EE5120 Linear Algebra: Tutorial 7, July-Dec 2018, Dr. Uday Khankhoje, EE IIT Madras
Covers 6.1,6.2,6.3 of GS

1. The following information is given about a 4×4 matrix A.

$$A[0.5, 0.5, , 0.5, 0.5]^T = 3[0.5, 0.5, 0.5, 0.5]^T,$$
$$A[0.5, -0.5i, -0.5, 0.5i]^T = 2.5[0.5, -0.5, 0.5, -0.5]^T,$$
$$A[0.5, -0.5, 0.5, -0.5]^T = [0.5, 0.5, -0.5, -0.5]^T,$$
$$A[0.5, 0.5i, -0.5, -0.5i]^T = 0.5[0.5, -0.5, -0.5, 0.5]^T,$$

Compute a rank i matrix A by considering only the first i equations from the above ($i = 1, 2, 3, 4$).

Solution: Added by Manoj.

Let $B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & -0.5i & -0.5 & 0.5i \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5i & -0.5 & -0.5i \end{bmatrix}$ and $B_2 = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & -0.5 & 0.5 \end{bmatrix}$. It can be verified that B_1 and B_2 are unitary matrices. Then, the given equations can be written as $AB_1 = B_2 D$, where $D = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$. It is similar to the SVD form where we have $AV = UD$, with $V = B_1$ and $U = B_2$. Hence, rank i matrix A can be written as $A = \sum_{k=1}^{i} d_k \mathbf{u}_k \mathbf{v}_k^H$, where $\mathbf{u}_l$ and $\mathbf{v}_l$ are the l^{th} columns of matrices U and V respectively. Thus, we have,

$$\text{Rank 1 matrix } A = 0.75 \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

$$\text{Rank 2 matrix } A = \begin{bmatrix} 1.375 & 0.75 + 0.625i & 0.125 & 0.75 - 0.625i \\ 0.125 & 0.75 - 0.625i & 1.375 & 0.75 + 0.625i \\ 1.375 & 0.75 + 0.625i & 0.125 & 0.75 - 0.625i \\ 0.125 & 0.75 - 0.625i & 1.375 & 0.75 + 0.625i \end{bmatrix},$$

$$\text{Rank 3 matrix } A = \begin{bmatrix} 1.625 & 0.5 + 0.625i & 0.375 & 0.5 - 0.625i \\ 0.375 & 0.5 - 0.625i & 1.625 & 0.5 + 0.625i \\ 1.125 & 1 + 0.625i & -0.125 & 1 - 0.625i \\ -0.125 & 1 - 0.625i & 1.125 & 1 + 0.625i \end{bmatrix},$$

$$\text{Rank 4 matrix } A = \begin{bmatrix} 1.75 & 0.5 + 0.5i & 0.25 & 0.5 - 0.5i \\ 0.25 & 0.5 - 0.5i & 1.75 & 0.5 + 0.5i \\ 1 & 1 + 0.75i & 0 & 1 - 0.75i \\ 0 & 1 - 0.75i & 1 & 1 + 0.75i \end{bmatrix}.$$

2. Consider a wireless communication system consisting of a transmitter (Tx) and a receiver (Rx). Tx wishes to transmit a data vector $\mathbf{x} \in \mathbb{C}^n$ to Rx. But Tx processes the data $\mathbf{x}$ before transmission. Thus, the final data which Tx transmits is $\tilde{\mathbf{x}} = P\mathbf{x}$, where $\tilde{\mathbf{x}}$ is the processed data and the $n \times n$ matrix P describes how Tx had processed the original data vector $\mathbf{x}$. The

transmitted signal vector $\tilde{\mathbf{x}}$ passes through the wireless channel represented by the matrix H of rank r. The channel matrix H is known both to Tx and Rx. The signal vector received by Rx is given by,

$$\tilde{\mathbf{y}} = H\tilde{\mathbf{x}} + \tilde{\mathbf{w}},$$

where $\tilde{\mathbf{y}} \in \mathbb{C}^m$ is the observation vector and $\tilde{\mathbf{w}}$ is the low power additive unknown noise vector. Then Rx processes $\tilde{\mathbf{y}}$ and obtains a vector $\mathbf{y}$ as, $\mathbf{y} = Q\tilde{\mathbf{y}}$. It is given that $\mathbf{w} = Q\tilde{\mathbf{w}}$ will still be a low power unknown noise vector. Finally, Rx has to predict the information (or the original data vector $\mathbf{x}$) sent by Tx, for which Rx computes the least squares estimate of $\mathbf{x}$, denoted as $\hat{\mathbf{x}}_{\text{LS}}$. Now, answer the questions that follow:

(a) Suppose $r = n = m$ and H has n-dimensional eigen-space. Let $H = S\Lambda S^{-1}$ be the eigen-value decomposition of H.

  (i) Write the expression for $\mathbf{y}$ in terms of $\mathbf{x}$ (the original data vector) and $\mathbf{w}$ (the processed noise vector), if $P = S$ and $Q = S^{-1}$.

  (ii) Compute a simplified expression for $\hat{\mathbf{x}}_{\text{LS}}$. Observe that $\hat{\mathbf{x}}_{\text{LS}}(i) = a_i\mathbf{y}(i), \forall i = 1, ..., n$ where a_i's are some scalars, i.e., each element of $\mathbf{x}$ can be estimated independent of the other.

  (iii) What are those a_i's?

  [Note: For a vector $\mathbf{g}$, $\mathbf{g}(i)$ refers to i^{th} entry in the vector $\mathbf{g}$].

(b) Let $r < n \leq m$.

  (i) Choose the matrices P and Q using SVD of H so that elements of $\mathbf{x}$ can be estimated independent of each other (just like the way in part (a)).

  (ii) As done in (a), write an expression for $\mathbf{y}$ in terms of $\mathbf{x}$ and $\mathbf{w}$.

  (iii) However, it is imperative to note here that you will not be able to estimate all the entries in $\mathbf{x}$. Guess why?

  (iv) As a result of (iii), Tx will send only those many data elements in $\mathbf{x}$ that can be estimated, with remaining entires set to zero. Can you now say how many data points can be transmitted and estimated successfully in this case? Say the number of data points that can be sent is N. Can it be any N elements of $\mathbf{x}$?

  (v) Having answered these, can you write a simplified design for matrices P and Q, i.e., fill only certain columns with what is needed and fill the remaining part of these matrices with zeros? Based on this design of P and Q, write the structure of the original data vector $\mathbf{x}$. Answer clearly.

(c) From the expression of $\mathbf{y}$ in part (b), the singular values of H, say $\left\{\sigma_k\right\}_{k=1}^{r}$, take the interpretation of the gain values offered by the channel H to the data elements present in $\mathbf{x}$. Suppose $0 \approx \sigma_i <<< \mathbf{w}(i)$, but clearly, $\sigma_i \neq 0, \forall i \in \mathcal{S}$, where $\mathcal{S} \subset \{1, ..., r\}$, and assuming that this information is known to Tx and Rx, what is the best way for the Tx to transmit data vectors so that Rx makes considerably a decent job of estimating them successfully with very less error.

(d) For this question, assume that Tx and Rx does not have the knowledge of H, but they know that rank of H is r. Both Tx and Rx obtain two estimates of H (somehow), say $\hat{H}_1$ and $\hat{H}_2$. Let $\left\{\alpha_k\right\}_{k=1}^{r}$ and $\left\{\beta_k\right\}_{k=1}^{r}$ be the sets of singular values of $\hat{H}_1$ and $\hat{H}_2$, such that $\alpha_1 = \beta_r = 0$. Let all the non-zero α_i's and β_i's be sufficiently large compared to the noise power level and $\alpha_i = \beta_i, \forall i = 2, ..., r-1$. Also assume that there is no error in estimating the unitary matrices that appears in the SVD of H. Again, the goal is to obtain estimates of entries (that can be possible) of $\mathbf{x}$ independent of each other (refer to condition in part (a) with scalars a_i's).

(i) How would you now precisely and compactly design matrices P and Q using the estimates **[case 1]** $\hat{H}_1$ alone, **[case 2]** $\hat{H}_2$ alone, and **[case 3]** $\hat{H}_1$ and $\hat{H}_2$ both?

(ii) Can the system successfully transmit and receive the same number of data points as found in part (b) in each of the above cases? Explain your answer clearly.

(iii) What would be the a_i's in part(a) here, for each of the above cases?

Solution: Added by Manoj.

(a) Here, $P = S$ and $Q = S^{-1}$. Hence, we have,

$$\begin{aligned}\mathbf{y} = Q\tilde{\mathbf{y}} &= Q\left(H\tilde{\mathbf{x}} + \tilde{\mathbf{w}}\right) = QHP\mathbf{x} + \mathbf{w} \\ &= S^{-1}\left(S\Lambda S^{-1}\right)S\mathbf{x} + \mathbf{w} \\ &= \Lambda\mathbf{x} + \mathbf{w}.\end{aligned}$$

Λ is the eigenvalue matrix with all eigenvalues present along its diagonal. Let i^{th} diagonal entry in Λ be λ_i. Note that H is given to be a full-rank matrix. This implies, $\lambda_i \neq 0$, $\forall i$ and matrix Λ is invertible. Thus, least squares estimate of $\mathbf{x}$ is given by,

$$\hat{\mathbf{x}}_{\text{LS}} = \Lambda^{-1}\mathbf{y}.$$

But since Λ is a diagonal matrix, so is its inverse, with the diagonal entries being reciprocal of diagonal entries of Λ. Thus, we get,

$$\hat{\mathbf{x}}_{\text{LS}}(i) = \frac{\mathbf{y}(i)}{\lambda_i},$$

thereby, proving that every element of $\mathbf{x}$ can be estimated independent of the other. (The scalars a_i's mentioned in the question are nothing but $\frac{1}{\lambda_i}$'s).

(b) H is an $m \times n$ matrix with rank r. Thus, its SVD can be written as, $H = U\Sigma V^H$, where V and U are $n \times n$ and $m \times m$ unitary matrices and Σ is an $m \times n$ matrix with the elements at $(i,i)^{th}$ position having singular values of H, for $i = 1, ..., r$ and all remaining entries of Σ being zero. In this case, optimal choices for P and Q matrices would be,

$$P = V \text{ and } Q = U^H.$$

Note that Q is an $m \times m$ matrix here. As a consequence, we get,

$$\begin{aligned}\mathbf{y} = QHP\mathbf{x} + \mathbf{w} &= U^H\left(U\Sigma V^H\right)V\mathbf{x} + \mathbf{w} \\ &= \Sigma\mathbf{x} + \mathbf{w},\end{aligned}$$

and $\mathbf{y}$ is an $m \times 1$ vector, which will be of the form $\mathbf{y}(i) = \sigma_i\mathbf{x}(i) + \mathbf{w}(i)$, $\forall i = 1, ..., r$ and $\mathbf{y}(i) = \mathbf{w}(i)$, $\forall i > r$, where σ_i's are the singular values (non-zero elements in D). Due to this, the received observation vector will have information only about r entries in $\mathbf{x}$. Because of this reason, only those r entries can be estimated at the receiver and not the remaining. So, as per the notation given in the question $N = r$. And, the least square estimate of $\mathbf{x}$ will be,

$$\hat{\mathbf{x}}_{\text{LS}}(i) = \frac{\mathbf{y}(i)}{\sigma_i}, \; i = 1, ..., r.$$

As required, we are able to estimate data points in $\mathbf{x}$ individually. Now, SVD of H can be written as $U_r \Sigma_r V_r^H$, where V_r and U_r are $n \times r$ and $m \times r$ sub-matrices of V and U respectively, containing the r columns the correspond to the r non-zero singular values, and Σ_r is an $r \times r$ diagonal matrix with σ_i's as its diagonal entries. Then, we can have,

$$P = [V_r \big| O_1] \text{ and } Q = [U_r \big| O_2]^H,$$

where O_1 and O_2 are $n \times n - r$ and $m \times m - r$ all-zero matrices. And, so the structure of $\mathbf{x}$ can be $[\mathbf{x}(1)\ \mathbf{x}(2)\ ...\ \mathbf{x}(r)\ 0\ ...\ 0]^T$. It can be verified that even for this design, we obtain the same $\mathbf{y}$ as mentioned above.

Note: The r entries cannot be at any random locations in $\mathbf{x}$. It should be those that correspond to the right singular vectors V_i that are associated with non-zero singular values σ_i's. Also, the above structure of P, Q and $\mathbf{x}$ is not unique. Any permutations of columns of P and Q are allowed, however, the r data points in $\mathbf{x}$ as to be permuted by the same way as the columns in P are. Similar arrangement relationship hold for $\mathbf{y}$ and Q.

(c) With the design of P and Q discussed in part (b), we have,

$$\mathbf{y}(i) = \sigma_i \mathbf{x}(i) + \mathbf{w}(i),\ i = 1, ..., r.$$

As a result, we get,

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{LS}}(i) &= \frac{\mathbf{y}(i)}{\sigma_i} = \frac{1}{\sigma_i}\Big(\sigma_i \mathbf{x}(i) + \mathbf{w}(i)\Big) \\ &= \mathbf{x}(i) + \frac{\mathbf{w}(i)}{\sigma_i},\ i = 1, ..., r.\end{aligned}$$

Since for every $i \in \mathcal{S}$, $0 \approx \sigma_i << \mathbf{w}(i) \Rightarrow \frac{\mathbf{w}(i)}{\sigma_i} >> \mathbf{x}(i)$, there is a chance that, $\hat{\mathbf{x}}_{\mathrm{LS}}(i) \approx \frac{\mathbf{w}(i)}{\sigma_i}$. So, none of these data points will be decoded correctly and the error will be huge. To have less error, it is suggested not to send any data points along the singular values σ_i^2, $\forall i \in \mathcal{S}$ as $\mathcal{S}$ is known to Tx and Rx.

(d) Given: $\hat{H}_i = U_r \hat{\Sigma}_i V_r^H$, where $\hat{\Sigma}_i$, $\forall i = 1, 2$ are $r \times r$ diagonal matrices with α_i's and β_i's being the diagonal elements for $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ respectively.

For **case 1**, since only $\hat{H}_1$ is to be used and $\alpha_1 = 0$, the singular vectors corresponding to it are not useful. Hence, matrices P and Q can be,

$$P = [\mathbf{v}_2\ \mathbf{v}_3\ ... \mathbf{v}_r \big| O_3] \text{ and } Q = [\mathbf{u}_2\ ...\ \mathbf{u}_r \big| O_4]^H,$$

where O_3 is an $n \times n - (r-1)$ all zero matrix and O_4 is an $m \times m - (r-1)$ all zero matrix. Further, $\mathbf{v}_i$ and $\mathbf{u}_i$ are i^{th} columns of V_r and U_r respectively. Even though Tx and Rx know about rank of H, if they have to design P and Q only using $\hat{H}_1$, since $\alpha_1 = 0$, Tx and Rx has no option other than not to use $\mathbf{v}_1$ and $\mathbf{u}_1$. As a result only $r - 1$ data points can be successfully transmitted and estimated in this case. And, the scalars a_i's are: $a_i = \frac{1}{\alpha_i}$, $\forall i = 2, ..., r$.

Similarly in **case 2**, we design P and Q only using $\mathbf{v}_1, ..., \mathbf{v}_{r-1}$ and $\mathbf{u}_1, ..., \mathbf{u}_{r-1}$ as $\beta_r = 0$. Even in this case, only $r - 1$ data points can be transmitted and estimated. In this case, $a_i = \frac{1}{\beta_i}$, $\forall i = 1, ..., r - 1$.

In **case 3**, since both the estimates of the channels can be used, P and Q can be the same matrices as used in part (b). r data points can be successfully transmitted and estimated. And, the scalars a_i's will be: $a_1 = \alpha_1$, $a_r = \beta_r$ and $a_i = \alpha_i$, $\forall i = 2, ..., r-1$.

3. Suppose the factorization below is an SVD of a matrix A with the entries in U and V rounded to two decimal places.

$$A = \begin{bmatrix} 0.40 & -0.78 & 0.47 \\ 0.37 & -0.33 & -0.87 \\ -0.84 & -0.52 & -0.16 \end{bmatrix} \begin{bmatrix} 7.10 & 0 & 0 \\ 0 & 3.10 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.30 & -0.51 & -0.81 \\ 0.76 & -0.64 & -0.12 \\ 0.58 & -0.58 & 0.58 \end{bmatrix}$$

(a) What is the rank of A?

(b) Use this decomposition of A, with no calculations, to write a basis for $C(A)$, the column space of A, and a basis for $N(A)$, the null space of A.

(c) Repeat parts (a) and (b) for the matrix B

$$B = \begin{bmatrix} -0.86 & -0.11 & -0.50 \\ 0.31 & 0.68 & -0.67 \\ 0.41 & -0.73 & -0.55 \end{bmatrix} \begin{bmatrix} 12.48 & 0 & 0 & 0 \\ 0 & 6.34 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.66 & -0.03 & -0.35 & 0.66 \\ -0.13 & -0.90 & -0.39 & -0.13 \\ 0.65 & 0.08 & -0.16 & -0.73 \\ -0.34 & 0.42 & -0.84 & -0.08 \end{bmatrix}$$

Solution: Added by Prajosh.

(a) A has two non-singlular values. So, rank of $A = 2$.

(b) u_1, u_2=

$$\begin{bmatrix} 0.40 \\ 0.37 \\ -0.84 \end{bmatrix}, \begin{bmatrix} -0.78 \\ -0.33 \\ -0.52 \end{bmatrix}$$

is a basis for col A. and

$$v_3 = \begin{bmatrix} 0.58 \\ -0.58 \\ 0.58 \end{bmatrix}$$

is a basis for Null A.

(c) Rank=2
u_1, u_2=

$$\begin{bmatrix} -0.86 \\ 0.31 \\ 0.41 \end{bmatrix}, \begin{bmatrix} -0.11 \\ 0.68 \\ -0.73 \end{bmatrix}$$

is a basis for col A.
v_3, v_4=

$$\begin{bmatrix} 0.65 \\ 0.08 \\ -0.16 \\ -0.73 \end{bmatrix}, \begin{bmatrix} -0.34 \\ 0.42 \\ -0.84 \\ -0.08 \end{bmatrix}$$

is a basis for Null A.

4. A is an $m \times n$ matrix with singular value decomposition $A = U\Sigma V^T$, where U is an $m \times m$ orthonormal matrix,Σ is an $m \times n$ diagonal matrix with r positive entries, and V is an $n \times n$ orthonormal matrix. Justify the following
   (a) Show that if A is square, then $|det(A)|$ is the product of the singular values of A.
   (b) Show that if P is an orthonormal $m \times m$ matrix, then PA has the same singular values of A.
   (c) Justify that the second singular value of a matrix A is the maximum of $||Ax||$ as x varies overall unit vectors orthonormal to v_1, with v_1 a right singular vector corresponding to first singular value of A.

Solution: Added by Prajosh.

(a) Determinant of an orthonormal matrix is $+1$ or -1.

$$1 = det(I) = det(U^T U) = det(U^T)det(U) = (det(U))^2$$

. Suppose that A is square and $A = U\Sigma V^T$. Then Σ is square, and

$$det(A) = det(U)det(\Sigma)det(V^T) = \pm det\Sigma = \pm\sigma_1.....\sigma_n$$

(b) Let $A = U\Sigma V^T$. If P and U are orthonormal, the matrix PU is also orthonormal. Because,
if P and U are orthonormal and each is invertible, PU is invertible and $(PU)^{-1} = U^{-1}P^{-1} = U^T P^T = (PU)^T$. So, the equation $PA = (PU)\Sigma V^T$ has the form required for a singular value decomposition. The diagonal entries of in Σ are the singular values of PA.

(c) The right singular vector v_1 is an eigenvector for the largest eigenvector λ_1 of $A^T A$. The second largest eigenvalue λ_2 is the maximum of $x^T(A^T A)x$ overall unit vectors orthogonal to v_1. Since $x^T(A^T A)x = ||Ax||^2$, the square root of λ_2, which is the second largest singular value of A, is the maximum of $||Ax||$ over all unit vectors orthogonal to v_1.

5. Prove that a symmetric positive definite matrix has a unique symmetric positive definite square root.

Solution: Added by siva
Let A be the given $n \times n$ symmetric positive definite matrix.
(i) Let's first prove that there exists a positive definite B such that $B^2 = A$.
Let $\lambda_1, \lambda_2 \ldots \lambda_n$ be the eigen values of A. Hence the eigen decomposition of A is $A = SDS^T$. Now, consider the matrix $B = SD'S^T$ where D' is diagonal and it's entries are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}$. Clearly $B^2 = A$ and B is symmetric positive definite.
(ii) Proof for uniqueness of B:
Let's assume a symmetric positive definite matrix with eigen decomposition $C = PTP^T$ such that $C^2 = A$. Let the eigen values of C be $\mu_1, \mu_2, \ldots, \mu_n$.
Consider the matrix $P^T AP$. Then, $P^T AP = P^T C^2 P = (P^T CP)^2 = T^2$. The matrix P diagonalizes A. Hence the entries of T must be the eigen values of A, i.e.,

$\mu_i^2 = \lambda_i \implies \mu_i = \sqrt{\lambda_i} \implies T = D'$.
We have $B^2 = A = C^2$

$$\begin{aligned}(SD'S^T)^2 &= (PD'P^T)^2\\ SD'^2S^T &= PD'^2P^T\\ (P^TS)D'^2 &= D'^2(P^TS)\\ QD'^2 &= D'^2Q \quad \text{ where } Q = P^TS\end{aligned}$$

Assuming all the eigen values of A are distinct, Q must be diagonal. (If the elements of D'^2 were not distinct, Q will not necessarily be diagonal but will be block diagonal of appropriate sizes. Even the following proof will hold. But for better understanding we will stick to distinct eigen values.)
We also have $QD' = D'Q \implies D' = Q^TD'Q$.
Consider the matrix B. $B = SD'S^T = S(Q^TD'Q)S^T = SS^TPD'P^TSS^T = PD'P^T = C$.
Hence proved.

6. The symbol $A \succ B$ means $A - B$ is positive definite, $A \prec B$ means $B - A$ is positve definite. Consider a symmetric positve definite matrix H.
   (a) Prove $mI \prec H \prec MI$ if and only if eigen values of H are bounded between m and M, where I is the identity matrix and $M > m > 0$.
   (b) Prove that the diagonal elements of H cannot be non-positive.
   (c) Given a symmetric matrix G, find appropriate m, M such that $mI \preceq G \preceq MI$.

Solution: Added by Siva Note that $A \prec B \iff \mathbf{x}^TA\mathbf{x} < \mathbf{x}^TB\mathbf{x} \ \ \forall \mathbf{x}$

(a) (i) Given $mI \prec H \prec MI$

$$\begin{aligned}&\implies \quad \mathbf{x}^TmI\mathbf{x} < \mathbf{x^T}H\mathbf{x} < \mathbf{x}^TMI\mathbf{x} \ \ \forall \mathbf{x}\\ &\implies m \parallel \mathbf{x} \parallel_2^2 < \mathbf{x}^TH\mathbf{x} < M \parallel \mathbf{x} \parallel_2^2 \ \ \forall \mathbf{x}\end{aligned}$$

Since the inequality is true for all $\mathbf{x}$, let's consider the unit eigen vector, $\mathbf{v}$ of H whose corresponding eigen value is λ

$$\begin{aligned}&\implies m \parallel \mathbf{v} \parallel_2^2 < \lambda \parallel \mathbf{v} \parallel_2^2 < M \parallel \mathbf{v} \parallel_2^2\\ &\implies \qquad m < \lambda < M\end{aligned}$$

(ii) Given $m < \lambda < M$
Since H is symmetric, any vector $\mathbf{x}$ can be written as a linear combination of the eigen vectors of H i.e., $\mathbf{x} = \sum_{i=1}^n c_i\mathbf{v_i}$, where $\mathbf{v_i}$ are the eigen vectors and n is the size of matrix.
Consider

$$\begin{aligned}\mathbf{x}^TH\mathbf{x} &= (\sum_{i=1}^n c_i\mathbf{v_i})^TH(\sum_{i=1}^n c_i\mathbf{v_i}) = (\sum_{i=1}^n c_i\mathbf{v_i})^T(\sum_{i=1}^n c_i\lambda_i\mathbf{v_i})\\ &= \sum_{i=1}^n c_i^2\lambda_i \ \ \forall \mathbf{x} \quad \text{(Since } \{\mathbf{v_i}\} \text{ is an orthogonal set)}\end{aligned}$$

Since eigen values are bounded, we have the following

$$
\begin{aligned}
& \sum_{i=1}^{n} c_i^2 m < \sum_{i=1}^{n} c_i^2 \lambda_i < \sum_{i=1}^{n} c_i^2 M \\
\implies & \sum_{i=1}^{n} c_i \mathbf{v_i}^T (m) \sum_{i=1}^{n} c_i \mathbf{v_i} < \sum_{i=1}^{n} c_i^2 \lambda_i < \sum_{i=1}^{n} c_i \mathbf{v_i}^T (M) \sum_{i=1}^{n} c_i \mathbf{v_i} \\
\implies & \mathbf{x}^T (mI) \mathbf{x} < \mathbf{x}^T H \mathbf{x} < \mathbf{x}^T (MI) \mathbf{x} \;\; \forall \mathbf{x} \\
\implies & mI \prec H \prec MI
\end{aligned}
$$

(b) Since H is positive definite, we have $\mathbf{x}^T H \mathbf{x} > 0 \;\; \forall \mathbf{x}$. Choosing $\mathbf{x}$ to be $\mathbf{e_i}$ which is a vector with 1 in it's i^{th} position and zero elsewhere, we get $\mathbf{e_i}^T H \mathbf{e_i} > 0 \implies H_{ii} > 0$

(c) We have $mI \preceq G \preceq MI$, where we intend to find m and M. Let λ_{min} and λ_{max} are the minimum and maximum eigen values of G respectively.
Consider $\mathbf{x}^T G \mathbf{x} = \sum_{i=1}^{n} c_i^2 \lambda_i$. This can be bounded as follows

$$
\begin{aligned}
& \sum_{i=1}^{n} c_i^2 \lambda_{min} \le \sum_{i=1}^{n} c_i^2 \lambda_i \le \sum_{i=1}^{n} c_i^2 \lambda_{max} \\
\implies & \mathbf{x}^T (\lambda_{min} I) \mathbf{x} \le \mathbf{x}^T G \mathbf{x} \le \mathbf{x}^T (\lambda_{max} I) \mathbf{x} \;\; \forall \mathbf{x} \\
\implies & \lambda_{min} I \prec G \prec \lambda_{max} I
\end{aligned}
$$

Hence m and M can be λ_{min} and λ_{max} respectively.

7. Decide between a minimum, maximum, or saddle point for the following functions, if they are stationary points.

   (a) $F = -1 + 4(e^x - x) - 5x\sin(y) + 6y^2$ at the point $x = y = 0$.

   (b) $F = (x^2 - 2x)\cos(y)$, at the point $x = 1, y = \pi$.

   (c) $F = \frac{1}{4}x^4 + x^2 y + y^2$ at the point $x = 1, y = 2$.

Hint: Use the second derivative matrices.

Solution: Gradient of the function at the stationary point is zero. For minima, maxima and saddle points the eigenvalues of second derivative matrices are real and positive, real and negative, and real and non-zero respectively; along with being stationary.

(a)

$$\nabla F = \begin{bmatrix} 4e^x - 4 - 5\sin(y) \\ -5x\cos(y) + 12y \end{bmatrix} \Rightarrow \nabla F_{(0,0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

So, It is a stationary point.

$$\nabla^2 F = \begin{bmatrix} 4e^x & -5\cos(y) \\ -5\cos(y) & 5x\sin(y) + 12 \end{bmatrix} \Rightarrow \nabla^2 F = \begin{bmatrix} 4 & -5 \\ -5 & 12 \end{bmatrix}$$

So, eigenvalues are 1.59,14.4. So, It is a minima.

(b)

$$\nabla F = \begin{bmatrix} (2x-2)\cos(y) \\ (x^2-2x)(-\sin(y)) \end{bmatrix} \Rightarrow \nabla F_{(1,\pi)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

So, It is a stationary point.

$$\nabla^2 F = \begin{bmatrix} 2\cos(y) & (2x-2)(-sin(y)) \\ (2x-2)(-sin(y)) & (x^2-2x)(-\cos(y)) \end{bmatrix} \Rightarrow \nabla^2 F = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

So, eigenvalues are -1,-2. So, It is a maxima.

(c)

$$\nabla F = \begin{bmatrix} x^3+2xy \\ x^2+2y \end{bmatrix} \Rightarrow \nabla F_{(1,2)} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \neq 0$$

So, It is not a stationary point as the gradient is not equal to zero.

8. (a) Solve the generalized eigenvalue problem i.e, $Ax = \lambda Bx$ for eigenvectors

$$\begin{bmatrix} 4 & 3 \\ 3 & 7 \end{bmatrix} x = \lambda \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} x$$

Hint: Solve $|A - \lambda B| = 0$ for eigenvalues and substitute them in $Ax = \lambda Bx$ for eigenvectors. Also, If B is invertible. Then multiplying on both sides by B^{-1} gives $B^{-1}Ax = \lambda x$.

(b) Solve the generalized eigenvector problem, i.e., $(A - \lambda I)^P x = 0$ for

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

Theory of generalized eigenvectors: $|A - \lambda I| = 0$ has to be solved for eigenvalues. For eigenvectors, you atmost need to solve for $(A - \lambda I)^k x = 0$, where k is the algebraic multiplicity of A. You start with P=1 and increment the P by 1, till you get all the eigenvectors.

Solution:

(a) $|A - \lambda B| = 0 \Rightarrow (4-\lambda)(7-\lambda) - (3-2\lambda)(3-2\lambda) = 0$

$$\Rightarrow 28 - 11\lambda + \lambda^2 - (9 - 12\lambda + 4\lambda^2) = 0 \Rightarrow 3\lambda^2 - \lambda - 19 = 0$$

Eigenvalues are: $1 \pm \frac{\sqrt{229}}{6}$.

Eigenvectors are $\begin{bmatrix} -1 \\ -0.55 \end{bmatrix}, \begin{bmatrix} 1 \\ -0.82 \end{bmatrix}$ respectively.

(b) For eigenvalues, $|A - \lambda I| = 0$, implies $\lambda = 1, 1, 3$.

For eigenvectors,

$$\lambda = 3, \quad (A - 3I)v_1 = 0, \Rightarrow v_1 = [1\,2\,2]^T$$

$$\lambda = 1, \quad (A - 1I)v_2 = 0, \Rightarrow v_2 = [1\,0\,0]^T$$

Now for 3rd eigenvector corresponding to $\lambda = 1$,

$$(A - 1I)^2 v_3 = 0, (A - 1I)v_3 \neq 0, \, and \, (A - 1I)v_3 = v_2 \Rightarrow x = [0\,1\,0]^T$$