EECS 127/227AT  Optimization Models in Engineering
Spring 2020  Midterm

This exam has a total of 114 points. However, a score of 100 on 114 will be considered a perfect score, so 14 points on the exam are bonus.

1. Convexity (12 points)
State whether the following functions/sets are convex and justify your answer. Answers without justification will receive no credit.

(a) (4 points) Function \( f(\vec{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \).

(b) (4 points) Set \( S = \{ (\vec{x}, y) \mid \|A\vec{x} - \vec{b}\|_2^2 \leq y \} \). Hint: Consider the epigraph of a function. Other proofs may also work.

(c) (4 points) Function \( f(\vec{x}) = \max_{\vec{b}} \left[ \vec{b}^\top A\vec{b} + \vec{x}^\top \vec{b} \right] \), where \( A \) is a fixed arbitrary matrix. Hint: Note that the maximization is over \( \vec{b} \).

2. Gradient descent (10 points)
Consider the function \( f : \mathbb{R}^n \to \mathbb{R} \), where
\[
f(\vec{x}) = \frac{1}{4} \|\vec{x}\|_2^4.
\]
Let \( \vec{x}^* = \arg \min_{\vec{x}} f(\vec{x}) \).
Recall that the gradient descent update equation for minimizing \( f \) is given by
\[
\vec{x}_{t+1} = \vec{x}_t - \eta \nabla f(\vec{x}_t),
\]
where \( \eta > 0 \) is the step size.

(a) (2 points) Find \( \vec{x}^* \). You need not show any work for this subpart.

(b) (8 points) Suppose \( \|\vec{x}_0\|_2 = c \neq 0 \). Find the range of \( \eta \) (in terms of \( c \)) such that gradient descent converges to \( \vec{x}^* \). Justify your answer.
Hint: If you are having trouble solving this part for general dimension \( n \), solve it for \( n = 1 \) for partial credit.
3. PCA (12 points)

In this problem, we will find the principal components of data points on a regularly spaced grid. Consider a set \( S \) of \( n = 15 \) data points that lie at each integer node of a \( 5 \times 3 \) grid:

\[
S = \left\{ \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \mid x_1 \in \{-2, -1, 0, 1, 2\}, \ x_2 \in \{-1, 0, 1\} \right\}.
\]

A plot of these points is shown in Fig. ??.

![Figure 1: Point data.](image)

Note that the empirical covariance matrix of these data points is given by

\[
C = \begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{3} \end{bmatrix}.
\]

(a) (6 points) Recall that for data with empirical covariance matrix \( C \), the variance \( \sigma^2(\vec{u}) \) along any unit vector \( \vec{u} \) is given by

\[
\sigma^2(\vec{u}) = \vec{u}^\top C \vec{u}.
\]

The data’s first principal component \( \vec{u}_1 \) is the unit vector direction that maximizes variance, i.e.,

\[
\vec{u}_1 = \operatorname{argmax}_{\|\vec{u}\|_2 = 1} \sigma^2(\vec{u}).
\]

**Compute both** \( \vec{u}_1 \) **and** \( \sigma^2(\vec{u}_1) \). **Show your work.**

(b) (6 points) Let \( \vec{x}_i \) for \( i = 1, \cdots, 15 \) represent the elements of set \( S \). Suppose we transform every point \( \vec{x} \in S \) by multiplying by an arbitrary orthonormal matrix \( W \) to generate new data points \( \vec{z}_i = W \vec{x}_i \), where \( i = 1, \ldots, 15 \) indexes over every element of \( S \). Let \( \vec{v}_1 \) denote the first principal component of the transformed data and let \( \vec{v}_2 \) denote its second principal component. **Find** \( \vec{v}_1 \) **and** \( \vec{v}_2 \) **in terms of** \( \vec{u}_1 \), \( \vec{u}_2 \), **and** \( W \).

**Hint:** It may be useful to find the new empirical covariance of this transformed data in terms of \( C \) and \( W \).

---

1You may find this scenario contrived, but it’s actually based on a real research problem encountered by one of your GSIs when analyzing point cloud data from a robot’s sensor. To figure out where the robot should place its gripper along a beam to pick it up, they used PCA!
4. All I need is Q (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

\[
A = Q_1 R_1 = \begin{bmatrix}
1 & q_{12} \\
0 & q_{22} \\
0 & q_{23} \\
\end{bmatrix} \begin{bmatrix}
? & r_{12} \\
0 & r_{22} \\
\end{bmatrix}.
\]

(1)

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

\[
A = QR = [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix},
\]

(2)

where \( Q_1 \) and \( R_1 \) are as defined in Equation (1). Consider the least squares problem

\[
p^* = \min_\tilde{x} \left\| A\tilde{x} - \tilde{b} \right\|_2^2
\]

(3)

for \( A \) given in Equation (1) and some \( \tilde{b} \in \mathbb{R}^{3} \). Consider the following two possible ways of rewriting this least squares problem in terms of \( Q_1, Q_2, \) and \( R_1 \):

**Strategy 1:**

\[
\left\| \tilde{b} - A\tilde{x} \right\|_2^2 \overset{(I)}{=} \left\| Q^\top \tilde{b} - Q^\top A\tilde{x} \right\|_2^2 = \left\| Q_1^\top \tilde{b} - R_1 \tilde{x} \right\|_2^2 + \left\| Q_2^\top \tilde{b} \right\|_2^2.
\]

**Strategy 2:**

\[
\left\| \tilde{b} - A\tilde{x} \right\|_2^2 = \left\| \tilde{b} - Q_1 R_1 \tilde{x} \right\|_2^2 \overset{(II)}{=} \left\| Q_1^\top \tilde{b} - Q_1^\top Q_1 R_1 \tilde{x} \right\|_2^2 \overset{(III)}{=} \left\| Q_1^\top \tilde{b} - R_1 \tilde{x} \right\|_2^2.
\]

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): \( \left\| \tilde{b} - A\tilde{x} \right\|_2^2 \overset{(I)}{=} \left\| Q^\top \tilde{b} - Q^\top A\tilde{x} \right\|_2^2 \)

ii. Equality (II): \( \left\| \tilde{b} - Q_1 R_1 \tilde{x} \right\|_2^2 \overset{(II)}{=} \left\| Q_1^\top \tilde{b} - Q_1^\top Q_1 R_1 \tilde{x} \right\|_2^2 \)

iii. Equality (III): \( \left\| Q_1^\top \tilde{b} - Q_1^\top Q_1 R_1 \tilde{x} \right\|_2^2 \overset{(III)}{=} \left\| Q_1^\top \tilde{b} - R_1 \tilde{x} \right\|_2^2. \)
(c) (5 points) Now consider a different matrix $A = QR$, unrelated to the matrix $A$ in previous parts. Here, let

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

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:

$$p^* = \min_{\vec{x}} \| A\vec{x} - \vec{b} \|_2^2.$$

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


5. Subspace projection (18 points)

Consider a set of points $\vec{z}_1, \ldots, \vec{z}_n \in \mathbb{R}^d$. The first principal component of the data, $\vec{w}^*$, is the direction of the line that minimizes the sum of the squared distances between the points and their projections on $\vec{w}^*$, i.e.,

$$\vec{w}^* = \arg\min_{\|\vec{w}\|_2 = 1} \sum_{i=1}^n \| \vec{z}_i - \langle \vec{w}, \vec{z}_i \rangle \vec{w} \|^2.$$

In this problem, we generalize to finding the $r$-dimensional subspace (instead of a 1-dimensional line) that minimizes the sum of the squared distances between the points $\vec{z}_i$ and their projections on the subspace. We assume that $1 \leq r \leq \min(n, d)$. We can represent an $r$-dimensional subspace by its orthonormal basis $(\vec{w}_1, \ldots, \vec{w}_r)$, and we want to solve:

$$\left(\vec{w}^*_1, \ldots, \vec{w}^*_r\right) = \arg\min_{\|\vec{w}_i\|_2 = 1} \sum_{i=1}^n \min_{\alpha_1, \ldots, \alpha_r} \left\| \vec{z}_i - \sum_{k=1}^r \alpha_k \vec{w}_k \right\|^2.$$

Note that the inner minimization projects the point $\vec{z}_i$ onto the subspace defined by $(\vec{w}_1, \ldots, \vec{w}_r)$. The variables $\alpha_k \in \mathbb{R}$. This means that for an arbitrary point $\vec{z}$, this inner minimization

$$\left(\alpha^*_1, \ldots, \alpha^*_r\right) = \arg\min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z} - \sum_{k=1}^r \alpha_k \vec{w}_k \right\|^2$$

has minimizers $\alpha^*_k = \langle \vec{w}_k, \vec{z} \rangle$.

(a) (6 points) With the following definition of matrices $Z$ and $W$:

$$Z = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ \vec{z}_1 & \ldots & \vec{z}_n \\ \downarrow & \cdots & \downarrow \end{bmatrix}, \quad W = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ \vec{w}_1 & \ldots & \vec{w}_r \\ \downarrow & \cdots & \downarrow \end{bmatrix},$$

show that we can rewrite the optimization problem in Equation (??) as:

$$\left(\vec{w}^*_1, \ldots, \vec{w}^*_r\right) = \arg\min_{\|\vec{w}_i\|_2 = 1 \atop \langle \vec{w}_i, \vec{w}_j \rangle = 0 \forall i \neq j} \sum_{i=1}^n \min_{\alpha_1, \ldots, \alpha_r} \left\| Z - WW^\top \right\|_F^2.$$

(b) (6 points) Let $\sigma_i$ refer to the $i$th largest singular value of $Z$, and $l = \min(n, d)$. First show that,

$$\min_{\|\vec{w}_i\|_2 = 1 \atop \langle \vec{w}_i, \vec{w}_j \rangle = 0 \forall i \neq j} \left\| Z - WW^\top \right\|_F^2 \geq \sum_{i=r+1}^l \sigma_i^2.$$

(c) (6 points) Again $\sigma_i$ refers to the $i$th largest singular value of $Z$, and $l = \min(n, d)$. Show that,

$$\min_{\|\vec{w}_i\|_2 = 1 \atop \langle \vec{w}_i, \vec{w}_j \rangle = 0 \forall i \neq j} \left\| Z - WW^\top \right\|_F^2 \leq \sum_{i=r+1}^l \sigma_i^2.$$

Hint: Find a $W$ that achieves this upper bound.
6. Duality (36 points)

Consider the function

\[ f(\bar{x}) = \bar{x}^\top A \bar{x} - 2 \bar{b}^\top \bar{x}. \]

First, we consider the unconstrained optimization problem

\[ p^* = \min_{\bar{x} \in \mathbb{R}^n} f(\bar{x}) = \min_{\bar{x} \in \mathbb{R}^n} \bar{x}^\top A \bar{x} - 2 \bar{b}^\top \bar{x}, \tag{6} \]

for a real \( n \times n \) symmetric matrix \( A \in \mathbb{S}^n \) and \( \bar{b} \in \mathbb{R}^n \). If the problem is unbounded below, then we say \( p^* = -\infty \). Let \( \bar{x}^* \) denote the minimizing argument of the optimization problem.

(a) (6 points) Suppose \( A \succeq 0 \) (positive semidefinite) and \( \bar{b} \in \mathcal{R}(A) \). Let \( \text{rank}(A) = n \). Find \( p^* \).

(b) (8 points) Suppose \( A \succeq 0 \) (positive semidefinite) and \( \bar{b} \in \mathcal{R}(A) \) as before. Let \( A \) be rank-deficient, i.e., \( \text{rank}(A) = r < n \). Let \( A \) have the compact/thin and full SVD as follows, with diagonal positive definite \( \Lambda_r \in \mathbb{R}^{r \times r} \):

\[ A = U_r \Lambda_r U_r^\top = \begin{bmatrix} U_r & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^\top \\ U_1^\top \end{bmatrix}. \]

Show that the minimizer \( \bar{x}^* \) of the optimization problem (6) is not unique by finding a general form for the family of solutions for \( \bar{x}^* \) in terms of \( U_r, U_1, \Lambda_r, \bar{b} \).

(c) (6 points) If \( A \not\succeq 0 \) (\( A \) not positive semi-definite) show that \( p^* = -\infty \) by finding \( \bar{v} \) such that \( f(\alpha \bar{v}) \to -\infty \) as \( \alpha \to \infty \).

(d) (6 points) Suppose \( A \succeq 0 \) (positive semidefinite) and \( \bar{b} \notin \mathcal{R}(A) \). Find \( p^* \). Justify your answer mathematically.

For parts (a) and (b), consider real \( n \times n \) symmetric matrix \( A \in \mathbb{S}^n \) and \( \bar{b} \in \mathbb{R}^n \). Let \( \text{rank}(A) = r \), where \( 0 \leq r \leq n \). Now we consider the constrained optimization problem

\[ p^* = \min_{\bar{x} \in \mathbb{R}^n} \bar{x}^\top A \bar{x} - 2 \bar{b}^\top \bar{x} \]

\[ \text{s.t. } \bar{x}^\top \bar{x} \geq 1, \tag{7} \]

(e) (4 points) Write the Lagrangian \( \mathcal{L}(\bar{x}, \lambda) \), where \( \lambda \) is the dual variable corresponding to the inequality constraint.

(f) (6 points) For any matrix \( C \in \mathbb{R}^{n \times n} \) with \( \text{rank}(C) = r \leq n \) and compact SVD

\[ C = U_r \Lambda_r V_r^\top, \]

we define the pseudoinverse as

\[ C^\dagger = V_r \Lambda_r^{-1} U_r^\top. \]

We use the “dagger” operator to represent this. If \( \bar{d} \) lies in the range of \( C \), then a solution to the equation \( C \bar{x} = \bar{d} \) can be written as \( \bar{x} = C^\dagger \bar{d} \), even when \( C \) is not full rank. Show that the dual problem to the primal problem (7) can be written as,

\[ d^* = \max_{\lambda \geq 0, \bar{b} \in \mathcal{R}(A-\lambda I)} \left\{ (A-\lambda I)^\dagger \bar{b} + \lambda \right\}. \]

Hint: To show this, first argue that when the constraints are not satisfied \( \min_{\bar{x}} \mathcal{L}(\bar{x}, \lambda) = -\infty \). Then show that when the constraints are satisfied, \( \min_{\bar{x}} \mathcal{L}(\bar{x}, \lambda) = -\bar{b}^\top (A-\lambda I)^\dagger \bar{b} + \lambda. \)