Linear Algebra

Dr. Aashish Mathur
Associate Professor
Department of Electrical Engineering
Indian Institute of Technology Jodhpur

Study Material

- G. Strang, *Linear Algebra and Its Applications*

- Gilbert Strang lectures on Linear Algebra (MIT)

https://www.youtube.com/watch?v=QVKj3LADCnA&list=PL49CF3715CB9EF31D

- Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra*

- Seymour Lipschutz and Marc Lipson, *Linear Algebra*, Schaum's Outlines

Mathematical Foundations

- $x \rightarrow a$: x approaches a
- $t \leftarrow t + 1$: in an algorithm – Assign to variable t the new value $t + 1$
- $\text{argmax}_x\, f(x)$: The value of x that leads to the maximum value of $f(x)$
- $\text{argmin}_x\, f(x)$: The value of x that leads to the minimum value of $f(x)$
- $m \bmod n$: $m\ modulo\ n$ – that is, the remainder when m is divided by n (e.g.: 7 mod 5 =2)
- $\ln(x)$: logarithm base e, or natural logarithm of x
- $\log(x)$: logarithm base 10 of x
- $\log_2(x)$: logarithm base 2 of x
- $\mathbf{1}_N$: Vector of length N solely of 1's
- $\text{diag}(a_1, a_2, \ldots, a_N)$: Matrix whose diagonal elements are $a_1, a_2, \ldots, a_N$ and off-diagonal elements are 0
- $||\boldsymbol{x}||$: Euclidean norm of vector $\boldsymbol{x}$
- $|\mathbf{A}|$: Determinant of $\boldsymbol{A}$ ($\boldsymbol{A}$ is a matrix)
- $|D|$: Cardinality of set D – i.e., the number of (possible distinct) discrete elements in it

$\boldsymbol{x}$: lower case bold font represents vectors and $\boldsymbol{A}$: uppercase bold font represents matrix

- Calligraphic font generally denotes sets or lists

$$D = \{x_1, x_2, \ldots\ldots, x_n\}$$

- Scalars

$s \in \mathbb{R}$, defining a real-valued scalar

$n \in \mathbb{N}$, defining a natural number scalar

- Vectors: A n −dimensional vector is assumed to be a column vector. Vector is denoted by bold font $\boldsymbol{x}$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad \boldsymbol{x} \in \mathbb{R}^n$$

- Matrices: It is a two-dimensional array of numbers $- \boldsymbol{A} \in \mathbb{R}^{m \times n}$

$$\boldsymbol{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \ldots\ldots\ldots & a_{1n} \\ a_{21} & & & a_{2n} \\ \vdots & & & \vdots \\ & & \ldots\ldots\ldots & \\ a_{m1} & & & a_{mn} \end{bmatrix}$$

- Complex conjugate transpose (Hermitian)
  - $\boldsymbol{x}^{\mathrm{H}} = (\boldsymbol{x}^{\mathrm{T}})^{*} = [x_1{}^{*}, x_2{}^{*}, \ldots\ldots, x_n{}^{*}]$
  - $\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{T}})^{*} = (\boldsymbol{A}^{*})^{\mathrm{T}}$

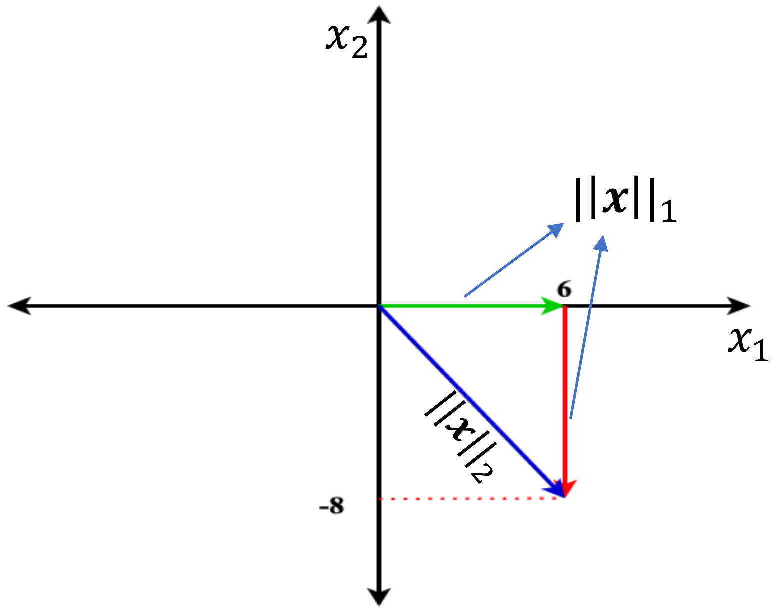
- If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$, then it is called Hermitian Matrix
  - Example: If $\boldsymbol{A} = \begin{bmatrix} 2 & 1+2\mathrm{j} \\ 1-2\mathrm{j} & 3 \end{bmatrix}$ then $\boldsymbol{A}^{\mathrm{H}} = \begin{bmatrix} 2 & 1+2\mathrm{j} \\ 1-2\mathrm{j} & 3 \end{bmatrix} = \boldsymbol{A}$

➢ Vector norms

- $||\boldsymbol{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad p = 1,2,\ldots$
- 1−norm: $||\boldsymbol{x}||_1 = \sum_{i=1}^{n} |x_i|$
- Euclidean (2−norm): $||\boldsymbol{x}||_2 = \left(\sum_{i=1}^{n} x_i^* x_i\right)^{1/2} = (\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})^{1/2} = \sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}$
- ∞ −norm: $||\boldsymbol{x}||_\infty = \max_i |x_i|$

➢ Examples – consider a vector $\boldsymbol{x} = \begin{bmatrix} 6 \\ -8 \end{bmatrix}$

- 1−norm: $||\boldsymbol{x}||_1 = |6| + |-8| = 14$
- 2−norm : $||\boldsymbol{x}||_2 = \sqrt{|6|^2 + |-8|^2} = \sqrt{100} = 10$
- ∞ −norm: $||\boldsymbol{x}||_\infty = \max_i\{|6|, |-8|\} = 8$



➢ Vector norms

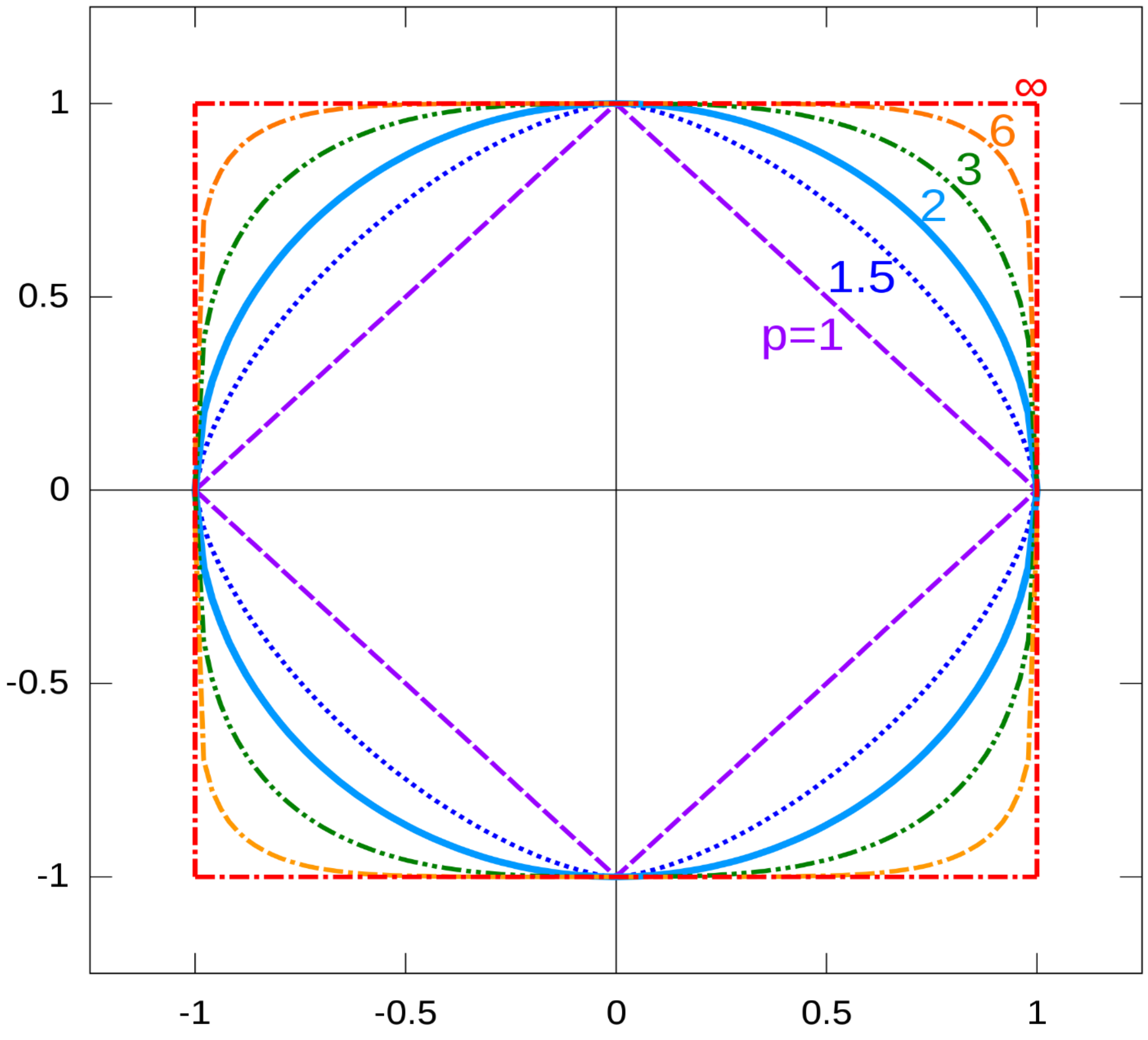
➢ Example- consider a vector $\boldsymbol{x} = \begin{bmatrix} 3 \\ -2 \\ -i \end{bmatrix}$

- $1-$norm: $||\boldsymbol{x}||_1 = \ |3| + |-2| + |-i| = 5 + 1$

- $2-$norm: $||\boldsymbol{x}||_2 = \sqrt{|3|^2 + |-2|^2 + |-i|^2} \ = \sqrt{9 + 4 + 1} \ = \sqrt{13}$

- $\infty -$norm: $||\boldsymbol{x}||_\infty = \ \max_i\{|3|, |-2|, |-i|\} = 3$

➢Pictorial representation of all the L_p- norms



➢Inner product

- $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{n} x_i\, y_i$
- $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = \sum_{i=1}^{n} x_i^*\, y_i$
- Example: $\boldsymbol{x} = (1, \mathrm{i})$, $\mathbf{y} = (3, 1)$

  Then inner product of $< \boldsymbol{x}, \boldsymbol{y} > = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = 1 \times 3 + i^* \times 1 = 3 + (-i) = 3 - i$

- Example: $\boldsymbol{x} = (1, 1, 3)$, $\boldsymbol{y} = (3, 1, 1)$

  Then inner product of $< \boldsymbol{x}, \boldsymbol{y} > = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 1 \times 3 + 1 \times 1 + 3 \times 1 = 7$

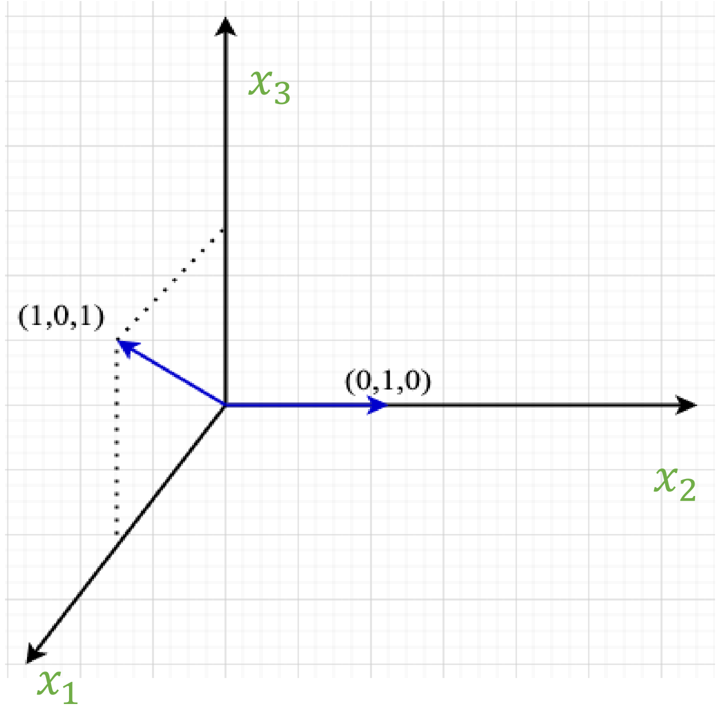
- We call a vector normalized if $||\boldsymbol{x}|| = 1$
- The angle θ between two n-dimensional vectors obeys

$$\cos\theta = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{||\boldsymbol{x}||\,||\boldsymbol{y}||}, \qquad |\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leq ||\mathbf{x}||\,||\mathbf{y}|| \quad \text{(Cauchy-Schwarz inequality)}$$

- $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0 \Rightarrow$ Vectors are orthogonal (perpendicular to each other)
  - Example: $\boldsymbol{x} = (1, 0, 1)$ $\boldsymbol{y} = (0, 1, 0)$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 1 \times 0 + 0 \times 1 + 1 \times 0 = 0$$

(1,0,1) and (0,1,0) are orthogonal to each other



- If the vectors are orthogonal and have unit norm=> orthonormal
- Outer Product
  - $$\boldsymbol{M} = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 \dots\dots & y_m \end{bmatrix} = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_m \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_m \end{bmatrix}$$
  - Example: $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$, then $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

- Frobenius norm: $||\boldsymbol{A}||_{\mathrm{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = \sqrt{\mathrm{trace}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}$

  - Example: $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & i \end{bmatrix}$. Then $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$

    $$||\boldsymbol{A}||_{\mathrm{F}} = \sqrt{\mathrm{trace}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})} = \sqrt{3}$$

- Linear independence

  - A set of vectors $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots ., \boldsymbol{x_n}\}$ is linearly independent if no vector in the set can be written as a linear combination of any of the others

    $$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ \ldots\ldots\ldots + c_n\boldsymbol{x}_n = 0 \ \Rightarrow \qquad c_1 = c_2 = \cdots = c_n = 0$$

    - Example: $\begin{bmatrix} 9 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 9 \end{bmatrix}$ are linearly independent since they are not multiples of each other and cannot be 0 other than c_1 and $c_2 = 0$

➢ Rank

- The rank of $\boldsymbol{A}$ is the number of independent rows or columns of $\boldsymbol{A}$

- The ranks of $\boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$, and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ are the same

- If $\boldsymbol{A}$ is square and full rank, there is unique inverse $\boldsymbol{A}^{-1}$ such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & \dots & \dots \dots & 1 \end{bmatrix}$$

- An $n \times n$ matrix $\boldsymbol{A}$ has rank n, then $\boldsymbol{A}$ is invertible $\Rightarrow$ $\det(\boldsymbol{A}) \neq 0$

- Example: $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 0 & 5 \end{bmatrix}$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -2 \\ 0 & -6 & -4 \end{bmatrix} \qquad R_2 \to R_2 - 2R_1\,, R_3 \to R_3 - 3R_1$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -2 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_3 \to R_3 - 2R_2$$

Echelon form: Number of non-zero rows = 2. Hence Rank = 2

➢ Unitary

- A square matrix $\boldsymbol{u}$ is called unitary if $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u} = \boldsymbol{I}$ and $\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}} = \boldsymbol{I}$

- $||\boldsymbol{u}|| = 1$, its rows and columns are orthonormal

- Example: $\boldsymbol{A} = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}$ is a Unitary matrix?

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1-i & 1+i \\ 1+i & 1-i \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \boldsymbol{I}_2$$

➢ Projections

- We want to find the projection point $\boldsymbol{p}$
- The point $\boldsymbol{p}$ must be some multiple $\boldsymbol{p} = \hat{x}\,\boldsymbol{a}$ of the given vector $\boldsymbol{a}$
- Every point on the line is a multiple of $\boldsymbol{a}$
- The problem is to compute the coefficient $\hat{x}$

$$(\boldsymbol{b} - \boldsymbol{p}) \perp \boldsymbol{a}$$

$$\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{b} - \hat{x}\boldsymbol{a}) = 0$$

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} - \hat{x}\,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 0$$

$$\hat{x} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$$

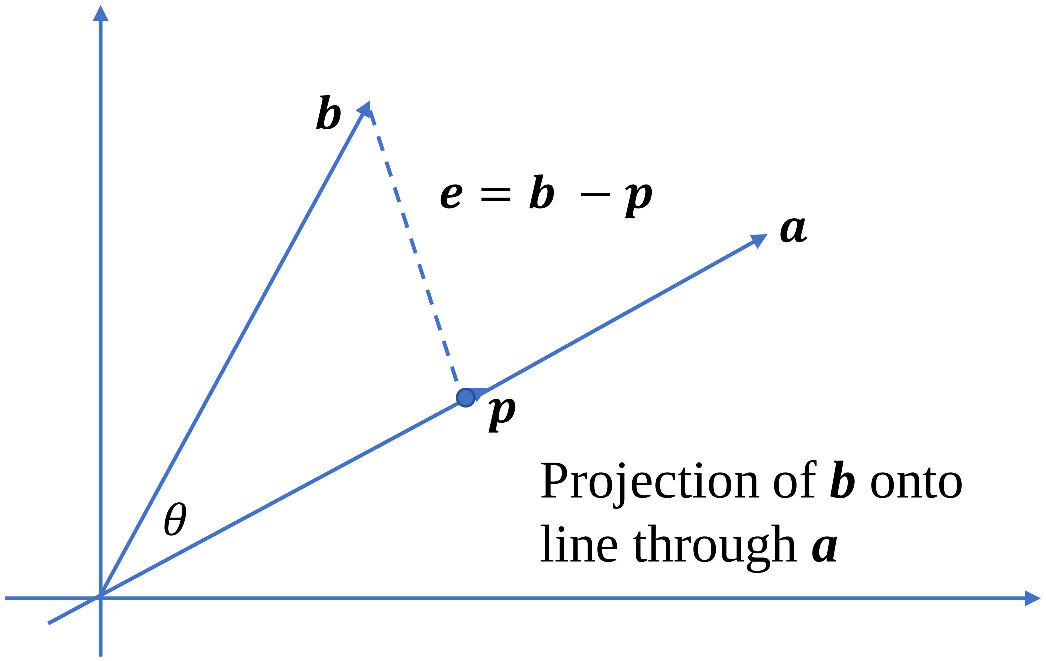


Fig: The projection $\boldsymbol{p}$ is the point (on the line through $\boldsymbol{a}$) closest to $\boldsymbol{b}$

- The projection of the vector $\boldsymbol{b}$ onto the line in the direction of $\boldsymbol{a}$ is $\boldsymbol{p} = \hat{x}\,\boldsymbol{a}$

$$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} \qquad \text{Projection onto a line}$$

- The vector $\boldsymbol{a}$ is put before the number: $\hat{x} = \dfrac{\boldsymbol{a}^{\mathrm{T}}\, \boldsymbol{b}}{\boldsymbol{a}^{T} \boldsymbol{a}}$

- Projection onto a line is carried out by a projection matrix $\boldsymbol{P}$

  So, the projection matrix is: $\boldsymbol{P} = \dfrac{\boldsymbol{a}\, \boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{T} \boldsymbol{a}}$

- $\boldsymbol{P}$ is the matrix that multiplies $\boldsymbol{b}$ and produces $\boldsymbol{p}$

Properties

1. **P** is a symmetric matrix

2. Its square is itself: $\boldsymbol{P}^2 = \boldsymbol{P}$

- Example: Find the projection of the vector $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ onto the line in the direction of $\boldsymbol{a} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$?

Solution: $\hat{x} = \dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = \dfrac{\begin{bmatrix} -4 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} -4 & 2 \end{bmatrix}\begin{bmatrix} -4 \\ 2 \end{bmatrix}} = \dfrac{-4}{20} = -\dfrac{1}{5}$

- The projection of the vector $\boldsymbol{b}$ onto the line in the direction of $\boldsymbol{a}$ is $\boldsymbol{p} = \hat{x}\,\boldsymbol{a} = -\frac{1}{5}\begin{bmatrix} -4 \\ 2 \end{bmatrix}$

➢Eigenvectors and Eigenvalues

$$\boldsymbol{M}\boldsymbol{x} = \lambda\boldsymbol{x}$$

$$(\boldsymbol{M} - \lambda\boldsymbol{I})\boldsymbol{x} = \mathbf{0}$$

- λ that makes $\boldsymbol{M} - \lambda\boldsymbol{I}$ singular is called an eigenvalue
- Solution vector $\boldsymbol{x} = \boldsymbol{e}_i$ and corresponding scalar $\lambda = \lambda_i$ are called the eigenvector and associated eigenvalue, respectively
- Characteristic equation: $|\boldsymbol{M} - \lambda\boldsymbol{I}| = \mathbf{0}$

➢Properties

- Trace$[\boldsymbol{M}] = \sum_{i=1}^{d} \lambda_i$
- Det$(\boldsymbol{M}) = |\boldsymbol{M}| = \prod_{i=1}^{d} \lambda_i$
- If a matrix is diagonal, then its eigenvalues are simply the non-zero entries on the diagonal, and the eigenvectors are the unit vectors parallel to the coordinate axes

➢Eigenvalue decomposition

$$\boldsymbol{M}[\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots.] = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots.]\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \ddots\end{bmatrix}$$

$\boldsymbol{MT} = \boldsymbol{T}\wedge \Rightarrow \boldsymbol{M} = \boldsymbol{T}\wedge\boldsymbol{T}^{-1}$ exist when $\boldsymbol{T}$ is invertible and when eigenvalues are distinct

- Example: Find the eigenvalue decomposition of matrix $\boldsymbol{M} = \begin{bmatrix}5 & 3\\ 3 & 5\end{bmatrix}$

Eigenvalue decomposition of matrix $\boldsymbol{M}$

- $\det(\boldsymbol{M} - \lambda\boldsymbol{I}) = \lambda^2 - 10\lambda + 25\ - 9 = \lambda^2 - 10\lambda + 16 = (\lambda - 2)(\lambda - 8) = 0$

$\lambda = 2, 8$ so $\wedge = \begin{bmatrix}2 & 0\\ 0 & 8\end{bmatrix}$

- For $\lambda = 2$, the eigenvector $(\boldsymbol{M} - 2\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \rightarrow \begin{bmatrix}3 & 3\\ 3 & 3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \boldsymbol{0}$

$3x_1 + 3x_2 = 0 \quad \rightarrow \quad x_1 = -x_2$

Eigenvector is given by $\boldsymbol{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

- For $\lambda = 8$ eigenvector $(\boldsymbol{M} - 8\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \rightarrow \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0}$

$$-3x_1 + 3x_2 = 0 \quad \rightarrow \quad x_1 = x_2$$

Eigenvector given by $x = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

- $\boldsymbol{T} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$
- $\boldsymbol{M} = \boldsymbol{T} \wedge \boldsymbol{T}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

➢Singular value decomposition

- For any matrix $\boldsymbol{M}$**,** there is a decomposition: $\boldsymbol{M} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\mathrm{H}}$
- $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices
- $\boldsymbol{\Sigma}$ is diagonal with positive real entries
- Columns $\boldsymbol{u}_i$ of $\boldsymbol{U}$ are called the left singular vectors
- Columns $\boldsymbol{v}_i$ of $\boldsymbol{V}$ are called the right singular vectors
- Diagonal entries σ_i of $\boldsymbol{\Sigma}$ are called the singular values. They are positive, real, and sorted

$$\sigma_1 \geq \sigma_2 \geq \ldots .\, 0; \quad \sigma_1 = \text{largest singular value}$$

- If $\boldsymbol{M}$ is an $m \times n$ matrix then,

$\boldsymbol{U}: m \times m$ $\qquad$ $\boldsymbol{\Sigma}: m \times n$ $\qquad$ $\boldsymbol{V}: \mathrm{n} \times n$

- Rank $(\boldsymbol{M}) =$ The number of non-zero singular values

➢ Steps for getting SVD for a Matrix

1. The ordering of the vectors comes from the ordering of the singular values (largest to smallest)

2. The columns of U are the eigenvectors of $\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}}$

3. The columns of V are the eigenvectors of $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$

4. The diagonal elements of $\boldsymbol{\Sigma}$ are the singular values $\sigma_i = \sqrt{\lambda_i}$

5. The relationships between $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ (with normalization):

$$\boldsymbol{M}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \qquad\qquad \boldsymbol{M}^{\mathrm{T}}\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i$$

The scaling factor comes from $||\boldsymbol{M}\boldsymbol{v}_i|| = \sigma_i = ||\boldsymbol{M}^{\mathrm{T}}\boldsymbol{u}_i||$

- Example: Find the SVD of $\boldsymbol{A}$, where $\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$?

Solution: First, we'll work with $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$

- The characteristic polynomial is $\det\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} - \lambda \boldsymbol{I}\right) = \lambda^2 - 34\lambda + 225 = (\lambda - 25)(\lambda - 9)$, so eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are $\lambda = 25$ and 9. Singular values are $\sigma_1 = \sqrt{25} = 5, \sigma_2 = \sqrt{9} = 3$,

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$$

- The eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$ are 25, 9, and 0, and since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric we know that the eigenvectors will be orthogonal

- For $\lambda = 25$, reduced matrix is $\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ so, $\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

- for $\lambda = 9$, reduced matrix is $\begin{bmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$ so, $\boldsymbol{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$

- For the last eigenvector, we could compute unit vector perpendicular to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ by

$$\boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{v}_3 = 0, \boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{v}_3 = 0$$

so, for $\lambda = 0$, eigenvector is given by $v_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \end{bmatrix}$

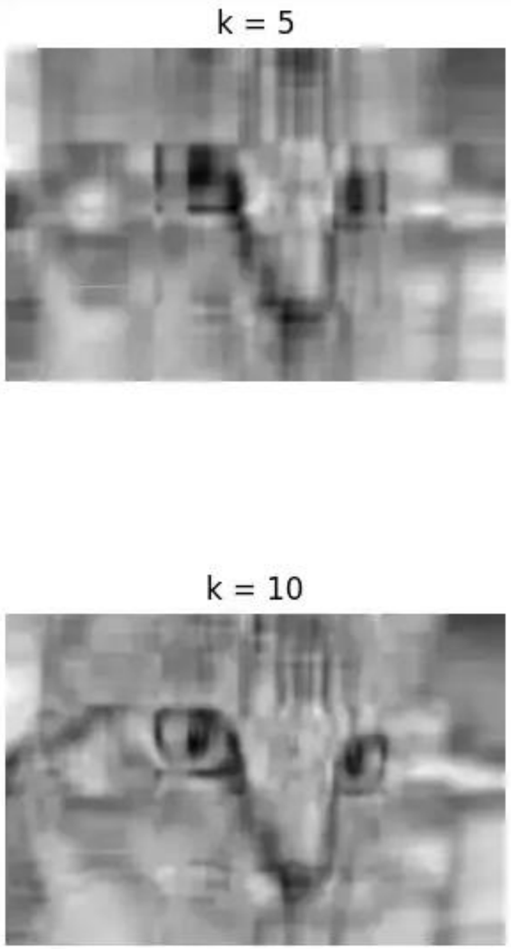
- $V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \end{bmatrix}$

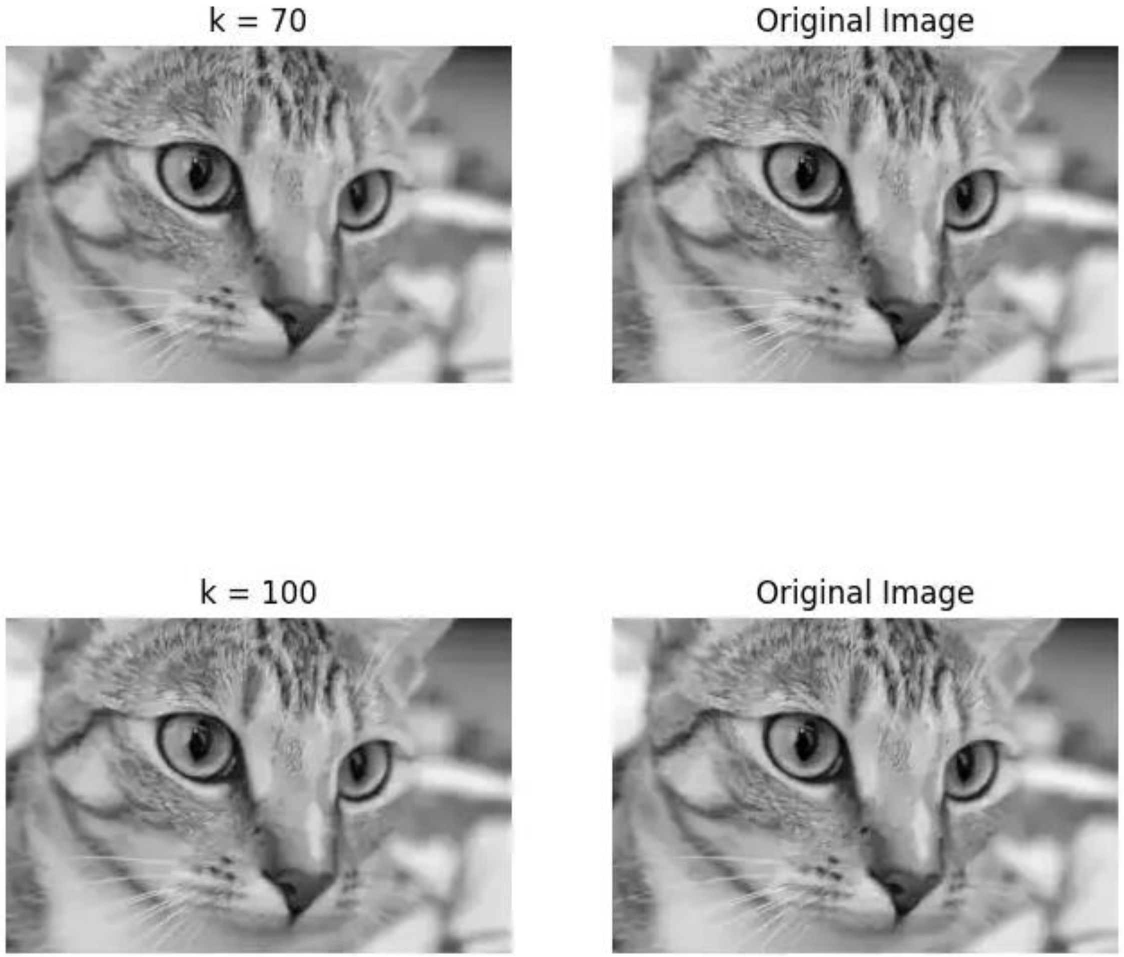
- We have to compute $\mathbf{U}$ by the formula $\sigma_i \mathbf{u}_i = \boldsymbol{A}\boldsymbol{v}_i$, $\boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

- $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}$

```python
from skimage.color import rgb2gray
from skimage import data
import matplotlib.pyplot as plt
import numpy as np
from scipy.linalg import svd


X = np.array([[3, 3, 2], [2, 3, -2]])
print(X)

U, singular, V_transpose = svd(X)

print("U: ", U)
print("Singular array", singular)
print("V^{T}", V_transpose)

singular_inv = 1.0 / singular
s_inv = np.zeros(X.shape)
s_inv[0][0] = singular_inv[0]
s_inv[1][1] = singular_inv[1]
M = np.dot(np.dot(V_transpose.T, s_inv.T), U.T)
print(M)
```

```python
cat = data.chelsea()
plt.imshow(cat)

gray_cat = rgb2gray(cat)

U, S, V_T = svd(gray_cat, full_matrices=False)
S = np.diag(S)
fig, ax = plt.subplots(5, 2, figsize=(8, 20))

curr_fig = 0
for r in [5, 10, 70, 100, 200]:
    cat_approx = U[:, :r] @ S[0:r, :r] @ V_T[:r, :]
    ax[curr_fig][0].imshow(cat_approx, cmap='gray')
    ax[curr_fig][0].set_title("k = " + str(r))
    ax[curr_fig, 0].axis('off')
    ax[curr_fig][1].set_title("Original Image")
    ax[curr_fig][1].imshow(gray_cat, cmap='gray')
    ax[curr_fig, 1].axis('off')
    curr_fig += 1
plt.show()
```


k = 5
Original Image
k = 10
Original Image


k = 70
Original Image
k = 100
Original Image

➢Eigenvalue decomposition and SVD

- SVD of $\boldsymbol{M} = \boldsymbol{U\Sigma V}^{\mathrm{H}}$

- $\boldsymbol{MM}^{\mathrm{H}} = \boldsymbol{U\Sigma}\,\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V\Sigma U}^{\mathrm{H}} = \boldsymbol{U\Sigma}^{2}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U} \wedge \boldsymbol{U}^{\mathrm{H}}$

- The eigenvalues of $\boldsymbol{MM}^{\mathrm{H}}$ are singular values of $\boldsymbol{M}$ squared

- Eigenvectors of $\boldsymbol{MM}^{\mathrm{H}}$ are the left singular vectors of $\boldsymbol{M}$

➢Matrix inversion

- $\boldsymbol{M}$: $d \times d$; $\boldsymbol{M}^{-1}: d \times d$; $\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{I}$

- $\boldsymbol{M}^{-1} = \frac{\mathrm{adj}[\boldsymbol{M}]}{|\boldsymbol{M}|}$; $adj[\boldsymbol{M}]$ = adjoint of $\boldsymbol{M}$

- If $\boldsymbol{M}$ is not square (or if $\boldsymbol{M}^{-1}$ does not exist because the columns of $\boldsymbol{M}$ are not linearly independent), we use pseudoinverse $\boldsymbol{M}^{\dagger}$

- If $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$ is non-singular , then

$$\boldsymbol{M}^{\dagger} = (\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1}\,\boldsymbol{M}^{\mathrm{T}} \quad \text{(or)} \quad \boldsymbol{M}^{\dagger} = (\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\mathrm{H}}$$

- Example: Find inverse of $\boldsymbol{M} = \begin{bmatrix} 2 & -2 \\ -3 & 5 \end{bmatrix}$?

$$\boldsymbol{M}^{-1} = \frac{\mathrm{adj}[\boldsymbol{M}]}{|\boldsymbol{M}|} = \frac{\begin{bmatrix} 5 & 2 \\ 3 & 2 \end{bmatrix}}{4} = \begin{bmatrix} \frac{5}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix}$$

- Example: Find Pseudoinverse of $\boldsymbol{M} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 3 \end{bmatrix}$?

$$(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1} = \begin{bmatrix} 14 & 20 \\ 20 & 34 \end{bmatrix}^{-1} = \begin{bmatrix} 0.447 & -0.263 \\ -0.263 & 0.184 \end{bmatrix}$$

$$\boldsymbol{M}^{\dagger} = (\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1}\,\boldsymbol{M}^{T} = \begin{bmatrix} 0.447 & -0.263 \\ -0.263 & 0.184 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 3 \end{bmatrix} = \begin{bmatrix} -0.342 & -0.158 & 0.553 \\ 0.289 & 0.211 & -0.237 \end{bmatrix}$$

➢Derivative of matrices

- Suppose $f(\boldsymbol{x})$ is a scalar valued function of d variables $x_i, i = 1, 2, \dots, d$ which we represent as the vector $\boldsymbol{x}$

$$\nabla f(\boldsymbol{x}) = \text{grad } f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_d} \end{bmatrix} = \text{Derivative or gradient of } f(\cdot) \text{ with respect to } \boldsymbol{x}$$

- If we have an n −dimensional vector-valued function $\boldsymbol{f}$ of d −dimensional vector $\boldsymbol{x}$

$$\boldsymbol{J}(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_d} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_n(\boldsymbol{x})}{\partial x_d} \end{bmatrix}$$

- If $\boldsymbol{J}(\boldsymbol{x})$ is square, its determinant is called simply the *Jacobian* or *Jacobian determinant*

- Example: $f(\boldsymbol{x}) = x_1 + x_2 + x_3 + x_1x_2x_3$ find the gradient of $f(\boldsymbol{x})$

$$\nabla f(\boldsymbol{x}) = \text{grad } f(\boldsymbol{x}) = \begin{bmatrix} 1 + x_2x_3 \\ 1 + x_1x_3 \\ 1 + x_1x_2 \end{bmatrix}$$

- Example: $\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1x_2x_3 \end{bmatrix}$ find the value of $\boldsymbol{f}'(\boldsymbol{x})$?

$$J(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} 1 & 1 & 1 \\ x_2x_3 & x_1x_3 & x_1x_2 \end{bmatrix}$$

- If the entries of matrix $\boldsymbol{M}$ depend upon a scalar parameter θ, then can take the derivative of $\boldsymbol{M}$ component by component, to get another matrix, as

$$\frac{\partial \boldsymbol{M}}{\partial \theta} = \begin{bmatrix} \frac{\partial m_{11}}{\partial \theta} & \frac{\partial m_{12}}{\partial \theta} & \cdots & \frac{\partial m_{1d}}{\partial \theta} \\ \frac{\partial m_{21}}{\partial \theta} & \frac{\partial m_{22}}{\partial \theta} & \cdots & \frac{\partial m_{2d}}{\partial \theta} \\ \vdots & \vdots & & \vdots \\ \frac{\partial m_{n1}}{\partial \theta} & \frac{\partial m_{n2}}{\partial \theta} & \cdots & \frac{\partial m_{nd}}{\partial \theta} \end{bmatrix}$$

- $\frac{\partial \boldsymbol{M}^{-1}}{\partial \theta} = ?$

We know that: $\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{I}$

$$\frac{\partial \boldsymbol{I}}{\partial \theta} = \frac{\partial \boldsymbol{M}\ \boldsymbol{M}^{-1}}{\partial \theta}$$

$$0 = \frac{\partial \boldsymbol{M}}{\partial \theta}\boldsymbol{M}^{-1} + \boldsymbol{M}\frac{\partial \boldsymbol{M}^{-1}}{\partial \theta}$$

$$\boldsymbol{M}\frac{\partial \boldsymbol{M}^{-1}}{\partial \theta} = -\frac{\partial \boldsymbol{M}}{\partial \theta}\boldsymbol{M}^{-1}$$

$$\frac{\partial \boldsymbol{M}^{-1}}{\partial \theta} = -\boldsymbol{M}^{-1}\frac{\partial \boldsymbol{M}}{\partial \theta}\boldsymbol{M}^{-1}$$

➢Vector derivative identities

- Consider a matrix $\boldsymbol{M}$ and a vector $\boldsymbol{y}$ that are independent of $\boldsymbol{x}$
- $\frac{\partial[\boldsymbol{M}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \boldsymbol{M}$
- $\frac{\partial[\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}]}{\partial \boldsymbol{x}} = \boldsymbol{y}$

$$[x_1 \quad x_2 \ldots\ldots \quad x_n]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1\, y_1 + \cdots + x_n\, y_n$$

After differentiation w.r.t. to $\boldsymbol{x} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{y}$

- Example: $\boldsymbol{M} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $\boldsymbol{x} = [x_1 \;\; x_2]^{\mathrm{T}}$ then find $\frac{\partial[\boldsymbol{Mx}]}{\partial \boldsymbol{x}}$?

$$\boldsymbol{Mx} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix}$$

If an n −dimensional vector-valued function $\boldsymbol{f}$ of d-dimensional vector $\boldsymbol{x}$

$$\boldsymbol{J}(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_d} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_n(\boldsymbol{x})}{\partial x_d} \end{bmatrix}$$

Then, $$\frac{\partial[\boldsymbol{Mx}]}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial[2x_1+x_2]}{\partial x_1} & \frac{\partial[2x_1+x_2]}{\partial x_2} \\ \frac{\partial[x_1+2x_2]}{\partial x_1} & \frac{\partial[x_1+2x_2]}{\partial x_1} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \boldsymbol{M}$$

- $\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x}]}{\partial \boldsymbol{x}} = ?$
- Steps

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \dots\dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & & \vdots \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= [(a_{11}x_1 + \ \dots + a_{n1}x_n), \ \dots\dots\dots\dots\dots\dots\dots, \ (a_{1n}x_1 + \ \dots + a_{nn}x_n)]\,\boldsymbol{x}$$

$$= [\textstyle\sum_{i=1}^{n} a_{i1}x_i \quad \dots\dots\dots\dots \quad \sum_{i=1}^{n} a_{in}x_i] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 \textstyle\sum_{i=1}^{n} a_{i1}x_i + \dots\dots\dots + x_n \sum_{i=1}^{n} a_{in}x_i$$

$$= \textstyle\sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij}x_i = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}x_i\,x_j$$

- $\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial[\boldsymbol{x}^{T}\boldsymbol{A}\boldsymbol{x}]}{\partial x_1} \\ \vdots \\ \frac{\partial[\boldsymbol{x}^{T}\boldsymbol{A}\boldsymbol{x}]}{\partial x_n} \end{bmatrix} = \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}x_i\,x_j\right)$

$$\Rightarrow \frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}_k} = \frac{\partial}{\partial \boldsymbol{x}_k}\left(x_1 \sum_{i=1}^{n} a_{i1}x_i + \ldots\ldots\ldots + x_k \sum_{i=1}^{n} a_{ik}x_i + \ldots\ldots\ldots + x_n \sum_{i=1}^{n} a_{in}x_i\right)$$

$$= x_1\, a_{k1} + \ldots\ldots\ldots + \left(x_k a_{kk} + \sum_{i=1}^{n} a_{ik}x_i\right) + x_n a_{kn}$$

$$= \sum_{j=1}^{n} a_{kj}x_j + \sum_{i=1}^{n} a_{ik}x_i$$

$$= [(k^{th}\text{ row of }\boldsymbol{A})\boldsymbol{x} + \left(\text{transpose of }k^{th}\text{ column of }\boldsymbol{A}\right)\boldsymbol{x}]$$

$$= [(k^{th}\text{ row of }\boldsymbol{A}) + (\text{transpose of }k^{th}\text{ column of }\boldsymbol{A})]\boldsymbol{x}$$

$$= \begin{matrix} [(1^{st}\text{ row of }\boldsymbol{A}) + (\text{transpose of }1^{st}\text{ column of }\boldsymbol{A})]\boldsymbol{x} \\ \vdots \\ [(n^{th}\text{ row of }\boldsymbol{A}) + \left(\text{transpose of }n^{th}\text{ column of }\boldsymbol{A}\right)]\boldsymbol{x} \end{matrix}$$

$$= \begin{bmatrix} [(1^{st}\text{ row of }\boldsymbol{A}) + (\text{transpose of }1^{st}\text{ column of }\boldsymbol{A})] \\ \vdots \\ [(n^{th}\text{ row of }\boldsymbol{A}) + \left(\text{transpose of }n^{th}\text{ column of }\boldsymbol{A}\right)] \end{bmatrix}\boldsymbol{x}$$

➢ $$\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \begin{bmatrix} 1^{st}\text{ row of }\boldsymbol{A} \\ \vdots \\ n^{th}\text{ row of }\boldsymbol{A} \end{bmatrix} + \begin{bmatrix} (\text{transpose of }1^{st}\text{ column of }\boldsymbol{A}) \\ \vdots \\ \left(\text{transpose of }n^{th}\text{ column of }\boldsymbol{A}\right) \end{bmatrix} = \left(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}\right)\boldsymbol{x}$$

➢If $\boldsymbol{A}$ is symmetric, then

$$\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial x_k} = 2\boldsymbol{A}\boldsymbol{x}$$

➢ $\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial x_k} = \frac{\partial}{\partial \boldsymbol{x}_k}(\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}x_i\, x_j) = \frac{\partial}{\partial \boldsymbol{x}_k}\,(\sum_{i=1}^{n} a_{ik}x_i x_k + \sum_{j=1}^{n} a_{kj}x_k x_j)$

$$= \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j$$

$$\left[\text{Differentiating a summation: } \frac{\partial}{\partial \boldsymbol{x}_k}\left(\sum_{i=1}^{n} a_{ik}x_i x_k\right) = \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{x}_k}(a_{ik}x_i x_k) = \sum_{i=1}^{n} a_{ik}x_i\right]$$

- Example: Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$ show $\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}, \frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}$

Solution: $\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \frac{\partial[4x_1^2+2x_1x_2+4x_2^2]}{\partial x_k} = \begin{bmatrix} 8x_1 + 2x_2 \\ 8x_2 + 2x_1 \end{bmatrix} = \begin{bmatrix} 8 & 2 \\ 2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$= 2\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2\boldsymbol{A}\boldsymbol{x}$$

From this example we can also verify

$$\frac{\partial[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}]}{\partial \boldsymbol{x}} = \begin{bmatrix} 8 & 2 \\ 2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left(\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}\right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\, \boldsymbol{x}$$

➢Second derivatives of a scalar function of a scalar x

➢Taylor series about a point

$$f(x) = f(x_0) + \frac{1}{1!}\frac{df(x)}{dx}\Bigg|_{x=x_0}(x - x_0)$$

$$+ \frac{1}{2!}\frac{d^2f(x)}{dx^2}\Bigg|_{x=x_0}(x - x_0)^2$$

$$+ \ldots\ldots\ldots + \frac{1}{(k-1)!}\frac{d^{k-1}}{dx^{k-1}}\Bigg|_{x=x_0}(x - x_0)^{k-1} + \ldots$$

- Example: $f(x) = \sin(x) = \frac{1}{1!}x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 \ldots\ldots.\text{at } x_0 = 0$

you can represent *any* infinitely-differentiable function as an infinite polynomial

- Analogously, if the scalar-valued function f is instead a function of a vector $\boldsymbol{x}$, we can expand $f(\boldsymbol{x})$ in a Taylor series around a point $\boldsymbol{x}_0$

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \left[\frac{\partial f}{\partial \boldsymbol{x}}\right]^{\mathrm{T}}_{\boldsymbol{x}=\boldsymbol{x}_0} (\boldsymbol{x} - \boldsymbol{x}_{\mathbf{0}}) + \frac{1}{2!} (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} \left[\frac{\partial^2 f}{\partial \boldsymbol{x}^2}\right]^{\mathrm{T}}_{\boldsymbol{x}=\boldsymbol{x}_0} (\boldsymbol{x} - \boldsymbol{x}_{\mathbf{0}}) + O\left(\left\|\boldsymbol{x} - \boldsymbol{x}_0\right\|^3\right)$$

  - $\boldsymbol{H} = \nabla^2 f(\boldsymbol{x})$ is the Hessian matrix, the matrix of second order derivatives of $f(\cdot)$, here evaluated at $\boldsymbol{x}_0$

- Example: $f(\boldsymbol{x}) = 2x_1^2 + x_2 + x_1 + 2x_2^2$ then find the Hessian matrix?

Hessian of a matrix can be given by $\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial f(\boldsymbol{x})}{\partial x_1 \partial x_2} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2^2} \end{bmatrix}$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial(2x_1^2+x_2+x_1+2x_2^2)}{\partial x_1} \\ \frac{\partial(2x_1^2+x_2+x_1+2x_2^2)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1 + 1 \\ 1 + 4x_2 \end{bmatrix}$$

$$\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

➢Least square solution

$$\boldsymbol{a}x = \boldsymbol{b}$$

$$E^2 = ||\boldsymbol{a}x - \boldsymbol{b}||^2 = (a_1x - b_1)^2 + (a_2x - b_2)^2 + \ldots\ldots + (a_mx - b_m)^2$$

$$\frac{dE^2}{dx}\Big|_{x=\hat{x}} = 0$$

$$2(a_1\hat{x} - b_1)a_1 + \ldots\ldots\ldots\ldots + 2(a_m\hat{x} - b_m)a_m = 0$$

$$a_1^2\hat{x} + a_2^2\hat{x} + \ldots\ldots\ldots a_m^2\hat{x} = a_1b_1 + \ldots\ldots + a_mb_m$$

$$\hat{x} = \frac{a_1b_1 + \ldots\ldots + a_mb_m}{a_1^2 + a_2^2 + \ldots\ldots\ldots + a_m^2} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$$

- Orthogonality of ***a*** and ***b***

$$\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{b} - \hat{x}\boldsymbol{a}) \Rightarrow \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} - \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 0$$

➢ Least square problems with several variables

- To project $\boldsymbol{b}$ onto a subspace – rather than just onto a line

  $\boldsymbol{Ax} = \boldsymbol{b}$

  $\boldsymbol{A}: m$ x n

- Choose $\widehat{\boldsymbol{x}}$ so as to minimize the error
- $E = ||\boldsymbol{Ax} - \boldsymbol{b}||$
- Projection $p = \boldsymbol{A}\,\widehat{\boldsymbol{x}}$
- Error vector: $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{A}\,\widehat{\boldsymbol{x}}$
- $\widehat{\boldsymbol{x}} = ?$

➢ Solution

- The error vector must be perpendicular to each column $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots\ldots, \boldsymbol{a}_n$ of $\boldsymbol{A}$

$$\begin{matrix} \boldsymbol{a}_1^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{A}\widehat{\boldsymbol{x}}) = 0 \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{A}\widehat{\boldsymbol{x}}) = 0 \end{matrix} \quad \text{or} \quad \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}-\boldsymbol{A}\widehat{\boldsymbol{x}} \end{bmatrix} = 0$$

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{A}\widehat{\boldsymbol{x}}) = 0 \Rightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

(OR)

$$E^2 = ||\boldsymbol{Ax} - \boldsymbol{b}||^2$$

$$E^2 = (\boldsymbol{Ax} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax} - \boldsymbol{b})$$

$$E^2 = (\boldsymbol{Ax} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax} - \boldsymbol{b}) = \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{b}^{\mathrm{T}}\right)(\boldsymbol{Ax} - \boldsymbol{b}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{Ax} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$$

$$\frac{dE^2}{d\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

$$\Rightarrow 2\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} - 2\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = 0$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{b}$$

- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible exactly when the columns of $\boldsymbol{A}$ are linearly independent

- Best estimate

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

- Projection

$$\boldsymbol{p} = \boldsymbol{A}\,\hat{\boldsymbol{x}}$$

$$\boldsymbol{p} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

- Projection matrix

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$$

  - The matrix that gives $\boldsymbol{p}$
  - $\boldsymbol{P}^2 = \boldsymbol{P}$
  - $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}$

- Example: Solve the following system of linear equations, using matrix inversion method: 5*x* + 2 *y* = 3, 3*x* + 2 *y* = 5 and find x and y.

The matrix form of the system is $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} = \begin{bmatrix} 5 & 2 \\ 3 & 2 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$

$|\boldsymbol{A}| = 10 - 6 = 4 \neq 0\,, \quad \boldsymbol{A}^{-1}$ exists and $\boldsymbol{A}^{-1} = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -3 & 5 \end{bmatrix}$

$$\boldsymbol{AA}^{-1} = \boldsymbol{I}_2 = \mathbf{A}^{-1}\boldsymbol{A} = (\mathbf{A}^{-1}\boldsymbol{A})^{-1}$$

If $\boldsymbol{Ax} = \boldsymbol{b}$

then, $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{B} = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -3 & 5 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$

Solution: $x = -1, y = 4$