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 $f(\vec{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

   Solution:

   Not a convex function because the Hessian matrix is not positive semi-definite.

   (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.*

   Solution: Convex set because the epigraph of a convex function is convex.

   (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 $\vec{b}$.*

   Solution: Convex function because the point-wise maximum of convex functions (in this case affine functions) in $\vec{x}$ is convex.

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 $\vec{x}^* \doteq \arg\min_{\vec{x}} f(\vec{x})$.

   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.
   Solution: $f(\vec{x}) = \frac{1}{4} \|\vec{x}\|_2^4 \geq 0$.
   Also, $f(\vec{0}) = 0$. Thus, $\vec{x}^* = \vec{0}$.

   (b) (8 points) Suppose $\|\vec{x}_0\|_2 = c \neq 0$. **Find the range of η (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.

Solution: Using chain rule, we can compute the gradient of f to get,

$$\nabla f(\vec{x}_t) = \frac{1}{4}.2 \, \|\vec{x}_t\|_2^2 \, \nabla(\|\vec{x}_t\|_2^2)$$
$$= \|\vec{x}_t\|_2^2 \, \vec{x}_t.$$

Using this along with part (a) and the gradient step,

$$|\vec{x}_{t+1} - \vec{x}^*| = \left|\vec{x}_t - \eta \, \|\vec{x}_t\|_2^2 \, \vec{x}_t - 0\right|$$
$$= |\vec{x}_t - 0| \left|(1 - \eta \, \|\vec{x}_t\|_2^2)\right|.$$

To guarantee convergence to $\vec{x}^*$ we require for all t,

$$\left|(1 - \eta \, \|\vec{x}_t\|_2^2)\right| < 1$$
$$\implies 0 < \eta < \frac{2}{\|\vec{x}_t\|_2^2}.$$

But observing that if $\left|(1 - \eta \, \|\vec{x}_t\|_2^2)\right| < 1$ then $|\vec{x}_{t+1}| < |\vec{x}_t|$ so the lowest upper bound for η will be for $t = 0$. Thus, we need,

$$0 < \eta < \frac{2}{\|x_0\|_2^2}.$$

or equivalently,

$$0 < \eta < \frac{2}{c^2}.$$

3. PCA (12 points)

In this problem, we will find the principal components of data points on a regularly spaced grid.[1]
Consider a set S of $n = 15$ data points that lie at each integer node of a 5×3 grid:

$$S = \left\{ \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \;\middle|\; x_1 \in \{-2, -1, 0, 1, 2\}, \; x_2 \in \{-1, 0, 1\} \right\}.$$
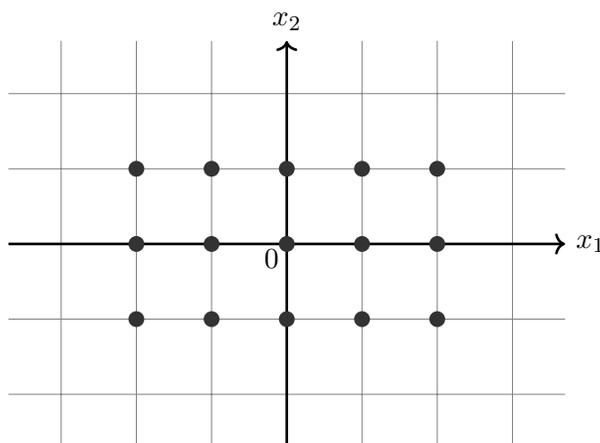
A plot of these points is shown in Fig. **??**.

Figure 1: Point data.

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.,

$$\vec{u}_1 = \underset{\|\vec{u}\|_2 = 1}{\operatorname{argmax}} \, \sigma^2(\vec{u}).$$

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

Solution: The first principal component $\vec{u}_1$ is the eigenvector corresponding to the largest eigenvalue of C, so $\vec{u}_1 = \vec{e}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$, and $\sigma^2(\vec{u}_1) = \vec{u}_1^\top C \vec{u}_1 = 2$. The plot of $\vec{u}_1$ is shown in Fig. **??**.

(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

---

[1]You 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!
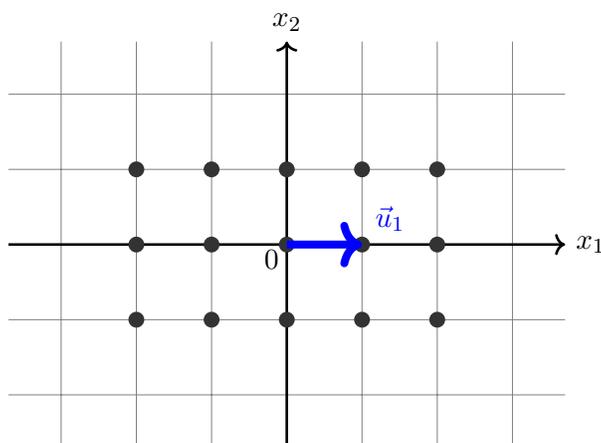
Figure 2: First principal component $\vec{u}_1$.

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.
Solution: We first calculate the transformed data's covariance matrix C_W:

$$
\begin{aligned}
C_W &= \frac{1}{n}\sum_{i=1}^{n}\vec{z}_i\vec{z}_i^\top \\
&= \frac{1}{n}\sum_{i=1}^{n}W\vec{x}_i\vec{x}_i^\top W^\top \\
&= W\left[\frac{1}{n}\sum_{i=1}^{n}\vec{x}_i\vec{x}_i^\top\right]W^\top \\
&= WCW^\top.
\end{aligned}
$$

We now wish to calculate first principal component v_1:

$$
\begin{aligned}
\vec{v}_1 &= \operatorname*{argmax}_{\|\vec{v}\|_2=1}\vec{v}^\top C_W\vec{v} \\
&= \operatorname*{argmax}_{\|\vec{v}\|_2=1}\vec{v}^\top WCW^\top\vec{v} \\
&= \operatorname*{argmax}_{\|W^\top\vec{v}\|_2=1}\vec{v}^\top WCW^\top\vec{v}
\end{aligned}
$$

where the third line follows because multiplication by an orthonormal matrix does not change the norm of a vector. Defining $\vec{y} \doteq W^\top\vec{v}$ (and thus $\vec{y}_1 \doteq W^\top\vec{v}_1$), we can write

$$
\vec{y}_1 = \operatorname*{argmax}_{\|\vec{y}\|_2=1}\vec{y}^\top C\vec{y},
$$

which is exactly equivalent to finding the first principal component of the untransformed data, so $\vec{y}_1 = \vec{u}_1$. Thus, $\vec{v}_1 = W\vec{y}_1 = W\vec{u}_1$. The second transformed principal component must be orthogonal to the first, so $\vec{v}_2 = W\vec{u}_2$.

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

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

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.

Solution:

Using the Gram Schmidt procedure we have,

$$r_{12} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$= 1.$$

Denoting $\vec{q} = \begin{bmatrix} q_{12} \\ q_{22} \\ q_{23} \end{bmatrix}$ and by using the fact that $\vec{q}$ must be unit-norm and that $r_{22} > 0$ we have,

$$r_{22}\vec{q} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - r_{12} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

which gives us $r_{22} = \sqrt{2}$ and $\vec{q} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

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

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \tag{2}$$

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

$$p^* = \min_{\vec{x}} \left\| A\vec{x} - \vec{b} \right\|_2^2 \tag{3}$$

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

Strategy 1:

$$\left\|\vec{b} - A\vec{x}\right\|_2^2 \stackrel{(I)}{=} \left\|Q^\top \vec{b} - Q^\top A\vec{x}\right\|_2^2$$
$$= \left\|Q_1^\top \vec{b} - R_1\vec{x}\right\|_2^2 + \left\|Q_2^\top b\right\|_2^2.$$

Strategy 2:

$$\left\|\vec{b} - A\vec{x}\right\|_2^2 = \left\|\vec{b} - Q_1 R_1 \vec{x}\right\|_2^2$$
$$\stackrel{(II)}{=} \left\|Q_1^\top \vec{b} - Q_1^\top Q_1 R_1 \vec{x}\right\|_2^2$$
$$\stackrel{(III)}{=} \left\|Q_1^\top \vec{b} - R_1 \vec{x}\right\|_2^2.$$

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

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

iii. Equality (III): $\left\|Q_1^\top \vec{b} - Q_1^\top Q_1 R_1 \vec{x}\right\|_2^2 \stackrel{(III)}{=} \left\|Q_1^\top \vec{b} - R_1 \vec{x}\right\|_2^2.$

Solution:

Equality (I) is CORRECT. Since Q is an orthogonal matrix we have $QQ^\top = I$. Thus,

$$\left\|\vec{b} - A\vec{x}\right\|_2^2 = (\vec{b} - A\vec{x})^\top (\vec{b} - A\vec{x})$$
$$= (\vec{b} - A\vec{x})^\top QQ^\top (\vec{b} - A\vec{x})$$
$$= \left\|Q^\top (\vec{b} - A\vec{x})\right\|_2^2$$

Equality (II) is INCORRECT. We can try to apply same approach as before but now a crucial difference is that Q_1 is **NOT** an orthogonal matrix. In fact $Q_1 Q_1^\top$ cannot be full rank since $\text{rank}(Q_1) = 2 < 3$, thus it cannot be I.

Equality (III) is CORRECT. Even though Q_1 is not an orthogonal matrix, its columns are orthonormal thus $Q_1^\top Q_1 = I$.

(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:

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

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 $\vec{b}$. **Solution:** Using Equation (**??**) from the correct solution in the previous part, we have

$$p^* = \left\| Q_2^\top \vec{b} \right\|_2^2$$

$$= \left(\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right)^2$$

$$= 4.$$

Further, $\vec{x}^* = R_1^{-1} Q_1^\top \vec{b}$. Thus,

$$\vec{y}^* = Q_1 R_1 \vec{x}^* = Q_1 Q_1^\top \vec{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

.

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.,

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

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 $1 \le r \le \min(n, d)$. We can represent an r-dimensional subspace by its orthonormal basis $(\vec{w}_1, \ldots, \vec{w}_r)$, and we want to solve:

$$(\vec{w}_1^*, \ldots, \vec{w}_r^*) = \underset{\substack{\|\vec{w}_i\|_2=1 \\ \langle \vec{w}_i, \vec{w}_j \rangle = 0 \ \forall i \ne j \\ 1 \le i,j \le r}}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z}_i - \sum_{k=1}^{r} \alpha_k \vec{w}_k \right\|^2. \tag{4}$$

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

$$(\alpha_1^*, \ldots, \alpha_r^*) = \underset{\alpha_1,\ldots,\alpha_r}{\operatorname{argmin}} \left\| \vec{z} - \sum_{k=1}^{r} \alpha_k \vec{w}_k \right\|^2$$

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

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

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

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

$$(\vec{w}_1^*, \ldots, \vec{w}_r^*) = \underset{\substack{\|\vec{w}_i\|_2=1 \\ \langle \vec{w}_i, \vec{w}_j \rangle = 0 \ \forall i \ne j \\ 1 \le i,j \le r}}{\operatorname{argmin}} \left\| Z - WW^\top Z \right\|_F^2. \tag{5}$$

Solution:

First, consider a single vector $\vec{z} \in \mathbf{R}^d$. For this vector, consider the optimization problem:

$$\min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z} - \sum_{i=1}^{r} \alpha_i \vec{w}_i \right\|_2^2.$$

We first expand the term inside the minimization problem as follows:

$$\left\| \vec{z} - \sum_{i=1}^{r} \alpha_i \vec{w}_i \right\|_2^2 = \|\vec{z}\|_2^2 + \left\| \sum_{i=1}^{r} \alpha_i \vec{w}_i \right\|_2^2 - 2 \langle \sum_{i=1}^{r} \alpha_i \vec{w}_i, \vec{z} \rangle = \|\vec{z}\|_2^2 + \sum_{i=1}^{r} \sum_{j=1}^{r} \alpha_i \alpha_j \langle \vec{w}_i, \vec{w}_j \rangle - 2 \sum_{i=1}^{r} \alpha_i \langle \vec{w}_i, \vec{z} \rangle$$

$$= \|\vec{z}\|_2^2 + \sum_{i=1}^{r} (\alpha_i^2 - 2\alpha_i \langle \vec{w}_i, \vec{z} \rangle)$$

where for the final equality, we have used the fact that $\langle \vec{w}_i, \vec{w}_j \rangle = 0$ for $i \neq j$ and $\|\vec{w}_i\| = 1$ for all i. By taking derivatives, we see that the optimal value for α_i is $\langle \vec{w}_i, \vec{z} \rangle$. From this, we can conclude that for a fixed vector, $\vec{z}$, we get:

$$\min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z} - \sum_{i=1}^{r} \alpha_i \vec{w}_i \right\|^2 = \left\| \vec{z} - \sum_{i=1}^{r} \langle \vec{w}_i, \vec{z} \rangle \vec{w}_i \right\|^2. \tag{6}$$

In this question you the optimizers α_j^* were given and it was sufficient to plug those in to arrive at this step. Now, observe that for a single vector, $\vec{z}$, we have:

$$WW^\top \vec{z} = W \begin{bmatrix} \langle \vec{w}_1, \vec{z} \rangle \\ \vdots \\ \langle \vec{w}_r, \vec{z} \rangle \end{bmatrix} = \sum_{i=1}^{r} \langle \vec{w}_i, \vec{z} \rangle \vec{w}_i.$$

Therefore, we get using the fact that the squared Frobenius norm of a matrix is the sum of the squared lengths of its columns:

$$\left\| Z - WW^\top Z \right\|_F^2 = \sum_{i=1}^{n} \left\| \vec{z}_i - WW^\top \vec{z}_i \right\|^2 = \sum_{i=1}^{n} \left\| \vec{z}_i - \sum_{j=1}^{r} \langle \vec{z}_i, \vec{w}_j \rangle \vec{w}_j \right\|^2.$$

From Equation **??**, we conclude that the above expression is equivalent to **??**.

Next, we will solve the optimization problem in Equation (**??**) using the SVD of Z.

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

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

Solution:

Let $Z = U\Sigma V^\top = \sum_{i=1}^{l} \sigma_i \vec{u}_i \vec{v}_i^\top$ denote the SVD of Z and let $Z_r = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top$. Note that for any $W \in \mathbb{R}^{d \times r}$, $WW^\top Z$ is a matrix of rank at most r. Therefore, we get from the Eckart-Young theorem that:

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

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

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

Hint: Find a W that achieves this upper bound.

Solution:

As before, let $Z = U\Sigma V^\top = \sum_{i=1}^{l} \sigma_i \vec{u}_i \vec{v}_i^\top$ denote the SVD of Z and $Z_r = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top$. By picking $\vec{w}_i = \vec{u}_i$ for $i \in [r]$ in (**??**), we get that:

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

From the previous part and this result, we conclude that an optimal solution to **??** are the top-r left singular vectors of Z which can be computed via the SVD of Z.

6. Duality (36 points)

Consider the function

$$f(\vec{x}) = \vec{x}^\top A \vec{x} - 2\vec{b}^\top \vec{x}.$$

First, we consider the unconstrained optimization problem

$$p^* = \min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) = \min_{\vec{x} \in \mathbb{R}^n} \vec{x}^\top A \vec{x} - 2\vec{b}^\top \vec{x} \tag{7}$$

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

(a) (6 points) Suppose $A \succeq 0$ (positive semidefinite) and $\vec{b} \in \mathcal{R}(A)$. Let $\text{rank}(A) = n$. **Find p^*.**
  Solution: If $\text{rank}(A) = n$, then $A \succ 0$, and therefore the objective is strictly convex. Setting the gradient to 0 we obtain,

$$\nabla_{\vec{x}} f(\vec{x}) = 2A\vec{x} - 2\vec{b} = 0$$
$$\implies A\vec{x} = \vec{b}$$
$$\implies \vec{x}^* = A^{-1}\vec{b}$$

Where in the last step, we used that fact that a full rank square matrix is invertible. Plugging this back into our objective function we get,

$$f(\vec{x}^*) = (\vec{b}^\top (A^{-1})^\top)A(A^{-1}\vec{b}) - 2\vec{b}^\top(A^{-1}\vec{b})$$
$$= \vec{b}^\top (A^\top)^{-1} \cancel{AA^{-1}}\vec{b} - 2\vec{b}^\top A^{-1}\vec{b}$$
$$= \vec{b}^\top A^{-1}\vec{b} - 2\vec{b}^\top A^{-1}\vec{b}$$
$$p^* = -\vec{b}^\top A^{-1}\vec{b}$$

(b) (8 points) Suppose $A \succeq 0$ (positive semidefinite) and $\vec{b} \in \mathcal{R}(A)$ 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 $\Lambda_r \in \mathbb{R}^{r \times r}$:

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

Show that the minimizer $\vec{x}^*$ of the optimization problem (**??**) is not unique by **finding a general form for the family of solutions for $\vec{x}^*$** in terms of $U_r, U_1, \Lambda_r, \vec{b}$.
  Solution: Since $A \succeq 0$, $f(\vec{x})$ is convex and we can attempt to find the minimizer by setting the gradient to zero. Doing this we obtain,

$$A\vec{x} = b, \tag{8}$$

as in the part (a) of this problem.
However, now this equation has infinite solutions since $\vec{b}$ lies in the range of A and A is rank-deficient. Indeed we can add any vector from the (non-trivial) nullspace of A to any particular

solution $\vec{x}_0$ of Equation (**??**) and get another solution.
By the Fundamental Theorem of Linear Algebra we have,

$$\vec{x} = U_r\vec{\alpha} + U_1\vec{\beta}$$
$$\vec{b} = U_r\vec{\gamma},$$

where we used the fact that $\vec{b} \in \mathcal{R}(A)$. Using this we obtain,

$$U_r\Lambda_r U_r^\top (U_r\vec{\alpha} + U_1\vec{\beta}) = U_r\vec{\gamma}$$

Since the columns of U_1 and U_r are orthogonal to each other and because $U_r^\top U_r = I, \Lambda_r$ is invertible we have,

$$U_r\Lambda_r U_r^\top U_r\vec{\alpha} = U_r\vec{\gamma}$$
$$\implies \vec{\alpha} = \Lambda_r^{-1}\vec{\gamma}$$
$$= \Lambda_r^{-1}U_r^\top \vec{b}.$$

Thus any solution to Equation (**??**) and hence a minimizer to the optimization problem (**??**) can be written as,

$$\vec{x}^* = U_r\Lambda_r^{-1}U_r^\top \vec{b} + U_1\vec{\beta}.$$

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

Solution: Since $A \nsucceq 0$ there exists an eigenvalue, eigenvector pair $(\mu, \vec{v})$ such that

$$\vec{v}^\top A\vec{v} = \mu < 0.$$

Assuming without loss of generality that $-2\vec{b}^\top \vec{v} \leq 0$ (If it is positive then multiply $\vec{v}$ by -1) we can take $\vec{x} = \alpha\vec{v}$ to obtain,

$$f(\vec{x}) = f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} + \alpha(-2\vec{b}^\top \vec{v}),$$

which goes to $-\infty$ as α goes to ∞ since $\vec{v}^\top A\vec{v} < 0$ and $-2\vec{b}^\top \vec{v} \leq 0$.

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

Solution: First, note that since A is symmetric, we have $\mathcal{R}(A) = \mathcal{R}(A^\top)$. We have $\vec{b} = \vec{v}_1 + \vec{v}_2$ with $\vec{v}_1 \in \mathcal{R}(A) = \mathcal{R}(A^\top)$ and $\vec{v}_2 \in \mathcal{N}(A)$ as $\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{N}(A)$ from the Fundamental Theorem of Linear Algebra. We cannot have $\vec{v}_2 = 0$ as otherwise we'd get $\vec{b} = \vec{v}_1 \in \mathcal{R}(A)$ which is a contradiction. Now, let $\vec{v} = \vec{v}_2$. We get from this:

$$f(\alpha\vec{v}) = \alpha^2\vec{v}^\top A\vec{v} - 2\alpha(\vec{v}_1 + \vec{v}_2)^\top \vec{v}_2 = 0 - 2\alpha \|\vec{v}_2\|^2$$

where we used the fact that $\vec{v}_2 \in \mathcal{N}(A)$ and $\vec{v}_1 \in \mathcal{R}(A)$. As $\alpha \to \infty$, we get that $f(\alpha\vec{v}) \to -\infty$ from which we conclude that $p^* = -\infty$.

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

$$p^* = \min_{\vec{x} \in \mathbb{R}^n} \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} \tag{9}$$
$$\text{s.t. } \vec{x}^\top \vec{x} \geq 1.$$

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

Solution:

$$\mathcal{L}(\vec{x}, \lambda) = \vec{x}^\top A\vec{x} - 2\vec{b}^\top \vec{x} + \lambda(1 - \vec{x}^\top \vec{x})$$
$$= \vec{x}^\top A\vec{x} - \vec{x}^\top \lambda\vec{x} - 2\vec{b}^\top \vec{x} + \lambda$$
$$= \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda$$

(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 $\vec{d}$ lies in the range of C, then a solution to the equation $C\vec{x} = \vec{d}$, can be written as $\vec{x} = C^\dagger \vec{d}$, even when C is not full rank. **Show that the dual problem to the primal problem (??) can be written as,**

$$d^* = \max_{\substack{\lambda \geq 0 \\ A - \lambda I \succeq 0 \\ \vec{b} \in \mathcal{R}(A - \lambda I)}} -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda.$$

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

Solution:

$$g(\lambda) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \lambda) = \min_{\vec{x}} \vec{x}^\top (A - \lambda I)\vec{x} - 2\vec{b}^\top \vec{x} + \lambda$$

Drawing from parts (c) and (d), we can see that if $A - \lambda I \not\succeq 0$ or if $A - \lambda I \succeq 0, \vec{b} \notin \mathcal{R}(A - \lambda I)$, then we can choose $\vec{x}$ to drive the Lagrangian to $-\infty$.

If the constraints are satisfied, however, then we can proceed like in part (b) by taking the gradient:

$$\nabla_{\vec{x}} \mathcal{L} = 2(A - \lambda I)\vec{x} - 2\vec{b} = 0$$
$$(A - \lambda I)\vec{x} = \vec{b}$$
$$\vec{x}^* = (A - \lambda I)^\dagger \vec{b}$$

where in the last step, we used the fact that the PSD contraint on $A - \lambda I$ is satisfied and $\vec{b}$ lies in the range of $A - \lambda I$, so we can use the pseudoinverse and the gradient-zero point is indeed the minimum.

Plugging this back into the Lagrangian, we get:

$$\mathcal{L}(\vec{x}^*, \lambda) = \vec{b}^\top ((A - \lambda I)^\dagger)^\top (A - \lambda I)(A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda$$
$$= \vec{b}^\top (A - \lambda I)^\dagger (A - \lambda I)(A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda$$
$$= \vec{b}^\top (A - \lambda I)^\dagger \vec{b} - 2\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda$$
$$= -\vec{b}^\top (A - \lambda I)^\dagger \vec{b} + \lambda$$

where we used the fact that $(A - \lambda I)^\dagger$ is symmetric and by properties of pseudo inverse,

$$(A - \lambda I)^\dagger (A - \lambda I)(A - \lambda I)^\dagger = (A - \lambda I)^\dagger.$$

Now, we have a full expression for our dual function:

$$g(\lambda) = \begin{cases} -b^\top (A - \lambda I)^\dagger b + \lambda & \text{if } A - \lambda I \succeq 0, b \in \mathcal{R}(A - \lambda I) \\ -\infty & \text{else} \end{cases}$$

The dual problem follows, as it is just a maximization of the dual function:

$$d^* = \max_{\lambda \geq 0} g(\lambda)$$