EECS 127/227AT | Optimization Models in Engineering | UC Berkeley | Spring 2024

Midterm

1. **Honor Code (0 pts)**

   Please copy the following statement in the space provided below and sign your name.

   As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules and do this exam on my own.

   If you do not copy the honor code and sign your name, you will get a 0 on the exam.

   Solution:

2. **Favorites (2 pts)**

   (a) (1 pts) What is your favorite book or book series?

   Solution: Any answer is fine.

   (b) (1 pts) Who is the speaker or writer of your favorite inspirational quote?

   Solution: Any answer is fine.

3. **SID (3 pts)**

   When the exam starts, write your SID at the top of every page. *No extra time will be given for this task.*

4. Singular Values (10 pts)

(a) (4 pts) Suppose $A \in \mathbb{R}^{3\times 2}$ is a matrix such that $A^\top A$ is given by

$$A^\top A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. \tag{1}$$

What are the singular values of A? *Justify your answer(s).*

Solution: The singular values are the square roots of the eigenvalues of $A^\top A$. The eigenvalues of $A^\top A$ are 5 and 3, since those are the diagonal entries of the diagonal matrix in the spectral decomposition.

Therefore, the singular values of A are $\sqrt{5}$ and $\sqrt{3}$.

(b) (6 pts) Suppose that $B \in \mathbb{R}^{3\times 2}$ has singular values 0, $\sqrt{2}$, and $\sqrt{7}$. Let $C = \begin{bmatrix} B & -B & 3I_3 \end{bmatrix} \in \mathbb{R}^{3\times 7}$, where $I_3 \in \mathbb{R}^{3\times 3}$ is the 3×3 identity matrix. **What are the singular values of C?** *Show your work and justify your answer(s).*

HINT: Consider the matrix $CC^\top \in \mathbb{R}^{3\times 3}$.

Solution: To find the singular values of C, we consider $CC^\top \in \mathbb{R}^{3\times 3}$. Note that we consider this matrix rather than $C^\top C \in \mathbb{R}^{7\times 7}$ because the former is smaller, and we get the following simplification:

$$CC^\top = \begin{bmatrix} B & -B & 3I \end{bmatrix} \begin{bmatrix} B \\ -B \\ 3I \end{bmatrix} = 2BB^\top + 9I. \tag{2}$$

By the shift and scale properties of eigenvalues, the eigenvalues of $CC^\top$ are $9 + 2\times$ the eigenvalues of $BB^\top$. Since the eigenvalues of $BB^\top$ are the squared singular values of B, we know that the eigenvalues of $BB^\top$ are 0, 2, and 7. Thus the eigenvalues of $CC^\top$ are 9, 13 and 23. Thus the nonzero singular values of C are 3, $\sqrt{13}$, and $\sqrt{23}$.

5. Convex Functions (10 pts)

(a) (4 pts) **Show that the function $f\colon \mathbb{R}^n \to \mathbb{R}$ given by $f(\vec{x}) \doteq \|\vec{x}\|_2^2$ is convex.**

NOTE: You may use the gradient and Hessian of f, which were computed in lecture and homework, but the convexity of f should be proved via "first principles" (zeroth/first/second order conditions, or other equivalent conditions for convexity).

Solution: The gradient and Hessian of f are

$$\nabla f(\vec{x}) = 2\vec{x}, \qquad \nabla^2 f(\vec{x}) = 2I. \tag{3}$$

The Hessian is positive semidefinite at each point $\vec{x}$ so f is convex.

(b) (6 pts) **Is the function $g\colon \mathbb{R}^n \to \mathbb{R}$ given by $g(\vec{x}) \doteq e^{\|\vec{x}\|_2^2}$ convex? If g is convex, prove it; if g is not convex, give an example $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $\theta \in [0,1]$ such that $g(\theta\vec{x} + (1-\theta)\vec{y}) > \theta g(\vec{x}) + (1-\theta)g(\vec{y})$.**

NOTE: One (short) solution to this problem does not use gradients or Hessians, but it is fine if yours does. In particular, the gradient and Hessian of g were derived in homework; if you want to use these quantities, please derive them here. You may use without proof the gradient and Hessian of $f(\vec{x}) \doteq \|\vec{x}\|_2^2$.

Solution: We give two solutions, one using properties of convex functions, and one which calculates the Hessian and shows it is PSD (this is more "brute-force").

Solution 1. Since the function $x \mapsto e^x$ is monotonically increasing and convex, and $\vec{x} \mapsto \|\vec{x}\|_2^2$ is convex by part (a), g is a composition of a monotonically increasing and convex function with a convex function, so it is convex.

Solution 2. We know that

$$\begin{aligned}\nabla g(\vec{x}) &= [D\exp\left(\|\vec{x}\|_2^2\right)][D\,\|\vec{x}\|_2^2] && (4)\\ &= \exp\left(\|\vec{x}\|_2^2\right)\cdot 2\vec{x} && (5)\\ &= 2e^{\|\vec{x}\|_2^2}\vec{x}. && (6)\end{aligned}$$

Therefore

$$\begin{aligned}\nabla^2 g(\vec{x}) &= D(\nabla g)(\vec{x}) && (7)\\ &= D(2e^{\|\vec{x}\|_2^2}\vec{x}). && (8)\end{aligned}$$

The components of this Jacobian are

$$\begin{aligned}[\nabla^2 g(\vec{x})]_{jk} &= \frac{\partial}{\partial x_k}(2e^{\|\vec{x}\|_2^2}\vec{x})_j && (9)\\ &= \frac{\partial}{\partial x_k} 2e^{\|\vec{x}\|_2^2} x_j && (10)\\ &= 2e^{\|\vec{x}\|_2^2}\frac{\partial x_j}{\partial x_k} + 2x_j\frac{\partial}{\partial x_k}e^{\|\vec{x}\|_2^2} && (11)\\ &= 2e^{\|\vec{x}\|_2^2}\frac{\partial x_j}{\partial x_k} + 2x_j[\nabla f(\vec{x})]_k && (12)\\ &= 2e^{\|\vec{x}\|_2^2}\frac{\partial x_j}{\partial x_k} + 4x_jx_ke^{\|\vec{x}\|_2^2} && (13)\end{aligned}$$

This matrix forms

$$\nabla^2 g(\vec{x}) = 2e^{\|\vec{x}\|_2^2}[I + 2\vec{x}\vec{x}^\top]. \tag{14}$$

We can show that this is PSD: take any $\vec{v} \in \mathbb{R}^n$, then

$$\vec{v}^\top[\nabla^2 g(\vec{x})]\vec{v} = 2e^{\|\vec{x}\|_2^2}\vec{v}^\top(I + 2\vec{x}\vec{x}^\top)\vec{v} = 2e^{\|\vec{x}\|_2^2}(\vec{v}^\top\vec{v} + 2(\vec{x}^\top\vec{v})^2) = 2e^{\|\vec{x}\|_2^2}(\|\vec{v}\|_2^2 + 2(\vec{x}^\top\vec{v})^2) \geq 0, \tag{15}$$

where the last inequality is because every single term in the expression is non-negative, so their product and sum must also be non-negative. This proves that g is convex.

6. Spectrahedron (7 pts)

Let $F_1, \dots, F_n \in \mathbb{R}^{m \times m}$ be symmetric matrices. Define the set $S \subseteq \mathbb{R}^n$, known as a *spectrahedron*, by

$$S \doteq \left\{ \vec{x} \in \mathbb{R}^n \,\middle|\, \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \sum_{i=1}^{n} x_i F_i \succeq 0 \right\}. \tag{16}$$

Here $A \succeq 0$ means that A is symmetric PSD. **Show that S is a convex set.**

HINT: You can use without proof that convex combinations of symmetric PSD matrices are symmetric PSD.

Solution: Let $\vec{x}, \vec{y} \in S$, let $\theta \in [0, 1]$, and let $\vec{z} = \theta\vec{x} + (1 - \theta)\vec{y}$. Then

$$\begin{aligned} \sum_{i=1}^{n} z_i F_i &= \sum_{i=1}^{n} (\theta x_i + (1-\theta) y_i) F_i && (17) \\ &= \theta \underbrace{\sum_{i=1}^{n} x_i F_i}_{\succeq 0} + (1-\theta) \underbrace{\sum_{i=1}^{n} y_i F_i}_{\succeq 0} && (18) \\ &\succeq 0 && (19) \end{aligned}$$

since the set of positive semidefinite matrices is closed under non-negative scalar multiples (like θ and $1 - \theta$) and addition.

7. Gradient Descent on Quadratic (6 pts)

Let $a, \eta \in \mathbb{R}$ be such that $\eta > 0$ and $0 < a < 1/\eta$. Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) \doteq \frac{1}{2}ax^2, \qquad \text{for all } x \in \mathbb{R}. \tag{20}$$

We run gradient descent on f with constant step size η and fixed initialization $x_0 = 1$ to get iterates $(x_t)_{t=0}^{\infty}$, i.e.,

$$x_{t+1} \doteq x_t - \eta \frac{\mathrm{d}f}{\mathrm{d}x}(x_t) \qquad \text{for all } t \geq 0, \qquad \text{and} \quad x_0 = 1. \tag{21}$$

Complete the following tasks:

- **compute the derivative of f (denoted $\frac{\mathrm{d}f}{\mathrm{d}x}$ or f');**
- **write the update rule for x_{t+1} in terms of x_t, a, and η;**
- **write an expression for x_t in terms of x_0, a, η, and t;**
- **and compute the limit $\lim_{t\to\infty} x_t$.**

Show your work and justify your answer(s).

Solution: Notice that

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x) = ax, \tag{22}$$

so that

$$x - \eta \nabla f(x) = (1 - \eta a)x. \tag{23}$$

Thus we have

$$x_{t+1} = (1 - \eta a)x_t, \tag{24}$$

and so

$$x_t = (1 - \eta a)^t x_0. \tag{25}$$

Since $x_0 = 1$, we have that $x_t = (1 - \eta a)^t$ for all $t \geq 0$. Since $a > 0$, we have $1 - \eta a < 1$, so that $x_t = (1 - \eta a)^t \to 0$.

8. Vector Calculus (14 pts)

(a) (8 pts) Let $\vec{a} \in \mathbb{R}^n$ be a fixed vector, and $b \in \mathbb{R}$ be a fixed scalar. **Compute the gradient and Hessian of the function** $f : \mathbb{R}^n \to \mathbb{R}$ **given by**

$$f(\vec{x}) \doteq \sin(\vec{a}^\top \vec{x} - b). \tag{26}$$

Show your work and justify your answer(s).

Solution: We use the chain rule. Write $f(\vec{x}) = g(\ell(\vec{x}))$ where $g(x) = \sin(x)$ and $\ell(\vec{x}) = \vec{a}^\top \vec{x} - b$. Then to compute the gradient, we have

$$\begin{aligned}
\nabla f(\vec{x}) &= [Df(\vec{x})]^\top && (27)\\
&= [D(g \circ \ell)(\vec{x})]^\top && (28)\\
&= [(Dg(\ell(\vec{x})))(D\ell(\vec{x}))]^\top && (29)\\
&= [\cos(\ell(\vec{x}))\vec{a}^\top]^\top && (30)\\
&= \cos(\vec{a}^\top \vec{x} - b) \cdot \vec{a}. && (31)
\end{aligned}$$

To compute the Hessian, we have

$$\begin{aligned}
\nabla^2 f(\vec{x}) &= D(\nabla f)(\vec{x}) && (32)\\
&= D(\cos(\vec{a}^\top \vec{x} - b) \cdot \vec{a}). && (33)
\end{aligned}$$

At this point we do component-wise derivatives:

$$\begin{aligned}
[\nabla^2 f(\vec{x})]_{ij} &= [D(\cos(\vec{a}^\top \vec{x} - b) \cdot \vec{a})]_{ij} && (34)\\
&= \frac{\partial(\cos(\vec{a}^\top \vec{x} - b) \cdot \vec{a})_i}{\partial x_j} && (35)\\
&= [D(\cos(\vec{a}^\top \vec{x} - b) \cdot \vec{a})]_{ij} && (36)\\
&= \frac{\partial(\cos(\vec{a}^\top \vec{x} - b) \cdot a_i)}{\partial x_j} && (37)\\
&= \frac{\partial(\cos(\vec{a}^\top \vec{x} - b))}{\partial x_j} \cdot a_i && (38)\\
&= \frac{\partial(\cos(\vec{a}^\top \vec{x} - b))}{\partial(\vec{a}^\top \vec{x} - b)} \cdot \frac{\partial(\vec{a}^\top \vec{x} - b)}{\partial x_j} \cdot a_i && (39)\\
&= -\sin(\vec{a}^\top \vec{x} - b) \cdot a_i \cdot a_j. && (40)
\end{aligned}$$

This gives

$$\nabla^2 f(\vec{x}) = -\sin(\vec{a}^\top \vec{x} - b)\vec{a}\vec{a}^\top. \tag{41}$$

(b) (6 pts) Let $\vec{u} \in \mathbb{R}^n$ be a fixed vector. **Compute the Jacobian of the function** $\vec{f} : \mathbb{R}^n \to \mathbb{R}^n$ **given by**

$$\vec{f}(\vec{x}) \doteq (\vec{u}^\top \vec{x})\vec{u}. \tag{42}$$

Show your work and justify your answer(s).

HINT: One (short) solution to this problem starts by rewriting $\vec{f}(\vec{x})$ as a matrix-vector product, but you can do this problem any way you want.

Solution: Note that we can write the projection as

$$\vec{f}(\vec{x}) = \vec{u}\vec{u}^\top \vec{x} \tag{43}$$

which is just a constant matrix times the input vector $\vec{x}$, so its Jacobian is just the matrix

$$D\vec{f}(\vec{x}) = \vec{u}\vec{u}^\top. \tag{44}$$

9. Factorizations of PSD Matrices (16 pts)

Let k, n be positive integers, with $k \leq n$. In this problem, we prove that $A \in \mathbb{R}^{n \times n}$ is a symmetric PSD matrix of rank k if and only if it can be written as $A = PP^\top$ for some matrix $P \in \mathbb{R}^{n \times k}$ which has full column rank.

(a) (8 pts) Let $A \in \mathbb{R}^{n \times n}$ be a symmetric PSD matrix with rank k. **Prove that there exists another matrix $P \in \mathbb{R}^{n \times k}$ with full column rank, i.e., $\operatorname{rank}(P) = k$, such that $A = PP^\top$.**

HINT: Recall that A is a square and symmetric $n \times n$ matrix, while P is a tall $n \times k$ matrix.

Solution: Let $A = \sum_{i=1}^{k} \lambda_i \vec{v}_i \vec{v}_i^\top$, and let $P = \begin{bmatrix} \sqrt{\lambda_1}\vec{v}_1 & \ldots & \sqrt{\lambda_k}\vec{v}_k \end{bmatrix}$. Then that $PP^\top = \sum_{i=1}^{k} \lambda_i \vec{v}_i \vec{v}_i^\top = A$, and P has full column rank because it has orthogonal columns.

(b) (8 pts) Let $P \in \mathbb{R}^{n \times k}$ be a matrix with full column rank, i.e., $\operatorname{rank}(P) = k$. **Prove that if we define $A \doteq PP^\top$, then $A \in \mathbb{R}^{n \times n}$ is a symmetric PSD matrix of rank k.**

HINT: We know two ways to show that $\operatorname{rank}(A) = k$. One uses the rank-nullity theorem and that $\mathcal{N}(B^\top B) = \mathcal{N}(B)$ for any matrix B in order to compute the rank of $A = PP^\top$. The other uses the SVD of P.

Solution: Indeed A is symmetric because

$$A^\top = (PP^\top)^\top = (P^\top)^\top (P^\top) = PP^\top. \tag{45}$$

To show that A is positive semidefinite, for each $\vec{x} \in \mathbb{R}^n$ we have

$$\vec{x}^\top A \vec{x} = \vec{x}^\top PP^\top \vec{x} = \left\| P^\top \vec{x} \right\|_2^2 \geq 0. \tag{46}$$

We now show that $\operatorname{rank}(A) = k$. Indeed as a sum of k dyads it is easy to show that $\operatorname{rank}(A) \leq k$, but this does not prove that the quantities are equal, deserving partial credit. We give two proofs here.

Proof 1.

A simple proof is by dimension-counting and the rank-nullity theorem:

$$\begin{aligned}
\operatorname{rank}(PP^\top) &= n - \dim(\mathcal{N}(PP^\top)) && (47)\\
&= n - \dim(\mathcal{N}(P^\top)) && (48)\\
&= n - (n - \dim(\mathcal{R}(P^\top))) && (49)\\
&= n - (n - \operatorname{rank}(P^\top)) && (50)\\
&= n - (n - \operatorname{rank}(P)) && (51)\\
&= n - (n - k) && (52)\\
&= k. && (53)
\end{aligned}$$

Here we know that $\mathcal{N}(PP^\top) = \mathcal{N}(P^\top)$ by the more general statement that for any matrix B we have $\mathcal{N}(B) = \mathcal{N}(B^\top B)$ (and take $B = P^\top$).

Proof 2.

A more quantitative proof uses the (compact) SVD, writing $P = U_k \Sigma_k V_k^\top$, where $U_k \in \mathbb{R}^{n \times k}$ and $V_k \in \mathbb{R}^{k \times k}$ have orthonormal columns and $\Sigma_k \in \mathbb{R}^{k \times k}$ is diagonal with positive entries. (We may do this precisely because P has full column rank, meaning that it has rank k). Then

$$PP^\top = (U_k \Sigma_k V_k^\top)(U_k \Sigma_k V_k^\top)^\top = U_k \Sigma_k V_k^\top V_k \Sigma_k^\top U_k^\top = U_k \Sigma_k \Sigma_k^\top U_k^\top = U_k \Sigma_k^2 U_k^\top \tag{54}$$

which is a rank-k matrix since Σ_k^2 has k nonzero entries on its diagonal.

10. ℓ^p Norms (8 pts)

Let n be a positive integer. Recall that for $1 \leq p \leq \infty$ the ℓ^p norm on $\mathbb{R}^n$ is defined as

$$\|\vec{x}\|_p \doteq \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} & \text{if } 1 \leq p < \infty \\ \max_{i \in \{1,\ldots,n\}} |x_i| & \text{if } p = \infty, \end{cases} \qquad \text{for all } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n. \tag{55}$$

Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Let $\vec{r}_i \in \mathbb{R}^n$ be the i^{th} row of A, i.e.,

$$A = \begin{bmatrix} \vec{r}_1^\top \\ \vdots \\ \vec{r}_m^\top \end{bmatrix}. \tag{56}$$

Prove the identity

$$\max_{\substack{\vec{v} \in \mathbb{R}^n \\ \|\vec{v}\|_2 = 1}} \|A\vec{v}\|_\infty = \max_{i \in \{1,\ldots,m\}} \|\vec{r}_i\|_2. \tag{57}$$

HINT: The Cauchy-Schwarz inequality may be useful. Think about when equality holds.

Solution: For any $\vec{v} \in \mathbb{R}^n$ we have

$$\begin{aligned} \|A\vec{v}\|_\infty &= \max_{i \in \{1,\ldots,m\}} |(A\vec{v})_i| && (58) \\ &= \max_{i \in \{1,\ldots,m\}} \left|\vec{r}_i^\top \vec{v}\right| && (59) \\ &\leq \max_{i \in \{1,\ldots,m\}} \|\vec{v}\|_2 \|\vec{r}_i\|_2 && (60) \\ &= \|\vec{v}\|_2 \max_{i \in \{1,\ldots,m\}} \|\vec{r}_i\|_2, && (61) \end{aligned}$$

where the inequality step is by Cauchy-Schwarz. Therefore

$$\begin{aligned} \|A\|_{2,\infty} &= \max_{\substack{\vec{v} \in \mathbb{R}^n \\ \|\vec{v}\|_2 = 1}} \|A\vec{v}\|_\infty && (62) \\ &\leq \max_{\substack{\vec{v} \in \mathbb{R}^n \\ \|\vec{v}\|_2 = 1}} \|\vec{v}\|_2 \max_{i \in \{1,\ldots,m\}} \|\vec{r}_i\|_2 && (63) \\ &= \max_{i \in \{1,\ldots,m\}} \|\vec{r}_i\|_2. && (64) \end{aligned}$$

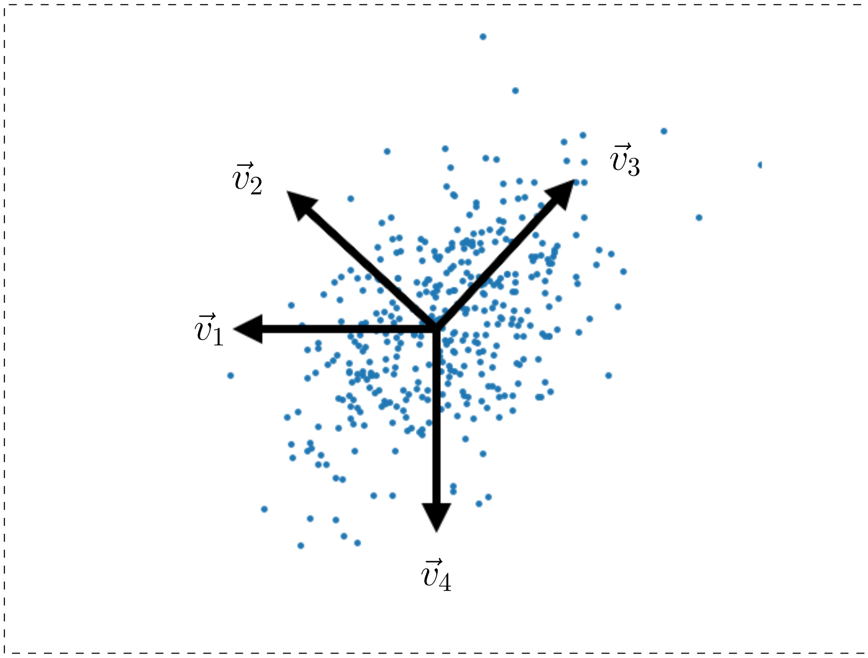
This upper bound is always achievable. Let $i^\star \in \operatorname{argmax}_{i \in \{1,\ldots,m\}} \|\vec{r}_i\|_2$. Then, a choice of

$$\vec{v} = \frac{\vec{r}_{i^\star}}{\|\vec{r}_{i^\star}\|_2} \tag{65}$$

achieves the upper bound. We derive this by noting that we just need to make the above invocation of Cauchy-Schwarz tight, which occurs when $\vec{v}$ is parallel to $\vec{r}_{i^\star}$.

11. PCA and Regression (34 pts)

(a) (4 pts) Given the following plot of data in $\mathbb{R}^2$ (i.e., each dot is a data point in $\mathbb{R}^2$) and candidate unit vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4 \in \mathbb{R}^2$, **identify the candidate vectors which could be the first principal component and second principal component of the data (and specify which is which).** *You do not need to show your work for this subpart.*



Solution: The vector $\vec{v}_3$ is the first principal component, since it aligns the most with the largest degree of variation in the data; alternatively, projecting onto it gives the minimum sum of squared errors, across all unit vectors. Then $\vec{v}_2$ is the second principal component since it is the only vector orthogonal to $\vec{v}_3$.

(b) (6 pts) Suppose we have pairs of data $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$, where $n > d$. As usual, we arrange these data points into a matrix and vector, i.e.,

$$X = \begin{bmatrix} \vec{x}_1^\top \\ \vdots \\ \vec{x}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times d}, \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n. \tag{66}$$

Assume that X is centered, i.e., each column has mean zero: $(1/n)\sum_{i=1}^n \vec{x}_i = \vec{0}_d$, where $\vec{0}_d$ is the zero vector in $\mathbb{R}^d$. Suppose that X has compact SVD given by $X = U_d \Sigma_d V_d^\top$ where

$$U_d = \begin{bmatrix} \vec{u}_1, \ldots, \vec{u}_d \end{bmatrix} \in \mathbb{R}^{n \times d}, \qquad V_d = \begin{bmatrix} \vec{v}_1, \ldots, \vec{v}_d \end{bmatrix} \in \mathbb{R}^{d \times d}, \qquad \Sigma_d = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix} \in \mathbb{R}^{d \times d} \tag{67}$$

where $\sigma_1 > \sigma_2 > \cdots > \sigma_d > 0$. **From this SVD, identify the top k principal components of the data $\{\vec{x}_1, \ldots, \vec{x}_n\} \subseteq \mathbb{R}^d$, where $k \leq d$.** *You do not need to show your work for this subpart.*

HINT: Recall that the first principal component solves the optimization problem $\underset{\vec{w} \in \mathbb{R}^d \colon \|\vec{w}\|_2 = 1}{\operatorname{argmax}} \vec{w}^\top X^\top X \vec{w}$.

Solution: Recall that the sample covariance is

$$\frac{1}{n}\sum_{i=1}^n \vec{x}_i \vec{x}_i^\top = \frac{1}{n} X^\top X. \tag{68}$$

The top k principal components are the top k eigenvectors of the sample covariance, given by

$$\frac{1}{n}X^\top X = V_r\left(\frac{\Sigma_r^2}{n}\right)V_r^\top. \tag{69}$$

These principal components are then the first k columns of V_r, namely $\vec{v}_1, \ldots, \vec{v}_k$.

(c) (4 pts) Suppose that $P = \begin{bmatrix}\vec{p}_1, \ldots, \vec{p}_k\end{bmatrix} \in \mathbb{R}^{d\times k}$ is a matrix with columns $\vec{p}_j$. Let $Z = XP$, and let the entries of Z be z_{ij}, i.e.,

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk}\end{bmatrix} \in \mathbb{R}^{n\times k}. \tag{70}$$

Give an expression for z_{ij} in terms of $\vec{x}_i$ and $\vec{p}_j$. *You do not need to show your work for this subpart.*

Solution: We have

$$\begin{aligned} & Z = XP && (71) \\ \implies & Z^\top = P^\top X^\top && (72) \\ \implies & \begin{bmatrix}\vec{z}_1 & \cdots & \vec{z}_n\end{bmatrix} = P^\top \begin{bmatrix}\vec{x}_1 & \cdots & \vec{x}_n\end{bmatrix} = \begin{bmatrix}P^\top\vec{x}_1 & \cdots & P^\top\vec{x}_n\end{bmatrix}. && (73)\end{aligned}$$

Then we have

$$\vec{z}_i = P^\top \vec{x}_i = \begin{bmatrix}\vec{p}_1^\top \\ \vdots \\ \vec{p}_k^\top\end{bmatrix}\vec{x}_i = \begin{bmatrix}\vec{p}_1^\top\vec{x}_i \\ \vdots \\ \vec{p}_k^\top\vec{x}_i\end{bmatrix}. \tag{74}$$

(d) (10 pts) Define the matrices U_k, V_k, and Σ_k as

$$U_k = \begin{bmatrix}\vec{u}_1, \ldots, \vec{u}_k\end{bmatrix} \in \mathbb{R}^{n\times k}, \qquad V_k = \begin{bmatrix}\vec{v}_1, \ldots, \vec{v}_k\end{bmatrix} \in \mathbb{R}^{d\times k}, \qquad \Sigma_k = \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_k\end{bmatrix} \in \mathbb{R}^{k\times k}. \tag{75}$$

Suppose that $P = V_k$, so that $Z = XV_k$. Let $\lambda \geq 0$, and let $\vec{\beta}^\star \in \mathbb{R}^k$ solve the ridge regression problem

$$\vec{\beta}^\star \doteq \operatorname*{argmin}_{\vec{\beta}\in\mathbb{R}^k}\left[\|Z\vec{\beta} - \vec{y}\|_2^2 + \lambda\|\vec{\beta}\|_2^2\right]. \tag{76}$$

Show that:

$$\vec{\beta}^\star = (\Sigma_k^2 + \lambda I_k)^{-1}\Sigma_k U_k^\top \vec{y}, \tag{77}$$

where $I_k \in \mathbb{R}^{k\times k}$ is the $k \times k$ identity matrix.

Solution: We give two approaches, one which saves a lot of work by evaluating terms in an optimal order, and another which is more brute-force.

Approach 1.

To compute $\vec{\beta}^\star$, we have

$$\begin{aligned} Z &= XP && (78) \\ &= XV_k && (79) \\ &= U_d\Sigma_d V_d^\top V_k && (80)\end{aligned}$$

$$= U_d\Sigma_d \begin{bmatrix} V_k^\top \\ V_{d-k}^\top \end{bmatrix} V_k \tag{81}$$

$$= U_d\Sigma_d \begin{bmatrix} V_k^\top V_k \\ V_{d-k}^\top V_k \end{bmatrix} \tag{82}$$

$$= U_d\Sigma_d \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} \tag{83}$$

$$= \begin{bmatrix} U_k & U_{d-k} \end{bmatrix} \begin{bmatrix} \Sigma_k & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & \Sigma_{d-k} \end{bmatrix} \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} \tag{84}$$

$$= U_k\Sigma_k, \tag{85}$$

where $U_{d-k} = \begin{bmatrix}\vec{u}_{k+1}, \dots, \vec{u}_d\end{bmatrix}$, $V_{d-k} = \begin{bmatrix}\vec{v}_{k+1}, \dots, \vec{v}_d\end{bmatrix}$, and $\Sigma_{d-k} = \begin{bmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix}$. Then

$$\vec{\beta}^\star = (Z^\top Z + \lambda I_k)^{-1} Z^\top \vec{y} \tag{86}$$

$$= ((U_k\Sigma_k)^\top (U_k\Sigma_k) + \lambda I)^{-1} (U_k\Sigma_k)^\top \vec{y} \tag{87}$$

$$= (\Sigma_k^\top \Sigma_k + \lambda I_k)^{-1} \Sigma_k^\top U_k^\top \vec{y} \tag{88}$$

$$= (\Sigma_k^2 + \lambda I_k)^{-1} \Sigma_k U_k^\top \vec{y}. \tag{89}$$

Approach 2.

We start by computing

$$\vec{\beta}^\star = (Z^\top Z + \lambda I)^{-1} Z^\top \vec{y} \tag{90}$$

$$= ((XV_k)^\top (XV_k) + \lambda I_k)^{-1} (XV_k)^\top \vec{y} \tag{91}$$

$$= (V_k^\top X^\top X V_k + \lambda I_k)^{-1} V_k^\top X^\top \vec{y}. \tag{92}$$

Now we plug in $X = U_d\Sigma_d V_d^\top$, and obtain

$$\vec{\beta}^\star = (V_k^\top (U_d\Sigma_d V_d^\top)^\top (U_d\Sigma_d V_d^\top) V_k + \lambda I_k)^{-1} V_k^\top (U_d\Sigma_d V_d^\top)^\top \vec{y} \tag{93}$$

$$= (V_k^\top V_d \Sigma_d U_d^\top U_d \Sigma_d V_d^\top V_k + \lambda I_k)^{-1} V_k^\top V_d \Sigma_d U_d^\top \vec{y} \tag{94}$$

$$= (V_k^\top V_d \Sigma_d^2 V_d^\top V_k + \lambda I_k)^{-1} V_k^\top V_d \Sigma_d U_d^\top \vec{y}. \tag{95}$$

As in the previous solution, we write

$$V_d^\top V_k = \begin{bmatrix} V_k^\top \\ V_{d-k}^\top \end{bmatrix} V_k = \begin{bmatrix} V_k^\top V_k \\ V_{d-k}^\top V_k \end{bmatrix} = \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix}. \tag{96}$$

This obtains

$$\vec{\beta}^\star = \left(\begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \Sigma_d^2 \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} + \lambda I_k \right)^{-1} \begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \Sigma_d U_d^\top \vec{y} \tag{97}$$

$$= \left(\begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \begin{bmatrix} \Sigma_k & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & \Sigma_{d-k} \end{bmatrix}^2 \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} + \lambda I_k \right)^{-1} \begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \Sigma_d U_d^\top \vec{y} \tag{98}$$

$$= \left(\begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \begin{bmatrix} \Sigma_k^2 & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & \Sigma_{d-k}^2 \end{bmatrix} \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} + \lambda I_k \right)^{-1} \begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \Sigma_d U_d^\top \vec{y} \tag{99}$$

$$= \left(\Sigma_k^2 + \lambda I_k\right)^{-1} \begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \Sigma_d U_d^\top \vec{y} \tag{100}$$

$$= \left(\Sigma_k^2 + \lambda I_k\right)^{-1} \begin{bmatrix} I_k & 0_{k\times(d-k)} \end{bmatrix} \begin{bmatrix} \Sigma_k & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & \Sigma_{d-k} \end{bmatrix} \begin{bmatrix} U_k & U_{d-k} \end{bmatrix}^\top \vec{y} \tag{101}$$

$$= \left(\Sigma_k^2 + \lambda I_k\right)^{-1} \Sigma_k U_k^\top \vec{y}. \tag{102}$$

Note: The quantities $(V_k^\top M V_k)^{-1}$ and $V_k^\top M^{-1} V_k$ are not equal in general, because V_k is not square and not invertible, and so we cannot use the false identity $(V_k^\top M V_k)^{-1} = V_k^\top M^{-1} V_k$ expression to simplify any calculations.

(e) (10 pts) Let $\vec{\alpha}^\star \in \mathbb{R}^d$ solve the original ridge regression problem, i.e.,

$$\vec{\alpha}^\star \doteq \operatorname*{argmin}_{\vec{\alpha}\in\mathbb{R}^d} \left[\|X\vec{\alpha} - \vec{y}\|_2^2 + \lambda\|\vec{\alpha}\|_2^2\right] = V_d(\Sigma_d^2 + \lambda I_d)^{-1}\Sigma_d U_d^\top \vec{y}, \tag{103}$$

where $I_d \in \mathbb{R}^{d\times d}$ is the $d \times d$ identity matrix. (You can assume without proof that the above equality is true.)

Compute

$$\|X\vec{\alpha}^\star - Z\vec{\beta}^\star\|_2^2, \tag{104}$$

in terms of the vectors $(\vec{u}_i)_{i=1}^d$ and $\vec{y}$, and the scalars $(\sigma_i)_{i=1}^d$ and λ. *Show your work and justify your answer(s).*

Solution: We use the compact SVD, though solutions can use any SVD (as long as the dimensions are kept correct). We have

$$X\vec{\alpha}^\star = (U_d\Sigma_d V_d^\top)(V_d(\Sigma_d^2 + \lambda I)^{-1}\Sigma_d U_d^\top \vec{y}) \tag{105}$$

$$= U_d\Sigma_d V_d^\top V_d(\Sigma_d^2 + \lambda I)^{-1}\Sigma_d U_d^\top \vec{y} \tag{106}$$

$$= U_d\Sigma_d(\Sigma_d^2 + \lambda I)^{-1}\Sigma_d U_d^\top \vec{y} \tag{107}$$

$$= U_d \begin{bmatrix} \frac{\sigma_1^2}{\sigma_1^2+\lambda} & & \\ & \ddots & \\ & & \frac{\sigma_d^2}{\sigma_d^2+\lambda} \end{bmatrix} U_d^\top \vec{y} \tag{108}$$

$$= \sum_{i=1}^{d} \frac{\sigma_i^2}{\sigma_i^2+\lambda} \vec{u}_i\vec{u}_i^\top \vec{y} \tag{109}$$

$$= \sum_{i=1}^{d} \frac{\sigma_i^2}{\sigma_i^2+\lambda} (\vec{u}_i^\top \vec{y})\vec{u}_i \tag{110}$$

$$Z\vec{\beta}^\star = (U_k\Sigma_k)((\Sigma_k^2 + \lambda I)^{-1}\Sigma_k U_k^\top \vec{y}) \tag{111}$$

$$= U_k\Sigma_k(\Sigma_k^2 + \lambda I)^{-1}\Sigma_k U_k^\top \vec{y} \tag{112}$$

$$= U_k \begin{bmatrix} \frac{\sigma_1^2}{\sigma_1^2+\lambda} & & \\ & \ddots & \\ & & \frac{\sigma_k^2}{\sigma_k^2+\lambda} \end{bmatrix} U_k^\top \vec{y} \tag{113}$$

$$= \sum_{i=1}^{k} \frac{\sigma_i^2}{\sigma_i^2+\lambda} \vec{u}_i\vec{u}_i^\top \vec{y} \tag{114}$$

$$= \sum_{i=1}^{k} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} (\vec{u}_i^\top \vec{y}) \vec{u}_i. \tag{115}$$

Thus

$$X\vec{\alpha}^\star - Z\vec{\beta}^\star = \sum_{i=k+1}^{d} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} (\vec{u}_i^\top \vec{y}) \vec{u}_i \tag{116}$$

$$\|X\vec{\alpha}^\star - Z\vec{\beta}^\star\|_2^2 = \sum_{i=k+1}^{d} \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda}\right)^2 (\vec{u}_i^\top \vec{y})^2. \tag{117}$$

Bonus 1: An alternative, equally valid way to compute $Z\vec{\beta}^\star$ is as follows:

$$Z\vec{\beta}^\star = XV_k\vec{\beta}^\star \tag{118}$$

$$= (U_d\Sigma_d V_d^\top)V_k(\Sigma_k^2 + \lambda I_k)^{-1}\Sigma_k U_k^\top \vec{y} \tag{119}$$

$$= U_d\Sigma_d(V_d^\top V_k)(\Sigma_k^2 + \lambda I_k)^{-1}\Sigma_k U_k^\top \vec{y} \tag{120}$$

$$= \begin{bmatrix} U_k & U_{d-k} \end{bmatrix} \begin{bmatrix} \Sigma_k & 0_{k\times(d-k)} \\ 0_{(d-k)\times k} & \Sigma_{d-k} \end{bmatrix} \begin{bmatrix} I_k \\ 0_{(d-k)\times k} \end{bmatrix} (\Sigma_k^2 + \lambda I_k)^{-1}\Sigma_k U_k^\top \vec{y} \tag{121}$$

$$= U_k\Sigma_k(\Sigma_k^2 + \lambda I_k)^{-1}\Sigma_k U_k^\top \vec{y}. \tag{122}$$

Bonus 2: to compute $\vec{\alpha}^\star$ in the first place, we used a similar calculation to part (d), obtaining

$$\vec{\alpha}^\star = (X^\top X + \lambda I)^{-1} X^\top \vec{y} \tag{123}$$

$$= ((U_d\Sigma_d V_d^\top)^\top (U_d\Sigma_d V_d^\top) + \lambda I)^{-1} (U_d\Sigma_d V_d^\top)^\top \vec{y} \tag{124}$$

$$= (V_d\Sigma_d^\top U_d^\top U_d\Sigma_d V_d^\top + \lambda I)^{-1} V_d\Sigma_d^\top U_d^\top \vec{y} \tag{125}$$

$$= (V_d\Sigma_d^2 V_d^\top + \lambda I)^{-1} V_d\Sigma_d U_d^\top \tag{126}$$

$$= (V_d(\Sigma_d^2 + \lambda I)V_d^\top)^{-1} V_d\Sigma_d U_d^\top \tag{127}$$

$$= V_d(\Sigma_d^2 + \lambda I)^{-1} V_d^\top V_d\Sigma_d U_d^\top \tag{128}$$

$$= V_d(\Sigma_d^2 + \lambda I)^{-1}\Sigma_d U_d^\top \vec{y}. \tag{129}$$

