Homework 7

Self grades are never due for this assignment.

1. (Sp '19 Midterm 2 #7) Gradient Descent Algorithm

Consider $g : \mathbb{R}^n \to \mathbb{R}$, $g(\vec{x}) = \frac{1}{2}\vec{x}^\top Q\vec{x} - \vec{x}^\top \vec{b}$, where Q is a symmetric positive definite matrix, i.e., $Q \in \mathbb{S}^n_{++}$.

(a) Write the update rule for the gradient descent algorithm

$$\vec{x}_{k+1} = \vec{x}_k - \eta\nabla g(\vec{x}_k), \tag{1}$$

where η is the step size of the algorithm, and bring it into the form

$$(\vec{x}_{k+1} - \vec{x}_\star) = P_\eta(\vec{x}_k - \vec{x}_\star), \tag{2}$$

where $P_\eta \in \mathbb{R}^{n\times n}$ is a matrix that depends on η. Find $\vec{x}_\star$ and P_η in terms of Q, $\vec{b}$ and η. *NOTE*: $\vec{x}_\star$ is a minimizer of g.

Solution: We have $\nabla g(\vec{x}) = Q\vec{x} - \vec{b}$ and

$$\vec{x}_{k+1} = \vec{x}_k - \eta(Q\vec{x}_k - \vec{b}) = \vec{x}_k - \eta Q(\vec{x}_k - Q^{-1}\vec{b}). \tag{3}$$

We can write

$$\vec{x}_{k+1} - Q^{-1}\vec{b} = \vec{x}_k - Q^{-1}\vec{b} - \eta Q(\vec{x}_k - Q^{-1}\vec{b}) = (I - \eta Q)(\vec{x}_k - Q^{-1}\vec{b}). \tag{4}$$

This shows that $\vec{x}_\star = Q^{-1}\vec{b}$ and $P_\eta = I - \eta Q$.

(b) Write a condition on the step size η and the matrix Q that ensures convergence of $\vec{x}_k$ to $\vec{x}_\star$ for every initialization of $\vec{x}_0$.

Solution: From part (a), we have

$$\vec{x}_k - \vec{x}_\star = (I - \eta Q)^k(\vec{x}_0 - \vec{x}_\star). \tag{5}$$

For every initialization $\vec{x}_0$, $(\vec{x}_k - \vec{x}_\star)$ converges to zero if (and only if) all eigenvalues of $(I - \eta Q)$ is in $(-1, 1)$:

$$-1 < 1 - \eta\lambda < 1 \quad \text{for each eigenvalue } \lambda \text{ of } Q. \tag{6}$$

Since Q is positive definite, all of its eigenvalues are positive, and the right hand side of the inequality is satisfied for all $\eta > 0$. For the left hand side of the inequality, we need

$$-1 < 1 - n\lambda \, \forall\lambda_Q \iff \eta < \frac{2}{\lambda_{\max}(Q)}. \tag{7}$$

(c) Assume all eigenvalues of Q are distinct. Let η_m denote the largest stepsize that ensures convergence for all initializations $\vec{x}_0$, based on the condition computed in part (b).

Does there exist an initialization $\vec{x}_0 \neq \vec{x}_\star$ for which the algorithm converges to the minimum value of g for certain values of the step size η that are larger than η_m?

Justify your answer. *HINT: The question asks if such initializations exist; not whether it is practical to find them.*

Solution: From part (a), we have

$$\vec{x}_k - \vec{x}_\star = (I - \eta Q)^k (\vec{x}_0 - \vec{x}_\star). \tag{8}$$

If we want

$$(I - \eta Q)^k (\vec{x}_0 - \vec{x}_\star) \to \vec{0} \quad \text{as } k \to \infty \tag{9}$$

for a specific initialization $\vec{x}_0$, the vector $(\vec{x}_0 - \vec{x}_\star)$ must lie in the eigenspaces of $(I - \eta Q)$ corresponding to the eigenvalues in the range $(-1, 1)$. This explanation gets full credit.

For example, if $\frac{2}{\lambda_1} < \eta < \frac{2}{\lambda_2}$, where λ_1 and λ_2 are the largest two eigenvalues of Q, we have $(I - \eta Q)^k (\vec{x}_0 - \vec{x}_\star) \to \vec{0}$ as long as $(\vec{x}_0 - \vec{x}_\star)$ does not have any component in the eigenspace corresponding to the minimum eigenvalue of $(I - \eta Q)$.

2. (Sp '19 Midterm 2 #3) Convexity of Sets

Determine if each set C given below is convex. Prove that each set is convex or provide an example to show that it is not convex. You may use any techniques used in class or discussion to demonstrate or disprove convexity.

(a) $C = \{\vec{x} \in \mathbb{R}^2 \mid x_1 x_2 \geq 0\}$, where $\vec{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$.

Solution: C **is not convex.** The set C is shown in Fig. **??**: From the figure, it is clear that this set is non-convex. For a formal proof, consider points $\vec{z}_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$ and $\vec{z}_2 = \begin{bmatrix} -1 & 0 \end{bmatrix}^\top$. We have $\vec{z}_1 \in C$ and $\vec{z}_2 \in C$. Then $\vec{z}_3 = \frac{\vec{z}_1 + \vec{z}_2}{2} = \begin{bmatrix} -0.5 & 0.5 \end{bmatrix}^\top \notin C$ since $(-0.5) \cdot 0.5 < 0$.

(b) $C = \{X \in \mathbb{S}^n \mid \lambda_{\min}(X) \geq 2\}$, where $\mathbb{S}^n$ is the set of symmetric matrices in $\mathbb{R}^{n \times n}$ and $\lambda_{\min}(X)$ is the minimum eigenvalue of X.

Solution: C **is convex.** Consider $X_1, X_2 \in C$. The minimum eigenvalue of X is given by

$$\lambda_{\min}(X) = \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \vec{z}^\top X \vec{z}. \tag{10}$$

Thus we have

$$\min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \vec{z}^\top X_1 \vec{z} \geq 2 \quad \text{and} \tag{11}$$

$$\min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \vec{z}^\top X_2 \vec{z} \geq 2. \tag{12}$$

C is convex if for any scalar $\theta \in [0, 1]$, we have $X_\theta \doteq \theta X_1 + (1-\theta) X_2 \in C$. Plugging in the above, we have

$$\begin{aligned}
\lambda_{\min}(X_\theta) &= \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \vec{z}^\top X_\theta \vec{z} && (13)\\
&= \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \vec{z}^\top (\theta X_1 + (1-\theta) X_2) \vec{z} && (14)\\
&= \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \left[\theta \vec{z}^\top X_1 \vec{z} + (1-\theta) \vec{z}^\top X_2 \vec{z}\right] && (15)\\
&\geq \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} \theta \vec{z}^\top X_1 \vec{z} + \min_{\vec{z} \in \mathbb{R}^n \,:\, \|\vec{z}\|_2 = 1} (1-\theta) \vec{z}^\top X_2 \vec{z} && (16)\\
&\geq \theta 2 + (1-\theta) 2 && (17)\\
&= 2. && (18)
\end{aligned}$$

Thus, $X_\theta \in C$, and therefore C is convex.

(c) Let $\mathcal{H}(\vec{w})$ denote the hyperplane with normal direction $\vec{w} \in \mathbb{R}^n$, i.e.,

$$\mathcal{H}(\vec{w}) = \{\vec{x} \in \mathbb{R}^n \mid \vec{x}^\top \vec{w} = 0\}. \tag{19}$$

Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be given by

$$P(\vec{x}) = \operatorname*{argmin}_{\vec{y} \in \mathcal{H}(\vec{w})} \|\vec{y} - \vec{x}\|_2 \,. \tag{20}$$

Let

$$C = \{P(\vec{x}) \mid \vec{x} \in \mathcal{B}\} \tag{21}$$

where $\mathcal{B} = \{\vec{x} \in \mathbb{R}^n \mid \|\vec{x}\|_2 \leq 1\}$.

Solution: C **is convex.** Let $Q \in \mathbb{R}^{n\times(n-1)}$ denote the matrix with columns forming a basis for $H(\vec{w})$. Then the optimization problem for $P(\vec{x})$ can be written as

$$P(\vec{x}) = Q\left[\operatorname*{argmin}_{\vec{w}\in\mathbb{R}^{n-1}} \|Q\vec{w} - \vec{x}\|_2^2\right] \tag{22}$$

and has the closed form solution $P(\vec{x}) = Q(Q^\top Q)^{-1}Q^\top\vec{x} = L\vec{x}$ for $L \doteq Q(Q^\top Q)^{-1}Q^\top$. Note that $P(\vec{x})$ is linear in $\vec{x}$.

Method 1:

$\mathcal{B}$ is a convex set and P is an affine operator. Affine transformations of convex sets are convex, so we conclude directly that C is convex.

Method 2:

Let $\vec{z}_1, \vec{z}_2 \in C$. This means there exist $\vec{x}_1, \vec{x}_2 \in \mathcal{B}$ such that $\vec{z}_1 = L\vec{x}_1$ and $\vec{z}_2 = L\vec{x}_2$. For $\theta \in [0,1]$, we consider $\vec{x}_\theta \doteq \theta\vec{x}_1 + (1-\theta)\vec{x}_2$. Because $\mathcal{B}$ is convex (since norm balls are convex), we have $\vec{x}_\theta \in \mathcal{B}$. Then,

$$\vec{z}_\theta \doteq \theta\vec{z}_1 + (1-\theta)\vec{z}_2 \tag{23}$$
$$= \theta L\vec{x}_1 + (1-\theta)L\vec{x}_2 \tag{24}$$
$$= L(\theta\vec{x}_1 + (1-\theta)\vec{x}_2) \tag{25}$$
$$= L\vec{x}_\theta \tag{26}$$
$$= P(\vec{x}_\theta). \tag{27}$$

Thus, $\vec{z}_\theta \in C$, so C is convex by the definition of convexity.

3. (Fa '22 Midterm #8) Matrix Square Root

Let $A, B \in \mathbb{S}^n_{++}$ be symmetric positive definite matrices.

As B is symmetric, it has an orthonormal eigendecomposition $B = V\Lambda V^\top$. Since B is positive definite, we can define its matrix square root as follows $B^{1/2} = V\Lambda^{1/2}V^\top$, where $\Lambda^{1/2}$ is a diagonal matrix whose entries are the square roots of the corresponding entries of Λ. We denote the inverse of $B^{1/2}$ as $B^{-1/2}$. Finally, define $C \doteq B^{-1/2}AB^{-1/2}$.

Prove that the maximum eigenvalue of C is $\lambda^\star$, where

$$\lambda^\star \doteq \max_{\vec{x}\neq\vec{0}} \frac{\vec{x}^\top A\vec{x}}{\vec{x}^\top B\vec{x}}. \tag{28}$$

Solution: Define $y = B^{1/2}x$ and hence $B^{-1/2}y = x$. As B is positive definite, $x \neq 0 \iff y \neq 0$. So we can rewrite the optimization problem as

$$\begin{aligned}
\max_{x\neq 0} \frac{x^\top Ax}{x^\top Bx} &= \max_{y\neq 0} \frac{(B^{-1/2}y)^\top A(B^{-1/2}y)}{(B^{-1/2}y)^\top B(B^{-1/2}y)} \\
&= \max_{y\neq 0} \frac{y^\top B^{-1/2}AB^{-1/2}y}{y^\top y}
\end{aligned}$$

Which is the same as finding the maximum eigenvalue and eigenvector of $C = B^{-1/2}AB^{-1/2}$ due to the definition via Rayleigh quotient.

4. (Sp '20 Midterm # 5) Subspace Projection

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

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

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

$$(\vec{w}_1^\star, \ldots, \vec{w}_r^\star) = \underset{\substack{\|\vec{w}_i\|_2=1 \\ \langle \vec{w}_i,\ \vec{w}_j \rangle = 0\ \forall i \neq j \\ 1 \leq i,j \leq r}}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z}_i - \sum_{k=1}^{r} \alpha_k \vec{w}_k \right\|^2 . \tag{29}$$

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

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

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

(a) 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 (29) as:

$$(\vec{w}_1^\star, \ldots, \vec{w}_r^\star) = \underset{\substack{\|\vec{w}_i\|_2=1 \\ \langle \vec{w}_i,\ \vec{w}_j \rangle = 0\ \forall i \neq j \\ 1 \leq i,j \leq r}}{\operatorname{argmin}} \left\| Z - W W^\top Z \right\|_F^2 . \tag{30}$$

Solution:

First, consider a single vector $\vec{z} \in \mathbb{R}^d$. 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:

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

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:

$$\min_{\alpha_1,\ldots,\alpha_r} \left\| \vec{z} - \sum_{i=1}^{r} \alpha_i \vec{w}_i \right\|^2 = \left\| \vec{z} - \sum_{i=1}^{r} \langle \vec{w}_i,\ \vec{z} \rangle \vec{w}_i \right\|^2. \tag{31}$$

In this question you the optimizers $\alpha_j^\star$ 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 31, we conclude that the above expression is equivalent to 29.

Next, we will solve the optimization problem in Equation (30) using the SVD of Z.

(b) Let σ_i refer to the i^{th} 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) Again σ_i refers to the i^{th} 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 (30), 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 29 are the top-r left singular vectors of Z which can be computed via the SVD of Z.

