\})$. Then $\vec{v}$ is orthogonal to the columns of U_r, which imply:

$$2v_1 + 2v_2 + v_3 = 0,$$

$$v_1 - v_2 = 0.$$

Thus, the following is a possible choice for $\vec{v}$ is: $\vec{v} = (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{4}{\sqrt{6}})$.

(d) Note that if $\vec{x}^\star, \vec{y}^\star \in \mathcal{B}_2 \times \mathcal{B}_2$ is a minimizer of the expression $\vec{x}^\top A\vec{y}$, then $(-\vec{x}^\star, \vec{y}^\star) \in \mathcal{B}_2 \times \mathcal{B}_2$ is a maximizer of the same expression. Thus:

$$\min_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y} = -\max_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y}.$$

Applying the Cauchy-Schwarz inequality and the definition of the spectral norm of A, we obtain that for any $\vec{x}_1, \vec{y}_1 \in \mathcal{B}_2$:

$$\begin{aligned}
\min_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y} &= -\max_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y} \\
&\geq -\|\vec{x}_1\|_2 \cdot \|A\vec{y}_1\|_2 \\
&\geq -\|\vec{x}_1\|_2 \cdot \|A\|_2 \cdot \|\vec{y}_1\|_2 \\
&\geq -\|A\|_2.
\end{aligned}$$

We claim that equality can be obtained by a specific value of $\vec{x} \in \mathcal{B}_2$ and $\vec{y} \in \mathcal{B}_2$. In particular, given orthogonal matrices $U, V \in \mathbb{R}^{3\times3}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{3\times3}$ such that $A = U\Sigma V^\top$ forms an SVD for A, we have $\vec{x}^\top A\vec{y} = \vec{x}^\top U\Sigma V^\top\vec{y}$. Since $\|A\|$ is the $(1,1)$-entry of Σ, the equality $\vec{x}^\top U\Sigma V^\top\vec{y} = \|A\|$ would hold if $\vec{x}^\top U = (1,0,0)^\top$ and $V^\top\vec{y} = (1,0,0)$, which is equivalent to requiring $\vec{x}$ and $\vec{y}$ to be the first column of U and the first column of V, respectively (i.e., the left singular vector and the right singular vector corresponding to the maximum singular value of A, respectively). We conclude that:

$$\min_{\vec{x},\vec{y}\in\mathcal{B}_2} \vec{x}^\top A\vec{y} = -\|A\|_2 = -3.$$

Problem ("All I need is Q", Spring 2020 Midterm)

(22 points) Consider a partially known matrix $A \in \mathbb{R}^{3\times 2}$, given by

$$A = \begin{bmatrix} ? & 1 \\ ? & 1 \\ ? & 1 \end{bmatrix}$$

where question marks denote unknown entries of A. We can write the compact QR decomposition of A in terms of $Q_1 \in \mathbb{R}^{3\times 2}$ and $R_1 \in \mathbb{R}^{2\times 2}$ as

$$A = Q_1 R_1 = \begin{bmatrix} 1 & q_{12} \\ 0 & q_{22} \\ 0 & q_{23} \end{bmatrix} \begin{bmatrix} ? & r_{12} \\ 0 & r_{22} \end{bmatrix} \tag{0.4}$$

for some unknown entry '?' and entries r_{12}, r_{22}, q_{12}, q_{22}, and q_{23}, which you will calculate below. Remember that the columns of Q_1 are orthonormal. Note that the '?' entries of A and R_1 are unknown and will remain unknown; you are NOT required to compute them.

(a) (5 points) Suppose $r_{22} > 0$. **Compute r_{12}, r_{22}, q_{12}, q_{22}, and q_{23}.** Show all your work.

(b) (12 points) Suppose we can write the full QR decomposition of A as

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ \vec{0}_{1\times 2} \end{bmatrix}, \tag{0.5}$$

where Q_1 and R_1 are as defined in Equation (0.4). Consider the least-squares problem:

$$p^\star = \min_{\vec{x}\in\mathbb{R}^2} \|A\vec{x} - \vec{b}\|_2^2$$

for A given in Equation (0.4) and some $\vec{b} \in \mathbb{R}^3$. Consider the following two ways of rewriting this least squares problem in terms of Q_1, Q_2, and R_1:

Strategy 1:

$$\begin{aligned} \|\vec{b} - A\vec{x}\|_2^2 &\overset{(I)}{=} \|Q^\top \vec{b} - Q^\top A\vec{x}\|_2^2 \\ &= \|Q_1^\top \vec{b} - R_1\vec{x}\|_2^2 + \|Q_2^\top \vec{b}\|_2^2. \end{aligned}$$

Strategy 2:

$$\begin{aligned} \|\vec{b} - A\vec{x}\|_2^2 &= \|\vec{b} - Q_1R_1\vec{x}\|_2^2 \\ &\overset{(II)}{=} \|\vec{Q}_1^\top b - Q_1^\top Q_1 R_1\vec{x}\|_2^2 \\ &\overset{(III)}{=} \|\vec{Q}_1^\top b - R_1\vec{x}\|_2^2 \end{aligned}$$

Determine whether the following labeled steps in the reformulations above are correct or incorrect and justify your answer. When evaluating the correctness of an equality, consider only that specific equality's correctness—i.e., ignore all earlier steps.

(i) Equality (I): $\|\vec{b} - A\vec{x}\|_2^2 \overset{(I)}{=} \|Q^\top\vec{b} - Q^\top A\vec{x}\|_2^2$.

(ii) Equality (II): $\|\vec{b} - Q_1 R_1 \vec{x}\|_2^2 \overset{(II)}{=} \|\vec{Q}_1^\top b - Q_1^\top Q_1 R_1 \vec{x}\|_2^2$.

(iii) Equality (III): $\|\vec{Q}_1^\top b - Q_1^\top Q_1 R_1 \vec{x}\|_2^2 \overset{(III)}{=} \|\vec{Q}_1^\top b - R_1 \vec{x}\|_2^2$.

(c) (5 points) Now consider a different matrix $A = QR$, unrelated to the matrix A in previous parts. Here, let

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$
$$R = \begin{bmatrix} R_1 \\ \vec{0}_{1\times 2} \end{bmatrix}$$

where $R \in \mathbb{R}^{3\times 2}$ and $R_1 \in \mathbb{R}^{2\times 2}$ is a completely unknown **invertible** upper triangular matrix. Let

$$\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Again consider the least squares optimization problem:

$$p^\star = \min_{\vec{x}\in\mathbb{R}^2} \|A\vec{x} - \vec{b}\|_2^2.$$

Find the optimal value $p^\star$. Your answer should be a real number; it should NOT be an expression involving A, Q, R, R_1, or $\vec{b}$.

Solution:

(a) Applying the rules of matrix-vector multiplication to (0.4), we obtain (specifically, we focus on the second column of A):

$$1 = r_{12} + r_{22}q_{12},$$
$$1 = q_{22}r_{22},$$
$$1 = q_{23}r_{22}.$$

Since the columns of Q_1 form an orthonormal set, we also have:

$$q_{12} = 0,$$
$$q_{12}^2 + q_{22}^2 + q_{23}^2 = 1.$$

Thus, we obtain $r_{12} = 1$. Rearranging the other equalities, we have:

$$q_{22} = \frac{1}{r_{22}},$$
$$q_{23} = \frac{1}{r_{22}},$$
$$q_{22}^2 + q_{23}^2 = 1,$$

from which we obtain $r_{22} = \sqrt{2}$, and $q_{22} = q_{23} = \frac{1}{\sqrt{2}}$. To summarize, $(r_{12}, r_{22}, q_{12}, q_{22}, q_{23}) = \left(1, \sqrt{2}, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

(b) (i) is true because $Q^\top$ is an orthogonal matrix, and is thus norm-preserving. (ii) is in general false, since Q_1 is *not* an orthogonal matrix (it is not even square). As an example, consider the scenario where $\vec{x} = (0, 0) \in \mathbb{R}^2$, in which case the values:

$$\|\vec{b} - Q_1 R_1 \vec{x}\|_2^2 = \|\vec{b}\|_2^2 = 1^2 + 2^2 + 3^2 = 14$$

and:

$$\|Q_1^\top \vec{b} - Q_1^\top Q_1 R_1 \vec{x}\|_2^2 = \|Q_1^\top \vec{b}\|_2^2 = 1^2 + \frac{2^2}{2} + \frac{3^2}{2} = \frac{15}{2}$$

are not equal. (iii) is true because the columns of Q_1 form an orthonormal set, which implies $Q_1^\top Q_1 = I_{2\times 2}$, where $I_{2\times 2}$ denotes the 2×2 identity matrix.

(c) From "**Strategy 1**" in part (b) of this question, we obtain the following lower bound, which is independent of the choice of $\vec{x}$:

$$\begin{aligned} \|\vec{b} - A\vec{x}\|_2^2 &= \|Q_1^\top \vec{b} - R_1 \vec{x}\|_2^2 + \|Q_2^\top \vec{b}\|_2^2 \\ &\geq \|Q_2^\top \vec{b}\|_2^2. \end{aligned}$$

Moreover, equality is attained at (and only at) the vector $\vec{x}^\star \in \mathbb{R}^3$ satisfying $Q_1^\top \vec{b} - R_1 \vec{x}$, i.e., at $\vec{x}^\star = R_1^{-1} Q_1^\top \vec{b}$. Thus, we have $p^\star = \|Q_2^\top \vec{b}\|_2^2$. For this problem, since $R_1 \in \mathbb{R}^{2\times 2}$, we have $Q_1 \in \mathbb{R}^{3\times 2}$, so Q_2 is the third column of Q, i.e., $Q_2 = (0, 1, 0) \in \mathbb{R}^3$. Thus:

$$\begin{aligned} p^\star &= \|Q_2^\top \vec{b}\|_2^2 \\ &= \left\| \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\|_2^2 \\ &= 4. \end{aligned}$$

Problem ("Vector Calculus", Spring 2023)

1. Let $A \in \mathbf{R}^{n\times n}$ be an $n \times n$ symmetric matrix. Compute the gradient with respect to $\vec{x}$ of the function $f\colon \mathbb{R}^n \setminus \{\vec{0}\} \to \mathbb{R}$ given by:

$$f(\vec{x}) \doteq \frac{\vec{x}^\top A\vec{x}}{\vec{x}^\top\vec{x}}. \tag{0.6}$$

Hint: Recall the quotient rule for finding the gradient of $h(\vec{x}) = \frac{n(\vec{x})}{d(\vec{x})}$ where n and d are scalar-valued functions:

$$\nabla h(\vec{x}) = \frac{d(\vec{x})\nabla n(\vec{x}) - n(\vec{x})\nabla d(\vec{x})}{(d(\vec{x}))^2}. \tag{0.7}$$

Solution: Define $n(\vec{x}) = \vec{x}^\top A\vec{x}$ and $d(\vec{x}) = \vec{x}^\top\vec{x}$. The gradients of these functions are

$$\nabla n(\vec{x}) = (A + A^\top)\vec{x} = 2A\vec{x} \tag{0.8}$$
$$\nabla d(\vec{x}) = 2\vec{x}. \tag{0.9}$$

Then $f(\vec{x}) = \frac{n(\vec{x})}{d(\vec{x})}$, so we have

$$\begin{aligned}\nabla f(\vec{x}) &= \frac{d(\vec{x})\nabla n(\vec{x}) - n(\vec{x})\nabla d(\vec{x})}{(d(\vec{x}))^2}\\ &= \frac{[\vec{x}^\top\vec{x}][2A\vec{x}] - [\vec{x}^\top A\vec{x}][2\vec{x}]}{(\vec{x}^\top\vec{x})^2}\\ &= 2\frac{A\vec{x}\vec{x}^\top\vec{x} - \vec{x}\vec{x}^\top A\vec{x}}{(\vec{x}^\top\vec{x})^2}\\ &= \frac{2}{\vec{x}^\top\vec{x}}\left(A - \frac{\vec{x}^\top A\vec{x}}{\vec{x}^\top\vec{x}}I\right)\vec{x}.\end{aligned}$$

2. Let $\vec{u} \in \mathbb{R}^n$. Compute the Jacobian with respect to $\vec{x}$ of the function $\vec{g}\colon \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\vec{g}(\vec{x}) \doteq \vec{x}(\vec{x}^\top\vec{u}). \tag{0.10}$$

Solution:

Define $h(\vec{x}) = \vec{u}^\top\vec{x}$. Then $\frac{\partial h}{\partial x_i}(\vec{x}) = u_i$. Also, define $g_i(\vec{x}) = h(\vec{x})x_i$ and we can compute the partial derivatives as:

$$\frac{\partial g_i}{\partial x_i}(\vec{x}) = h(\vec{x}) + x_i\frac{\partial h}{\partial x_i}(\vec{x}) = \vec{u}^\top\vec{x} + u_i x_i \tag{0.11}$$
$$\frac{\partial g_i}{\partial x_j}(\vec{x}) = x_i\frac{\partial h}{\partial x_j}(\vec{x}) = x_i u_j. \tag{0.12}$$

If we stack these partial derivatives in a Jacobian matrix it follows that:

$$D\vec{g}(\vec{x}) = \vec{x}\vec{u}^\top + (\vec{u}^\top\vec{x})I. \tag{0.13}$$

Problem ("Low-rank Matrix Completion", Spring 2023)

Consider a matrix $A \in \mathbb{R}^{m\times n}$. If some entries are corrupted, one principled way to identify A is to find the matrix $B \in \mathbb{R}^{m\times n}$ of minimal rank that agrees with A on all known entries. This can be formulated as an optimization problem whose objective function is $\mathsf{rank}(B)$. Because the $\mathsf{rank}(\cdot)$ function is not continuous, we use the intuition that a low-rank matrix will only have a few nonzero singular values, and instead use the sum-of-singular-values function as the objective:

$$f(B) \doteq \sum_{i=1}^{\mathsf{rank}(B)} \sigma_i\{B\} \tag{0.14}$$

where $\sigma_i\{B\}$ is the i^{th} largest singular value of B. In this problem we will explore some properties of f.

(a) **Prove that**

$$f(B) \le \max_{\substack{C\in\mathbb{R}^{m\times n}\\ \|C\|_2\le 1}} \mathsf{Tr}\left(C^\top B\right). \tag{0.15}$$

Hint. Expand B into its SVD. Try to find a $D \in \mathbb{R}^{m\times n}$ such that $\|D\|_2 = 1$ and $\mathsf{Tr}\left(D^\top B\right) = f(B)$.

Hint. You may use the cyclic property of traces *without proof*. If XYZ and ZXY are valid matrix products then $\mathsf{Tr}(XYZ) = \mathsf{Tr}(ZXY)$.

Solution:

Let $r \doteq \mathsf{rank}(B)$. Let $B = U_r\Sigma_r V_r^\top$ be the compact SVD of B. Let $D = U_r V_r^\top$. Note that $\|D\|_2 = 1$. This is because: (i) the SVD of D is expressed as $D = U_r I_{r\times r} V_r^\top$ where $I_{r\times r}$ is an identity matrix and (ii) the 2-norm of a matrix is defined to be maximum singular value of that matrix.

Next, note that

$$\begin{aligned}
\max_{\substack{C\in\mathbb{R}^{m\times n}\\ \|C\|_2\le 1}} \mathsf{Tr}\left(C^\top B\right) &\ge \mathsf{Tr}\left(D^\top B\right) \\
&= \mathsf{Tr}\left(V_r U_r^\top U_r \Sigma_r V_r^\top\right) \\
&= \mathsf{Tr}\left(V_r \Sigma_r V_r^\top\right) \\
&= \mathsf{Tr}\left(V_r^\top V_r \Sigma_r\right) \\
&= \mathsf{Tr}\left(\Sigma_r\right) \\
&= \sum_{i=1}^{r} \sigma_i\{B\} \\
&= f(B).
\end{aligned}$$

(b) **Prove that**

$$f(B) \ge \max_{\substack{C\in\mathbb{R}^{m\times n}\\ \|C\|_2\le 1}} \mathsf{Tr}\left(C^\top B\right). \tag{0.16}$$

Hint. Let $r \doteq \mathsf{rank}(B)$ and expand B into its *outer product* SVD, i.e., $B = \sum_{i=1}^{r} \sigma_i\{B\}\vec{u}_i\vec{v}_i^\top$.

Hint. You may use the cyclic and linearity properties of traces *without proof.* If XYZ and ZXY are valid matrix products then $\mathsf{Tr}(XYZ) = \mathsf{Tr}(ZXY)$. Also, $\mathsf{Tr}(\alpha X + \beta Y) = \alpha \mathsf{Tr}(X) + \beta \mathsf{Tr}(Y)$ for $\alpha, \beta \in \mathbb{R}$.

Solution:

Let $r = \mathsf{rank}(B)$. Let $B = \sum_{i=1}^{r} \sigma_i\{B\}\vec{u}_i\vec{v}_i^\top$ be an outer product SVD of B. For any $C \in \mathbb{R}^{m\times n}$ such that $\|C\|_2 \le 1$, we have

$$\begin{aligned}
\mathsf{Tr}\left(C^\top B\right) &= \mathsf{Tr}\left(C^\top \left(\sum_{i=1}^{r} \sigma_i\{B\}\vec{u}_i\vec{v}_i^\top\right)\right) \\
&= \mathsf{Tr}\left(\sum_{i=1}^{r} \sigma_i\{B\} C^\top \vec{u}_i\vec{v}_i^\top\right) \\
&= \sum_{i=1}^{r} \sigma_i\{B\} \mathsf{Tr}\left(C^\top \vec{u}_i\vec{v}_i^\top\right) \\
&= \sum_{i=1}^{r} \sigma_i\{B\} \mathsf{Tr}\left(\vec{v}_i^\top C^\top \vec{u}_i\right) \\
&= \sum_{i=1}^{r} \sigma_i\{B\} (\vec{v}_i^\top C^\top \vec{u}_i) \\
&\le \sum_{i=1}^{r} \sigma_i\{B\} \|C\vec{v}_i\|_2 \|\vec{u}_i\|_2 \\
&\le \sum_{i=1}^{r} \sigma_i\{B\} \underbrace{\|C\|_2}_{\le 1} \underbrace{\|\vec{v}_i\|_2}_{=1} \underbrace{\|\vec{u}_i\|_2}_{=1} \\
&\le \sum_{i=1}^{r} \sigma_i\{B\} \\
&= f(B).
\end{aligned}$$

This holds for all C such that $\|C\|_2 \le 1$, so taking the max over C gets

$$f(B) \ge \max_{\substack{C \in \mathbb{R}^{m\times n} \\ \|C\|_2 \le 1}} \mathsf{Tr}\left(C^\top B\right) \tag{0.17}$$

as desired.

Optional Problem ("Symmetric Matrices", Spring 2023)

1. Let $A \in \mathbf{R}^{n\times n}$ be a square matrix. Prove that if A is symmetric then A^{2k} is symmetric positive semidefinite for all integers $k > 1$.

   Solution:

   Many different proofs. Diagonalizing A we get $A = U\Lambda U^\top$. Then $A^k = U\Lambda^{2k}U^\top$ and $\Lambda^{2k} \succeq 0$.

2. Prove that if $A \in \mathbf{R}^{n\times n}$ is symmetric then its *matrix exponential*, defined as $e^A \in \mathbf{R}^{n\times n}$ given by

$$e^A = I + A + \frac{1}{2}A^2 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k \tag{0.18}$$

   is symmetric positive definite.

   Solution: Diagonalizing $A = U\Lambda U^\top$, we get

$$\begin{aligned}
e^A &= \sum_{k=0}^{\infty} \frac{1}{k!}A^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}(U\Lambda U^\top)^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}U\Lambda^k U^\top \\
&= U\left(\sum_{k=0}^{\infty} \frac{1}{k!}\Lambda^k\right)U^\top \\
&= U\left(\sum_{k=0}^{\infty} \frac{1}{k!}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}^k\right)U^\top \\
&= U\begin{bmatrix}\sum_{k=0}^{\infty}\frac{\lambda_1^k}{k!} & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty}\frac{\lambda_n^k}{k!}\end{bmatrix}U^\top \\
&= U\begin{bmatrix}e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n}\end{bmatrix}U^\top.
\end{aligned}$$

   This is a symmetric matrix whose eigenvalues are $e^{\lambda_i} > 0$, hence it is positive definite.