Problem 1. When solving

$$\min_x \frac{1}{2} x^T M x + y^T x$$

using a gradient descent, where M is an $n \times n$ symmetric positive definite matrix, we have

$$\|x^{(k)} - x_*\| \leq \|I - sM\|^k \|x^{(0)} - x_*\|.$$

where s is the step size. What is the eigenvalue of $I - sM$ in terms of $\lambda_i(M)$ and s? In class, we choose the step size to be

$$\frac{1}{\lambda_{\max}(M)}.$$

Is there a better choice of s (meaning larger than the step size we choose in class) that will lead to a linear convergence? If so please explain.

Solution: To prove the relationship between the M and $I - sM$ we see that:

$$I - sM = U(I - s\Lambda)U^{-1}$$

which implies that:

$$\lambda_{(1-sM)} = 1 - s\lambda_{(M)}$$

To find a more optimal step size, we want to prove the following lemma:

Lemma 1. *The operator norm for a symmtric positive define matrix can be simplied as follows*

$$\|M\| = \max_i |\lambda_i|$$

Proof. We can see the following. Let $V \in \mathbb{R}^n$, we can see that by definition:

$$\|M\| = \sup_{\|x\|=1} \|Mx\|$$

Thus, using the eigenvalue decomposition, we find that:

$$\|M\| = \sup_{\|x\|=1} \|Mx\| = \sup_{\|x\|=1} \|U\Lambda U^{-1}x\|$$

Let $\xi = U^{-1}x$. Thus, we see that:

$$\sup_{\|x\|=1} \|U\Lambda U^{-1}x\| = \sup_{\|\xi\|=1} \|U\Lambda\xi\| = \sup_{\|\xi\|=1} \|\Lambda\xi\|$$

Thus, the suprememum is obtained when $\xi = e_j$ where j corresponds to the index of eigenvalue with the highest magnitutde. Therefore, we see that:

$$\|M\| = \max_i |\lambda_i|$$

$\square$

Thus, we proceed with the proof of finding a better step size. Note that it is given

$$\|x^{(k)} \leq \|I - sM\|^K \|x^{(0)} - x_*\|$$

In class, it was shown that if $s \in (0, \frac{2}{\lambda_{max}(M)})$, then the above linearly converges. Thus, we aim to find an value in here that leads to smaller value of $\|I - sM\|$ relative to $s = \frac{1}{\lambda_{max}(M)}$. Thus, we want to

$$\min \max_i \{|1 - s\lambda_i|\}$$

However, note the linear nature of the objective function. Thus, we can simplify the problem as follows:

$$\min \max\{|1 - s\lambda_{min}(M)|, |1 - s\lambda_{max}(M)|\}$$

Thus, we see that the $\max\{|1 - s\lambda_{min}(M)|, |1 - s\lambda_{max}(M)|\}$ is minimized when

$$|1 - s\lambda_{min}(M)| = |1 - s\lambda_{max}(M)|$$

Thus, we aim to find an s that sastifies the upper bound. We can see that

$$\begin{aligned}
-(1 - s\lambda_{min}(M)) &= 1 - s\lambda_{max}(M) \\
s\lambda_{max}(M) + s\lambda_{min}(M) &= 2 \\
s &= \frac{2}{\lambda_{min}(M) + \lambda_{max}(M)}
\end{aligned}$$

This bound is better than that derived from class as:

$$\frac{2}{\lambda_{min}(M) + \lambda_{max}(M)} > \frac{1}{\lambda_{max}(M)}$$

which implies faster convergence. □

Problem 2. For a $n \times n$ matrix A, let

$$A_k = A_{:,1:k} A_{1:k,1:k}^{-1} A_{1:k,:},$$

assume $A_{1:k,1:k}$ is invertible. Suppose A is a rank-k matrix. Prove that $A = A_k$.

(a) Prove that each column of A comes from a unique linear combination of $A_{:,1:k}$.

(b) Now prove the claim.

(a)

Solution: By assumption $A_{1:k,1:k}$ is invertible, which implies that the first k columns of are linearly independent. Thus, this means that the first k columns are the range of A, as A is rank k. Thus, for any column of A, there exists an unique linear combination of the columns of $A_{:,1:k}$ □

(b)

Solution: Note by the above assumption, this implies that there exists a matrix $C \in \mathbb{R}^{k \times n}$ such that

$$A = A_{:,1:k} C$$

Note that we can subsection the first k rows from each matrix for the above equation, as the structure of C does not change. We can see that this statement holds as

$$A_{:,j} = A_{:,1:k} C_{:,j} \iff A_{1:k,j} = A_{1:k,1:k} C_{:,j} \iff A_{1:k,:} = A_{1:k,1:k} C$$

Rearranging the terms above yields that:

$$C = A_{1:k,1:k}^{-1} A_{1:k,:}$$

Thus, we see that

$$A = A_k = A_{:,1:k} A_{1:k,1:k}^{-1} A_{1:k,:}$$

□

Problem 3. Show that

$$Mv, \ldots, M^k v$$

are linearly independent, if the Hermitian matrix M has no repeated eigenvalues, and every entry of v is non-zero. (Hint: One can assume the fact that if a polynomial $p(t) = c_0 + c_1 t + \cdots + c_{k-1} t^{k-1}$ of degree $k - 1$ has k different roots, then it must be the zero polynomial and $c_0, \ldots, c_{k-1} = 0$.)

Solution: Note the Eigenvalue decomposition where $M = U\Lambda U^*$. To prove linear independence, we set the equation as follows:

$$c_1 Mv + c_2 M^2 v + \cdots + c_k M^k v = 0$$

We substitute the Eigenvalue decomposition as follows:

$$p(M) = c_1(U\Lambda U^*) + c_2(U\Lambda^2 U^*) + \ldots c_k(U\Lambda^k U^*)$$

$$\implies p(M) \cdot v = c_1(U\Lambda U^*)v + c_2(U\Lambda^2 U^*)v + \ldots c_k(U\Lambda^k U^*)v$$

We can rearrange the above polynomial to see that:

$$p(M) \cdot v = U(c_1\Lambda + c_2\Lambda + \cdots + c_k\Lambda^k)U^* v$$

Thus it suffices to analyze the roots of the following polynomial:

$$p(\Lambda) = c_1\Lambda + c_2\Lambda^2 + \ldots c_k\Lambda^k$$

By assumption, there are n eigenvalues where $k < n$. We can expand this out to each elementof the diagnol matrix, where:

$$p(\lambda_i) = c_1\lambda_i + c_2\lambda_i^2 \ldots c_k\lambda_i^k = 0$$

This implies that are at least $k + 1$ roots to a k power polynomial, as indicated by the $k + 1$ eigenvalues. Thus this means that $c_0 = c_1 = \cdots = c_k = 0$. Thus, this means that

$$Mv, \ldots, M^k v$$

are linearly independent. $\square$

Problem 4. Suppose we have the following updating scheme

$$q^{k+1} = \gamma q^k + (1-\gamma)\nabla f(x^k)$$
$$x^{k+1} = x^k - \alpha q^{k+1}.$$

(a) Let f be the function in Problem 1. Let M be a diagonal matrix with diagonals being $\lambda_1, \ldots, \lambda_n$ and $y = 0$. How does $z_i^{k+1} = \left[x_i^{k+1}, q_i^{k+1}\right]^T \in \mathbb{R}^2$ depend on $z_i^k = \left[x_i^k, q_i^k\right]^T$? Write it as $z_i^{k+1} = A_i z_i^k$.

(b) What are the eigenvalues of the 2×2 matrix A_i? One can write $A_i = B^{-1}C_iB$ where

$$B = \begin{bmatrix} 1 & 0 \\ 1 & \alpha \end{bmatrix}.$$

What is B^{-1}? What is the relationship between the eigenvalues of C_i and A_i?

(c) Show that x^k converges to $x_* = 0$, for $\alpha(1-\gamma) = 1/\lambda_{\max}(M)$, $\gamma = (\sqrt{\lambda_{\max}(M)} - \sqrt{\lambda_{\min}(M)})/\sqrt{\lambda_{\max}(M)}$. **Points to consider:** One can understand the convergence of x^k by how z_i^k converges to $z_{i,*}$. Write down z_i^k in terms of z_i^0. The convergence of z_i^k can be understood in terms of $A_i = B^{-1}C_iB$ (consider applying a change of coordinate to $z_i^k \in \mathbb{R}^2$ with the matrix B, and prove the convergence of the transformed variable). Are the eigenvalues of A_i complex or real?

(d) Please generalize the convergence proof in the last part, to general M and y as in Problem 1. The answer should be short.

(a)

Solution: Note that $\nabla f(x)_i = \lambda_i(M)$ due to the diagonal nature of M, which implies that:

$$q_i^{k+1} = \gamma q^k + (1-\gamma)\lambda_i(M) \quad x_i^{k+1} = x_i^k - \alpha q_i^{k+1}$$

Doing the appropriate substitutions, we find that:

$$A_i = \begin{bmatrix} 1 - \alpha(1-\gamma)\lambda_i(M) & -\alpha\gamma \\ (1-\gamma)\lambda_i(M) & \gamma \end{bmatrix}$$

□

(b)

Solution: Applying the formula for the inverse of a 2 by 2 matrix is as follows:

$$B^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{\alpha} & \frac{1}{\alpha} \end{bmatrix}$$

Note that this implies that A_i and C_i have the same eigenvalues as the two matrices are similar to each other. □

(c)

Solution: Note the following:

$$\begin{aligned} z_i^{k+1} &= A_i z_i^k \\ z_i^{k+1} &= B^{-1} C_i B z_i^k \\ B z_i^{k+1} &= C_i B z_i^k \end{aligned}$$

Let $Y_i = Bz_i$, thus we see that

$$B z_i^{k+1} = C_i B z_i^k \iff Y_i^{k+1} = C_i Y_i^k$$

Thus, it suffices to analyze the behavior of C_i, as if Y converges, then z converges, which implies that x converges to its optimal value. We first begin with the calculation of C_i, which is

$$\begin{aligned} C_i &= B A_i B^{-1} \\ &= \begin{bmatrix} 1 & 0 \\ 1 & \alpha \end{bmatrix} \begin{bmatrix} 1 - \alpha(1-\gamma)\lambda_i(M) & -\alpha\gamma \\ (1-\gamma)\lambda_i(M) & \gamma \end{bmatrix} \cdot \frac{1}{\alpha} \begin{bmatrix} \alpha & 0 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 1 & \alpha \end{bmatrix} \cdot \frac{1}{\alpha} \begin{bmatrix} \alpha(1 + \gamma - \alpha(1-\gamma)\lambda_i(M)) & -\alpha\gamma \\ \alpha & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 + \gamma - \alpha(1-\gamma)\lambda_i(M) & -\gamma \\ 1 & 0 \end{bmatrix} \end{aligned}$$

When we aim to calculate the eigenvalues, we see the following:

$$\det(C_i - \lambda I) = \det\left(\begin{bmatrix} 1 + \gamma - \alpha(1-\gamma)\lambda_i(M) - \lambda & -\gamma \\ 1 & -\lambda \end{bmatrix} \right)$$

$$= \lambda^2 - (1 + \gamma - \alpha(1-\gamma)\lambda_i(M))\lambda + \gamma$$

To check if the eigenvalues are real, we check the discriminant:

$$(1 + \gamma - \alpha(1-\gamma)\lambda_i(M))^2 - 4\gamma = \left(1 + \gamma - \frac{\lambda_i(M)}{\lambda_{\max}(M)}\right)^2 - 4\gamma$$

$$= \left(1 + \frac{\sqrt{\lambda_{max}(M)} - \sqrt{\lambda_{min}(M)})}{\sqrt{\lambda_{max}(M)}} - \frac{\lambda_i(M)}{\lambda_{\max}(M)}\right)^2 - 4\left(\frac{\sqrt{\lambda_{max}(M)} - \sqrt{\lambda_{min}(M)})}{\sqrt{\lambda_{max}(M)}}\right) > 0$$

which implies that all eigenvalues of C are real. Now to solve for the roots of the polynomial. For notational sake, let $h = 1 + \gamma - \alpha(1-\gamma)\lambda_i(M)$. Thus, we see that:

$$\lambda = \frac{h \pm \sqrt{h^2 - 4\gamma}}{2}$$

□