Problem 1.

Part (a): Write a detailed description of the singular value decomposition (SVD) of a matrix $A \in \mathbb{R}^{m\times n}$. In your description, be sure to touch on the following points:

A. When does the SVD exist?

B. Is the SVD unique?

C. How does the SVD reveal the four fundamental spaces $\text{col}(A)$, $\ker(A)$, $\text{col}(A^T)$, $\ker(A^T)$, and how does it reveal the column rank of A?

D. One other property of the SVD that you find interesting.

Proof. Given a matrix $A \in \mathbb{C}^{m\times n}$, SVD gives us a decomposition in the form of:

$$A = U\Sigma V^T,$$

where:

- $U \in \mathbb{C}^{m\times m}$ is a unitary matrix (i.e., $U^T U = UU^T = I_m$),
- $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with non-negative real entries on the diagonal, called the singular values of A. They are represented by $\sigma_1, \sigma_2, \ldots, \sigma_r$ (where $r = \min(m, n)$)
- $V \in \mathbb{C}^{n\times n}$ is a unitary matrix (i.e., $V^T V = VV^T = I_n$).

The columns of U and V are called the left and right singular vectors of A, respectively.

A: Existence The SVD exists for any matrix $A \in \mathbb{C}^{m\times n}$ (or $\mathbb{R}^{m\times n}$) regardless of its dimensions, rank, or invertibility.

B: Uniqueness The SVD is not unique in general; while the singular values (entries of Σ) are uniquely determined up to their order.

C. Fundamental subspaces and rank:

- The **column space** of A is given by the span of the first r columns of U.
- The **null space** of A, denoted $\ker(A)$, is given by the span of the last $n - r$ columns of V
- The **row space** of A, denoted $\text{col}(A^T)$, is spanned by the first r columns of V.
- The **left null space** of A, denoted $\ker(A^T)$, spanned by the last $m - r$ columns of U

The rank of A is equal to the number of non-zero singular values, i.e., $\text{rank}(A) = r$.

D. Interesting property of the SVD: As we learned in class SVD provides a nice low-rank approximation of any matrix (nonnegatative linear combination of rank-1 matrices). Specifically, for any $k \leq r$, the matrix:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^H$$

is the rank-k approximation of A that minimizes $\|A - A_k\|_F$ (Frobenius norm) or $\|A - A_k\|_2$ (spectral norm). This is particularly useful in dimensionality reduction techniques like PCA. $\square$

Part (b): Present (on a chalkboard, or via a notetaking app on Zoom) your description of the SVD to at least one other student in Math 56. List who you presented to and anyone who you listened to. Feel free to revise your description of the SVD with any feedback from your peer(s).

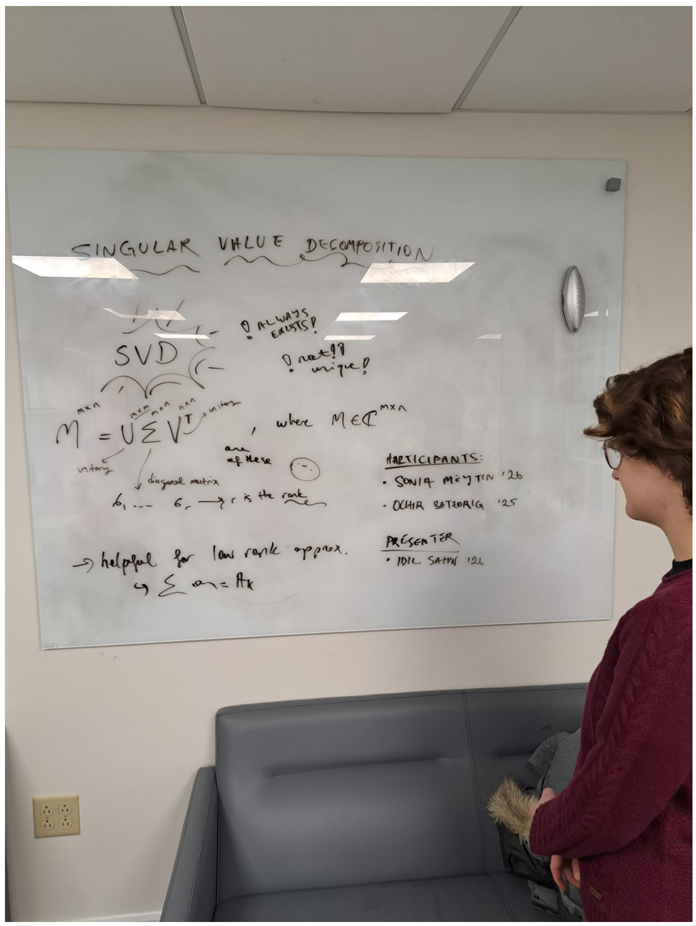

Figure 1: Presenting SVD

Problem 2.

Part (a): Let $A \in \mathbb{R}^{n \times n}$, and let $Q \in \mathbb{R}^{n \times n}$ be an arbitrary orthogonal matrix. Show that the similarity transformation QAQ^T has the same eigenvalues of A, i.e., the eigenvalues are not disturbed.

Proof. Let $B = QAQ^T = QAQ^{-1}$ as we know $Q^{-1} = Q^T$. Let λ be an eigenvalue of A, so that $Av = \lambda v$ for some nonzero vector v (the corresponding eigenvector). Now consider $u = Qv$, which implies $Q^{-1}u = v$. Since Q is an orthogonal matrix and hence invertible, we know that:

$$Av = \lambda v \implies A(Q^{-1}u) = \lambda(Q^{-1}u)$$

Multiplying both sides of this equation by Q, we get:

$$QA(Q^{-1}u) = Q\lambda(Q^{-1}u) = \lambda Q(Q^{-1}u) = \lambda u.$$

This shows that u is an eigenvector of QAQ^T with the same eigenvalue λ. As this holds true for all eigenvalues of A, we conclude that A and QAQ^T must have the same eigenvalues. □

Part (b): Let $A \in \mathbb{R}^{m\times n}$, and let $Q_L \in \mathbb{R}^{m\times m}$ and $Q_R \in \mathbb{R}^{n\times n}$ be arbitrary orthogonal matrices. Show that $Q_L A Q_R$ has the same singular values of A, i.e., left or right multiplication by orthogonal matrices does not disturb the singular values.

Proof. Let the singular value decomposition (SVD) of A be:

$$A = U\Sigma V^T,$$

where:

- $U \in \mathbb{R}^{m\times m}$ is an orthogonal matrix ($U^T U = I$),
- $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix ($V^T V = I$),
- $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of A (with $r = \min(m, n)$).

Now consider the matrix $B = Q_L A Q_R$. Substituting the SVD of A into B, we get:

$$B = Q_L A Q_R = Q_L(U\Sigma V^T)Q_R = B = (Q_L U)\Sigma(V^T Q_R)$$

Let $\tilde{U} = Q_L U$ and $\tilde{V} = Q_R V$. Since Q_L and Q_R are orthogonal matrices, recall that $Q_L^T Q_L = I$ and $Q_R^T Q_R = I$. This gives us:

$$\tilde{U}^T\tilde{U} = (Q_L U)^T(Q_L U) = U^T Q_L^T Q_L U = U^T U = I,$$

$$\tilde{V}^T\tilde{V} = (Q_R V)^T(Q_R V) = V^T Q_R^T Q_R V = V^T V = I.$$

Thus we observe that $\tilde{U}$ and $\tilde{V}$ are also orthogonal matrices. Therefore, the SVD of B is:

$$B = \tilde{U}\Sigma\tilde{V}^T.$$

Note that Σ (the diagonal matrix of singular values) is unchanged, the singular values of $B = Q_L A Q_R$ are the same as those of A. This proves our statement. □

Problem 3.

Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times n}$, $x \in \mathbb{R}^n$. In "Big O" notation, how many flops are required to compute the matrix-vector product Ax using the standard algorithm? What about the matrix-matrix product AB? Provide reasoning for your answer.

Proof. We consider this proof in two parts:

- Matrix-vector product Ax: Note that we have to calculate the dot product of v with each row of A (m rows, m dot products), and each dot product involves n multiplications and $n-1$ additions (v and also each row n elements). This gives us $(2n-1)m = O(nm)$ floating point operations.
- Matrix-matrix product AB: Note that we have to calculate the dot product of each column of B with each row of A (m rows, n columns, mn dot products), and each dot product involves n multiplications and $n-1$ additions (each column of B and also each row of A has n elements). This gives us $(2n-1)mn = O(n^2m)$ floating point operations.

□

Problem 4.

Let $u, x \in \mathbb{R}^n$.

Part (a): What is the rank of $U = uu^T$? What are its eigenvalues?

Proof. Consider this in two parts:

- Recall that the rank of U equals the number of its linearly independent columns. Since $U = uu^T$, remembering the column picture of matrix multiplication, every column of U is a scalar multiple of u. Thus, U has only one linearly independent column, giving us $\text{rank}(U) = 1$.
- We know since it has rank 1, we will have only one nonzero eigenvalue and eigenvector. Now observe that:
$$Uu = (uu^T)u = u(u^Tu) = u\|u\|^2$$
Since Uu this gives us a scalar multiple of u, we note that $\|u\|^2$ is the only nonzero eigenvalue and the remaining eigenvalues are equal to zero.

□

Part (b): In "Big O" notation, how many flops are required to compute $z = Ux$ when computed as $z = (uu^T)x$? What about when computed as $z = u(u^Tx)$?

Proof. Consider this in two parts:

- $z = (uu^T)x$: We first compute uu^T, which has n multiplications and for each element of u^T, giving us n^2 operations. As proven in problem 3, the product Ux where $U = uu^T$ has $O(n^2)$ flops. Adding them together, we get $n^2 + O(n^2) = O(n^2)$ flops.
- $z = u(u^Tx)$: We first compute u^Tx, which is a scalar (dot product) with n multiplications and $n-1$ additions, giving us $2n-1$ operations. Given u^Tx is a scalar, the next step $u(u^Tx)$ only involves n operations (one per element of u) This gives us $2n-1+n = O(n)$ flops.

□

Problem 5.

Let $U, V \in \mathbb{R}^{n\times k}$ with $\text{rank}(U) = \text{rank}(V) = k$. What is $\text{rank}(UV^T)$?

Proof. For any $x \in \mathbb{R}^n$, observe that if $UV^Tx = 0$ then $V^Tx = 0$, as U is a full-rank matrix. This implies that x is in the left nullspace of V^T; which has dimension $n-k$ as V^T has rank k (rank nullity theorem). Subsequently, the size of the nullspace of UV^Tx is also $n-k$. Recall by rank nullity theorem $rank(UV^T) + dim(ker(UV^T)) = n \implies rank(UV^T) = n-(n-k) = k$. □

Problem 6.

Show that $\|x\|_1 \leq n\|x\|_\infty$ for any $x \in \mathbb{R}^n$.

Proof. Recall that the ℓ_1-*norm* of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

and the ℓ_∞-*norm* of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|.$$

Observe that by definition, $|x_i| \leq \max_{1\leq i\leq n} |x_i|$ for all i. This gives us:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \leq \sum_{i=1}^{n} \max_{1\leq i\leq n} |x_i| = n \cdot \max_{1\leq i\leq n} |x_i| = n\|x\|_\infty$$

This proves our statement. □

Problem 7.

Let $C \in \mathbb{R}^{n\times n}$ be a symmetric positive definite (SPD) matrix and let

$$\langle u, v\rangle_C := u^TCv$$

be the C-weighted inner product for vectors $u, v \in \mathbb{R}^n$. Verify that $\|\cdot\|_C = \sqrt{\langle\cdot,\cdot\rangle_C}$ satisfies all properties of a norm on $\mathbb{R}^n$.

Proof. In order to prove that it satisfies all properties of a norm on $\mathbb{R}^n$; we need to prove the following three conditions:

A. **Non-negativity:** $\|u\|_C \geq 0$ and $\|u\|_C = 0 \iff u = 0$.

$$\|u\|_C = \sqrt{u^TCu}.$$

Since C is symmetric positive definite (SPD), $u^TCu \geq 0$ for all $u \in \mathbb{R}^n$, and $u^TCu = 0$ if and only if $u = 0$. Taking the square root preserves non-negativity, thus:

$$\|u\|_C \geq 0 \quad \text{and} \quad \|u\|_C = 0 \iff u = 0.$$

B. **Scalar multiplication/Positive homogeneity:** For any scalar $s \in \mathbb{R}$ and vector $u \in \mathbb{R}^n$, we need $\|su\|_C = |s|\|u\|_C$.

Let $s \in \mathbb{R}$ and $u \in \mathbb{R}^n$. Then:

$$\|su\|_C = \sqrt{\langle su, su\rangle_C} = \sqrt{(su)^T C(su)}.$$

Expanding the terms gives us:

$$(\alpha u)^T C(\alpha u) = \alpha^2 u^T Cu.$$

Taking the square root gives us:

$$\|su\|_C = \sqrt{s^2 u^T Cu} = |s|\sqrt{u^T Cu} = |s|\|u\|_C.$$

C. **Triangle inequality:** For any $u, v \in \mathbb{R}^n$, $\|u+v\|_C \leq \|u\|_C + \|v\|_C$. By definition of the weighted C-product, we know that:

$$\|u+v\|_C = \sqrt{\langle u+v, u+v\rangle_C} = \sqrt{(u+v)^T C(u+v)}.$$

Expanding the terms:

$$(u+v)^T C(u+v) = (u^T + v^T)C(u+v) = u^T Cu + u^T Cv + v^T Cu + v^T Cv$$

Now notice that we know $(u^T Cv)^T = v^T C^T u = v^T Cu$ as C is a symmetric matrix. Thus, combining all symmetric terms gives us:

$$(u+v)^T C(u+v) = u^T Cu + 2u^T Cv + v^T Cv$$

By the Cauchy-Schwarz inequality for the C-weighted inner product:

$$2u^T Cv \leq 2\sqrt{u^T Cu}\sqrt{v^T Cv}.$$

Thus we can bound the expression with:

$$(u+v)^T C(u+v) \leq u^T Cu + 2\sqrt{u^T Cu}\sqrt{v^T Cv} + v^T Cv$$

Notice that the right hand side forms a perfect square. Rewriting it so, we get:

$$(u+v)^T C(u+v) \leq (\sqrt{u^T Cu} + \sqrt{v^T Cv})^2$$

Taking the square root to get the C-weighted norm, we get:

$$\|u+v\|_C = \sqrt{(u+v)^T C(u+v)} \leq \sqrt{u^T Cu} + \sqrt{v^T Cv} = \|u\|_C + \|v\|_C$$

Bringing all three properties together, we see that the C-weighted norm satisfies all properties of a norm on $\mathbb{R}^n$.

$\square$

Problem 8.

Let $D = \operatorname{diag}(d_1, \ldots, d_n)$ where $d_i > 0$ for each i, and let $\|\cdot\|_D = \sqrt{\langle \cdot, \cdot \rangle_D}$. Show that $\|\cdot\|_D$ is equivalent to $\|\cdot\|_2$, i.e., find constants C_1 and C_2 such that

$$C_1\|x\|_2 \leq \|x\|_D \leq C_2\|x\|_2 \quad \text{for all } x \in \mathbb{R}^n.$$

Proof. Recall that the L_2-norm of a vector $x \in \mathbb{R}^n$ is given by:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Recall also that $\|x\|_D = \sqrt{x^T Dx}$ following the definition from problem 7. Let $d_{\min} = \min_{1\leq i\leq n} d_i$ and $d_{\max} = \max_{1\leq i\leq n} d_i$. For any $x \in \mathbb{R}^n$, since D is a diagonal matrix, $x^T Dx$ can be expressed as:

$$x^T Dx = \sum_{i=1}^{n} d_i x_i^2.$$

Since $d_i \leq d_{\max}$ for all i, we have:

$$x^T Dx \leq d_{\max} \sum_{i=1}^{n} x_i^2 = d_{\max}\|x\|_2^2.$$

Taking the square root:

$$\|x\|_D = \sqrt{x^T Dx} \leq \sqrt{d_{\max}}\|x\|_2.$$

Thus, $C_2 = \sqrt{d_{\max}}$. Similarly, given we know $d_{\min} \leq d_i$ for all i, we have:

$$x^T Dx \geq d_{\min} \sum_{i=1}^{n} x_i^2 = d_{\min}\|x\|_2^2.$$

Taking the square root:

$$\|x\|_D = \sqrt{x^T Dx} \geq \sqrt{d_{\min}}\|x\|_2.$$

Thus, $C_1 = \sqrt{d_{\min}}$. Bringing these together, we get $\sqrt{d_{\min}}\|x\|_2 \leq \|x\|_D \leq \sqrt{d_{\max}}\|x\|_2$, which proves our statement.

□

Problem 9.

Let C and $\|\cdot\|_C$ be as in Problem 7. Show that $\|\cdot\|_C$ is equivalent to $\|\cdot\|_2$, i.e., find constants C_1 and C_2 such that

$$C_1\|x\|_2 \leq \|x\|_C \leq C_2\|x\|_2 \quad \text{for all } x \in \mathbb{R}^n.$$

Proof. Recall that since C is symmetric positive definite, all its eigenvalues are real and positive. Let the eigenvalues of C be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$. Define:

$$\lambda_{\min} = \min_i \lambda_i, \quad \lambda_{\max} = \max_i \lambda_i.$$

Since C is positive definite, we have $\lambda_{\min} > 0$. For any $x \in \mathbb{R}^n$, using the spectral decomposition of C, we can write:

$$x^T C x = x^T Q \Lambda Q^T x,$$

where Q is an orthogonal matrix and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Now observe that $Q^T x = v$ for some vector v, furthermore $v^T = x^T Q$ This simplifies our expression to

$$(x^T Q)\Lambda(Q^T x) = v^T \Lambda v$$

Observe that since Λ is a diagonal matrix with positive values, the proof follows directly from Problem 8, with constants $C_1 = \sqrt{\lambda_{\min}}$ and $C_2 = \sqrt{\lambda_{\max}}$.

$\square$