EE2100: Matrix Analysis

Review Notes - 2[1]

Topics covered :

1. Introduction to Vectors
2. Properties of vectors

---

1. Vectors can be thought of as an ordered set of data. It should be noted that the ordered set of data could come from literally anything. A few examples of data that can be represented by vectors are
   - the coordinates of a point in space,
   - attributes of a physical object,
   - Sampled Signals: In the domain of signal processing, it is common practice to represent discrete time signals as vector. The value of the signal corresponding to each discrete instant can be represented as a vector. A few examples of processing signals using vector/matrix operations will be covered at a later point in the course.
   - the price of a stock,
   - In the context of control systems, it is a common practice to study the dynamics of the systems using a compact representation commonly referred to as the state space models. In the state-space model, the states of the system are typically denoted as a vector (known as the "State Vector"). It is interesting to note that, the data in the state vector changes with time.
2. The most common way to write vectors is by using a enclosing the ordered set of data using curved/rectangular brackets. For example, a vector with 3 values is typically written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

3. Vectors are commonly denoted by lower case boldface letters such as $\mathbf{x}$, $\mathbf{y}$. In some cases, the vectors are represented by a lower case letter with an overhead arrow (like $\vec{x}$).
4. The values in the vectors are commonly called as **entries/elements/coefficients/components**. The nature of vectors that are encountered in the course will have elements that are either Real numbers or complex numbers.
5. The number of elements of a vector defines its **length/dimension/size**. We will stick to using "length" to indicate the number of elements in a given vector.

[1] References: 1. S Boyd and L Vandenberghe, "Introduction to Applied Linear Algebra", Cambridge university press

6. A common way to indicate the nature of elements in a vector and its length is by adopting the set theory convention. A vector of real elements of length N can be presented as $\mathbf{x} \in \mathcal{R}^N$. Similarly, a vector of complex values of length N can be presented as $\mathbf{x} \in \mathcal{C}^N$.

7. The i^{th} entry/component of a vector $\mathbf{x}$ is denoted by x_i or $x[i]$. Notice here that there is no overhead arrow while representing the entries of the vector. The indexing of vectors goes from 1 to N (unlike some programming languages wherein the indexing is from 0).

8. Some common vectors (along with the notations employed) that we keep encountering in this course are

   - Zero vector (denoted by $\vec{0}$ or $\mathbf{0}$) is a vectors with each entry/element as 0.
   - Ones vector (denoted by $\vec{1}$ or $\mathbf{1}$) is a vectors with each entry/element as 1.
   - Unit vector (denoted by $\vec{e}_i$ or $\mathbf{e}_i$) is a vector where only the i^{th} element is equal to 1 and the rest of the elements are zero. For example, $\mathbf{e}_2 \in R^4$ is used to denote the following vector

$$\mathbf{e}_2/\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

   Mathematically, a unit vector $\mathbf{e}_j$ can be written as

$$e_{j,k} = \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases} \quad \text{(or)} \quad e_j[k] = \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases}$$

   - A unit vector in the direction of a $\mathbf{x}$ (note that the direction is an aspect that we have not yet defined in this course) is denoted by $\mathbf{e_x}$ or $\vec{e}_{\vec{x}}$.

9. Consider two vectors of same length i.e., $\mathbf{x}, \mathbf{y} \in \mathcal{R}^N$. The **addition** of two vectors of same length results in a vector of same length and is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{bmatrix} \in \mathcal{R}^N \tag{1}$$

   Notice that, we have overloaded the operator $+$.

10. It can easily be shown that the vector addition obeys the following laws

    - Commutative i.e., $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
    - Associative i.e., $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

- Addition with a zero vector i.e., $\mathbf{x} + \mathbf{0} = \mathbf{x}$.

11. The **scalar multiplication** or scalar vector product (with α representing a scalar, i.e., $\alpha \in \mathcal{R}$) is defined as

$$\alpha \mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_N \end{bmatrix} \in \mathcal{R}^N \tag{2}$$

12. It can easily be shown that the scalar multiplication obeys the following laws

- Commutative i.e., $\alpha \mathbf{x} = \mathbf{x} \alpha$.
- Associative i.e., $(\alpha \beta) \mathbf{x} = \alpha (\beta \mathbf{x})$
- Distributive over addition i.e., $\alpha (\mathbf{x} + \mathbf{y}) = \alpha \mathbf{x} + \alpha \mathbf{y}$.

13. Consider two vectors of same length i.e., $\mathbf{a}, \mathbf{b} \in \mathcal{R}^N$. The **inner product** or **dot product** of two vectors of same length results in a scalar and is defined as

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{N} a_i b_i \tag{3}$$

The inner product between two vectors is usually denoted by $\mathbf{a} \cdot \mathbf{b}$ or $\mathbf{a}^T \mathbf{b}$ or $< \mathbf{a}, \mathbf{b} >$. It can be easily shown that the inner product obeys the following laws.

- Commutative i.e., $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$.
- Associative with scalar multiplication i.e., $(\alpha \mathbf{a}) \cdot \mathbf{b} = \alpha (\mathbf{a} \cdot \mathbf{b})$
- Distributive with vector addition i.e., $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$.

14. The dot product of a vector with itself gives

$$\mathbf{a} \cdot \mathbf{a} = \sum_{i=1}^{N} a_i^2 \tag{4}$$

This is commonly referred to as the square of the magnitude (a more precise definition will follow later) or euclidean distance.

15. The dot product of a vector with the unit vector e_i extracts the i^{th} component of the vector i.e.,

$$\mathbf{a} \cdot \mathbf{e}_i = a_i \tag{5}$$

EE2100: Matrix Analysis

Review Notes - 3

Topics covered :

1. Norm and Direction of a vector
2. Cauchy Schwarz inequality

---

1. Let $\mathbf{x}, \mathbf{y} \in \mathcal{C}^N$. The **dot product** of two vectors with complex entries is defined as

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{N} x_i y_i^* \tag{1}$$

2. The **Norm** of a vector is mathematical function (that operates on a vector i.e., $\|\vec{x}\| : \mathcal{R}^N \to \mathcal{R}$) which satisfies the following properties

   - $\|\vec{x}\| > 0$ for all $\vec{x} \neq \vec{0}$,
   - $\|\alpha\vec{x}\| = |\alpha| \, \|\vec{x}\|$ for all $\alpha \in \mathcal{R}$ and,
   - $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (commonly referred to as triangle inequality)

3. The function defined by (2) satisfies all the properties of the norm and is commonly referred to as norm of order p or L_p norm. The parameter p in (2) is called the order of the norm.

$$\|\vec{x}\|_p = \sqrt[p]{\sum_{i=1}^{N} |x_i|^p} \tag{2}$$

   - The norm of order 1 (i.e., L_1 norm) denoted by $\|\vec{x}|_1$ is defined as

$$\|\vec{x}\|_1 = \sum_{i=1}^{N} |x_i| \tag{3}$$

   If all the elements of the vector $\mathbf{x}$ are positive i.e., $x_i \geq 0 \; \forall i$, the L_1 norm is related to the average value (denoted by $\bar{x}$) of the entries i.e.,

$$\bar{x} = \frac{1}{N}\|\vec{x}\|_1 \tag{4}$$

   The L_1 norm of the vector is also referred to as "Taxi-Cab" norm. It is easy to show that the L_1 norm satisfies all the properties of the norm. The proof for triangle inequality in the L_1 norm is as follows.

$$\begin{aligned} \|\vec{x} + \vec{y}\|_1 &= |x_1 + y_1| + \cdots + |x_N + y_N| \\ &\leq |x_1| + |y_1| + \cdots + |x_N| + |y_N| \;\; (\text{since } |x_i + y_i| \leq |x_i| + |y_i|) \\ &\leq \|\vec{x}\|_1 + \|\vec{y}\|_1 \end{aligned} \tag{5}$$

- The norm of order 2 (i.e., L_2 norm) denoted by $\|\vec{x}|_2$ is defined as

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^{N} |x_i|^2} \; = \; \sqrt{\sum_{i=1}^{N} x_i^2} \tag{6}$$

If the elements of the correspond to the sampled value of a signal, the L_1 norm is related to the root mean square (RMS) value of the signal (denoted by x_{rms}) i.e.,

$$x_{rms} = \sqrt{\frac{1}{N}\sum_{1}^{N} x_i^2} \; = \; \frac{1}{\sqrt{N}}\|\vec{x}\|_2 \tag{7}$$

The L_2 norm of the vector is also referred to as "Euclidean" norm.

- The norm of order ∞ (i.e., L_∞ norm) is defined as

$$\|\vec{x}\|_\infty = \max |x_i| \tag{8}$$

In this course, if the order of the norm is not explicitly specified, we treat it as an L_2 norm.

4. Consider $\mathbf{x} \in \mathcal{R}^2$.

- The locus of all the vectors in R^2 which have an L_1 norm of a is characterized by the equation $|x_1| + |x_2| = a$. The locus is a diamond in R^2 (shown in Fig. 1).
- On the other hand, the locus of all the vectors in R^2 which have an L_2 norm of a is characterized by the equation $x_1^2 + x_2^2 = a$. The locus is a circle in R^2 (shown in Fig. 1).
- The locus of all the vectors in R^2 which have an L_∞ norm of a is characterized by the relation $\max(|x_1|, |x_1|) = a$. The locus is a square in R^2 (shown in Fig. 1).

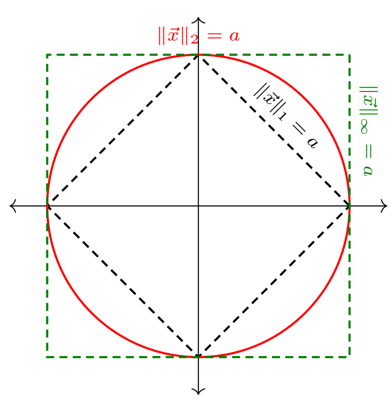


Fig. 1: Locus of vectors in $\mathcal{R}^2$ with constant norms

5. A vector that is scaled by a factor $\alpha \in \mathcal{R}$ lies along the original vector and its norm is scaled by a factor $|\alpha|$. Consider a vector $\mathbf{x} \in \mathcal{R}^N$. A vector of unit length along $\mathbf{x}$ is called the unit vector along $\vec{x}$ or the

direction of $\mathbf{x}$ and is denoted by $\vec{e}_x$ or $\mathbf{e}_x$. Mathematically,

$$\mathbf{e}_x = \frac{1}{\|x\|}\mathbf{x} \tag{9}$$

6. **Cauchy-Schwarz inequality**: The Cauchy-Schwarz inequality (relates the norm and the inner product of two vectors) states that

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\| \tag{10}$$

To prove, consider two vectors $\vec{x},\vec{y} \in \mathcal{R}^N$. Let $\vec{z}$ be defined as $\mathbf{z} = \mathbf{e}_x - \mathbf{e}_y$. Since the norm of a vector is always positive or zero,

$$\begin{aligned}
\|\vec{z}\|^2 \;=\; \vec{z}\cdot\vec{z} &\geq 0 \\
(\mathbf{e}_x - \mathbf{e}_y)\cdot(\mathbf{e}_x - \mathbf{e}_y) &\geq 0 \\
\|\mathbf{e}_x\|^2 + \|\mathbf{e}_y\|^2 - 2\mathbf{e}_x\cdot\mathbf{e}_y &\geq 0 \\
2 - 2\mathbf{e}_x\cdot\mathbf{e}_y &\geq 0 \\
\mathbf{e}_x\cdot\mathbf{e}_y &\leq 1 \\
\implies \mathbf{x}\cdot\mathbf{y} &\leq \|\mathbf{x}\|\|\mathbf{y}\|
\end{aligned} \tag{11}$$

Note that (11) does not completely prove the Cauchy Schwarz inequality. Consider $\mathbf{y}_1 = \mathbf{y}$. According to (11),

$$\mathbf{x}\cdot\mathbf{y}_1 \leq \|\mathbf{x}\|\|\mathbf{y}_1\| \implies \mathbf{x}\cdot\mathbf{y} \geq -\|\mathbf{x}\|\|\mathbf{y}\| \tag{12}$$

Using (11) and (12), the following relation can be deduced.

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\| \tag{13}$$

7. The triangle inequality for L_2 norm can be proved by using the Cauchy-Schwarz inequality. Consider,

$$\begin{aligned}
\|\vec{x}+\vec{y}\| &= \sqrt{(\vec{x}+\vec{y})\cdot(\vec{x}+\vec{y})} \\
&= \sqrt{\|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\underbrace{\vec{x}\cdot\vec{y}}_{\leq\|x\|\|y\|}} \\
&\leq \sqrt{\|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\|x\|\|y\|} \;\leq\; \sqrt{(\|\vec{x}\| + \|\vec{y}\|)^2} \\
&\leq \|\vec{x}\| + \|\vec{y}\|
\end{aligned} \tag{14}$$

EE2100: Matrix Analysis

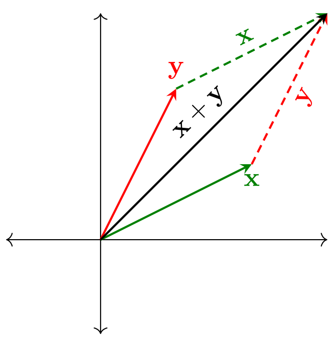
Review Notes - 4

Topics covered :

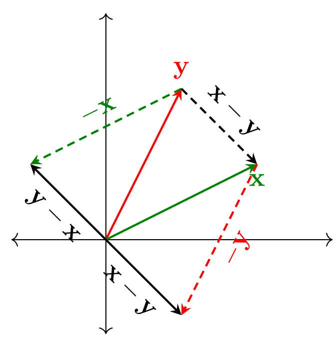
1. Angle between two vectors
2. Projection

---

1. Consider two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{R}^2$. The two vectors can be visualized in a two dimensional space as shown in Fig. 1. The vectors obtained by adding the two vectors and taking the difference between them are shown in Fig. 1(b) and Fig. 1(b) respectively. One key take away from the geometric interpretation of the sum



(a) Vector addition



(b) Vector subtraction

Fig. 1: Representation of vector sum and difference in $\mathcal{R}^2$

and difference of two vectors is that the original vectors along with their sum/difference form a triangle.

2. Consider two vectors $\mathbf{x},\ \mathbf{y} \in \mathcal{R}^N$. The position vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ constitute a triangle in $\mathcal{R}^N$. The magnitude of corresponding sides is given by $\|\mathbf{x}\|$, $\|\mathbf{y}\|$ and $\|\mathbf{x}-\mathbf{y}\|$ (see Fig. 2). The angle between $\mathbf{x}$ and

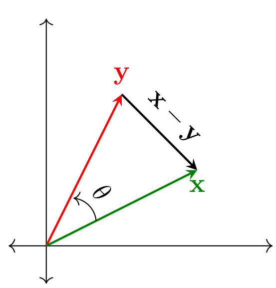


Fig. 2: Application of cosine rule

$\mathbf{y}$ (denoted by θ) can be computed by applying the cosine rule. According to cosine rule,

$$\begin{aligned}
\|\mathbf{x}-\mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}\cdot\mathbf{y} &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
\cos\theta &= \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}
\end{aligned} \tag{1}$$

Since $-1 \leq \cos\theta \leq 1$, a relation between the dot product of two vectors and their norm can be written as $|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ (Cauchy-Schwarz inequality).

3. The cosine of angle between the two vectors is equal to 1 i.e., $\cos\theta = 1$ when $\mathbf{y} = \alpha\mathbf{x}$ where $\alpha \in \mathcal{R}^+$ ($\mathcal{R}^+$ indicates that α is a scalar that can take only positive values or zero). Similarly, the cosine of angle between the two vectors is equal to -1 i.e., $\cos\theta = -1$ when $\mathbf{y} = \alpha\mathbf{x}$ where $\alpha \in \mathcal{R}^-$ ($\mathcal{R}^-$ indicates that α is a scalar that can take only negative values or zero).

4. Consider two vectors $\mathbf{x},\ \mathbf{y} \in \mathcal{R}^N$. If the dot product between the two vectors is zero i.e., $\mathbf{x}\cdot\mathbf{y} = 0$, the angle between the vectors is $\frac{\pi}{2}$ (or $\frac{3\pi}{2}$) (i.e., $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to each other). Set of vectors whose inner product is zero are referred to as orthogonal vectors. The set of vectors whose magnitude is 1 and are orthogonal to each other are referred to as orthonormal vectors.

5. Consider $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathcal{R}^N$. Let θ_{xy} and θ_{xz} denote the angle between the vectors ($\mathbf{x}$,$\mathbf{y}$) and ($\mathbf{x}$,$\mathbf{z}$) respectively. If $|\cos\theta_{xy}| < |\cos\theta_{xz}|$ then, we can state that $\mathbf{z}$ is more closer to $\mathbf{x}$ than $\mathbf{y}$.

6. **Projection** of one vector on another: Consider two vectors $\mathbf{a},\ \mathbf{b} \in R^N$. The vector $\mathbf{p}$ (say) along $\mathbf{b}$ is called the vector projection of $\mathbf{a}$ on $\mathbf{b}$ (denoted by $\mathbf{Proj_b a}$) if the distance from $\mathbf{a}$ to $\mathbf{p}$ is minimum. Mathematically, the vector projection can be defined as

$$\mathbf{Proj_b a} := \min_{\mathbf{p}}\ \|\mathbf{a} - \mathbf{p}\| \tag{2}$$

Since $\mathbf{p}$ is along $\mathbf{b}$, it can be represented as $\mathbf{p} = \alpha\vec{e}_{\mathbf{b}}$, where α is called the scalar projection of $\mathbf{a}$ on $\mathbf{b}$. The

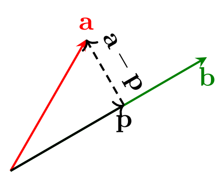


Fig. 3: Geometrical representation of projection

shortest distance is achieved when $\mathbf{a} - \mathbf{p}$ is orthogonal to $\mathbf{e_b}$ (recollect the arguments covered in the class). Hence,

$$\begin{aligned} (\mathbf{a} - \mathbf{p})\cdot\mathbf{e_b} &= 0 \\ \mathbf{a}\cdot\mathbf{e_b} &= \mathbf{p}\cdot\mathbf{e_b} = \alpha \\ \text{scalar projection: } \alpha &= \mathbf{a}\cdot\vec{e}_b \end{aligned} \tag{3}$$

The vector projection of $\mathbf{a}$ onto $\mathbf{b}$ is given by

$$\mathbf{Proj_b a} = (\mathbf{a}\cdot\mathbf{e_b})\,\mathbf{e_b} \tag{4}$$

7. **Representation of a vector as a linear combination of orthogonal vectors**: Consider three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{R}^2$ (see Fig. 4), with $\mathbf{b} \perp \mathbf{c}$ (i.e., $\mathbf{b}$ and $\mathbf{c}$ are orthogonal to each other). From Fig. 4, it can be

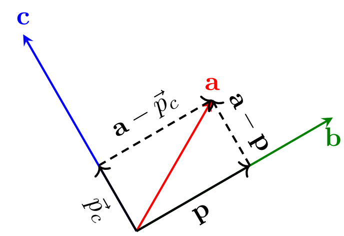


Fig. 4: Projection of vector onto two orthogonal vectors

seen that

$$\mathbf{a} = \underbrace{\mathbf{p}}_{\text{I}} + \underbrace{(\mathbf{a}-\mathbf{p})}_{\text{II}} \tag{5}$$

Since $(\mathbf{a}-\mathbf{p}) \perp \mathbf{e_b}$ and $\mathbf{b} \perp \mathbf{c}$, we can deduce that $(\mathbf{a}-\mathbf{p})$ is parallel to $\mathbf{c}$ and hence, it may be possible to represent Term II in Eq. (5) in terms of $\mathbf{c}$ or $\vec{e}_\mathbf{c}$. (One needs to be a little careful with this argument in N-dimensional space. We will look at this aspect at a later point in course). Accordingly, Eq. (5) can be written as

$$\mathbf{a} = \underbrace{\mathbf{p}}_{\text{I}} + \underbrace{\alpha_c \mathbf{e_c}}_{\text{II}} \tag{6}$$

The value of α_c can be computed by taking inner product with $\mathbf{e_c}$ on both sides of Eq. (6).

$$\mathbf{a}\cdot\mathbf{e_c} = \underbrace{\mathbf{p}\cdot\mathbf{e_c}}_{0} + \underbrace{\alpha_c\mathbf{e_c}\cdot\mathbf{e_c}}_{\alpha_c} \implies \alpha_c = \mathbf{a}.\mathbf{e_c} \tag{7}$$

It is interesting to note that α_c is equal to scalar projection of $\mathbf{a}$ onto $\mathbf{c}$. The vector $\mathbf{a}$ can thus be written in terms of $e_\mathbf{b}$ and $\mathbf{e_c}$ as

$$\mathbf{a} = \alpha_b \mathbf{e_b} + \alpha_c \mathbf{e_c} \tag{8}$$

Equation (8) expresses $\mathbf{a}$ in terms of $\mathbf{e_b}$ and $\mathbf{e_c}$ i.e., the vector $\mathbf{a}$ is expressed as a linear combination of $\mathbf{e_b}$ and $\mathbf{e_c}$. This is the first instance where we come across this idea of linear combination (a fundamental concept that we will define a bit later in the course). The values α_b and α_c are often referred to as coefficients of linear combination.

EE2100: Matrix Analysis

Review Notes - 5

Topics covered :

1. Binary operations and Fields
2. Vector space

---

1. A binary operation is defined as an operation that when performed on a set (say $\mathbb{A}$) results in an element from the same set. The elements of the set can be real/complex numbers or functions etc.

   - The operators $+$, $-$, $\times$ and $/$ are all binary operators on a set of real numbers ($\mathbb{R}$ or $\mathcal{R}$) or a set of complex numbers ($\mathbb{C}$ or $\mathcal{C}$).
   - If $\mathbb{N}$ is a set of natural numbers, the operators $+$ and $\times$ are binary while the operators $-$ and $/$ are not.
   - If $\mathbb{Z}$ is a set of integers (positive, negative and zero), the operators $+$, $-$ and $\times$ are binary.
   - If $\mathbb{D}$ is a set of Boolean values i.e., $\mathbb{D} = \{\text{T,F}\}$, the operators OR, AND and XOR (even others) are binary.

   Mathematically, a binary operation on a set $\mathcal{A}$ is often represented using the following convention

$$\begin{aligned} \text{operator} &: (\text{Set})\text{operator}(\text{set}) \rightarrow (\text{set}) \\ + &: \mathbb{R} + \mathbb{R} \rightarrow \mathbb{R} \\ \times &: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \end{aligned} \tag{1}$$

2. A Field is a set (typically denoted by $\mathbb{F}$) with two defined binary operators (commonly denoted by $+$ and $\cdot$ and referred to as addition and multiplication respectively; note that $+ : \mathbb{F} + \mathbb{F} \rightarrow \mathbb{F}$ and $\cdot : \mathbb{F} \cdot \mathbb{F} \rightarrow \mathbb{F}$) that satisfy the following conditions.

   (a) Commutative property for the two defined binary operators i.e., addition and multiplication: $\forall a, b \in \mathbb{F}$, $a + b = b + a$ and $a \cdot b = b \cdot a$

   (b) Associative property for the two defined binary operators i.e., addition and multiplication: $\forall a, b, c \in \mathbb{F}$, $(a + b) + c = a + (b + c)$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

   (c) Existence of additive and multiplicative identities: $\forall a \in \mathbb{F}$, there is an element denoted by 0 such that $a + 0 = a$ and an element denoted by 1 such that $a \cdot 1 = a$. The elements denoted by 0 and 1 are called the additive and multiplicative identities respectively.

   (d) Existence of additive and multiplicative inverses: $\forall a \in \mathbb{F}$ and a non-zero $b \in \mathbb{F}$, there exists elements $c, d \in \mathbb{F}$ such that $a + c = 0$ and $b \cdot d = 1$. The elements c and d are called additive and multiplicative inverses respectively.

3. The fields are commonly denoted as $(\mathbb{F}, +, \cdot)$, where $\mathbb{F}$ denotes the set under consideration and $+, \cdot$ indicate the binary operators defined over the set.

   - $(\mathbb{C}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ with the standard $+$ and $\cdot = \times$ is a Field.
   - $(\mathbb{N}, +, \cdot)$ with the standard $+$ and $\cdot = \times$ is a not Field due to lack of additive identity and additive and multiplicative inverses while $(\mathbb{Z}, +, \cdot)$ with the standard $+$ and $\cdot = \times$ is a not Field due to lack of multiplicative inverse.
   - $(\mathbb{D}, \text{XOR}, \text{AND})$, with XOR indicating the $+$ operator and AND indicating the $\cdot$ operator is a Field.

4. A Vector space (commonly denoted by $\mathbb{V}$) over a Field $\mathbb{F}$ is a set which has a binary operator $+ : \mathbb{V} + \mathbb{V} \to \mathbb{V}$ (i.e., for any $x, y \in \mathbb{V}$, $x + y \in \mathbb{V}$) and an operator $\cdot : \mathbb{F} \cdot \mathbb{V} \to \mathbb{V}$ (i.e., for any $\alpha \in \mathbb{F}$ and $x \in \mathbb{V}$, $\alpha \cdot x \in \mathbb{V}$) that satisfy the following conditions.

   (a) Commutative property for vector addition: $\forall x, y \in \mathbb{V}$, $x + y = y + x$.

   (b) Associative property for vector addition: $\forall x, y, z \in \mathbb{V}$, $(x + y) + z = x + (y + z)$.

   (c) Existence of additive identity: $\forall x \in \mathbb{V}$, there is an element denoted by $0_v \in \mathbb{V}$ such that $x + 0_v = x$. The element denoted by 0_v is called the additive identity.

   (d) Existence of additive inverse: $\forall x \in \mathbb{V}$, there is an element $y \in \mathbb{V}$ such that $x +_v y = 0$. The element y is called the additive inverse.

   (e) Existence of multiplicative identity: $\forall x \in \mathbb{V}$, there is an element $1_f \in \mathbb{F}$ such that $1_f \cdot x = x$.

   (f) Associative property 2: $\forall \alpha, \beta \in \mathbb{F}$ and $\forall x \in \mathbb{V}$, $(\alpha \cdot \beta) \cdot x = \alpha \cdot (\beta \cdot x)$

   (g) Distributive property 1: $\forall \alpha, \beta \in \mathbb{F}$ and $\forall x \in \mathbb{V}$, $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$.

   (h) Distributive property 2: $\forall \alpha \in \mathbb{F}$ and $\forall x, y \in \mathbb{V}$, $\alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y)$.

   The operators $+$ and $\cdot$ defined over $\mathbb{V}$ are referred to as vector addition and scalar multiplication respectively.

5. Note that the definition of vector space relies on operator overloading.

EE2100: Matrix Analysis

Review Notes - 6

Topics covered :

1. Theorems on Vector Space
2. Concept of Subspace of a Vector space
3. Linear combination and Span

---

1. If $\mathbb{V}$ is a Vector space defined over $\mathbb{F}$ with two operators $+$ and $\cdot$ defined over the vector space, then the following statements are True.

 - For any $x, y, z \in \mathbb{V}$, if $x + y = x + z$, then $y = z$. (commonly referred to as Cancellation law of Vector addition)

 Proof: Since $\mathbb{V}$ is a vector space and $x \in \mathbb{V}$, there exists an element in $u \in \mathbb{V}$ such that $x + u = 0$ (notice here, the 0 is a 0 element of the vector space). Accordingly,

$$\begin{aligned} y &= y + 0 \;=\; y + (x + u) \\ &= (y + x) + u \;=\; (x + y) + u \text{ [by Commutative and Associative properties]} \\ &= (x + z) + u \;=\; (z + x) + u \text{ [by Commutative property]} \\ &= z + (x + u) \;=\; z \text{ [by Associative property]} \end{aligned} \tag{1}$$

 - The 0 i.e., Additive identity of a vector space is unique.

 Proof by Contradiction: Let $0_1, 0_2 \in \mathbb{V}$ be two elements of a vector space such that for any $x \in \mathbb{V}$, $x + 0_1 = x$ and $x + 0_2 = x$. Accordingly

$$\begin{aligned} x + 0_1 &= x \;=\; x + 0_2 \\ 0_1 &= 0_2 \text{ [by cancellation property]} \end{aligned} \tag{2}$$

 Hence, additive identity is unique.

 - The additive inverse of an element in vector space is unique.

 Proof by Contradiction: Let $u_1, u_2 \in \mathbb{V}$ be the two additive inverse elements of $x \in \mathbb{V}$. Then by definition $x + u_1 = 0$ and $x + u_2 = 0$. Since the additive identity is unique,

$$\begin{aligned} x + u_1 &= x + u_2 \\ u_1 &= u_2 \text{ [by cancellation property]} \end{aligned} \tag{3}$$

 Hence, additive inverse of an element in vector space is unique.

2. The set of all vectors of length N from a field F are denoted by F^N. In this course, the field F we are interested in is either the field of real numbers or the field of complex numbers. Consequently, the vectors that we come across these course come from $\mathcal{R}^N$ or $\mathcal{C}^N$.

3. If $\mathbb{V}$ is a collection of vectors in $\mathcal{R}^N$ or $\mathcal{C}^N$ (notice here that the information about the field over which the vectors are defined is already captured in the notation) with vector addition and scalar multiplication defined as (4) constitutes a vector space.

$$\text{for } a, b \in \mathbb{V},\ a + b\ = \begin{bmatrix} a_1 + b_1 \\ \cdots \\ a_N + b_N \end{bmatrix} \text{and, for } \alpha \in \mathbb{F},\, a \in \mathbb{V},\ \alpha \cdot a\ = \begin{bmatrix} \alpha a_1 \\ \cdots \\ \alpha a_N \end{bmatrix} \tag{4}$$

   Notice in (4), the $\cdot$ operator is used to indicate scalar multiplication (don't confuse it with dot product).

4. In most of the future discussions, the statement $\mathbf{x} \in \mathcal{R}^n$ or $\mathbf{x} \in \mathcal{R}^N$ indicates that $\mathbf{x}$ is an element from the vector space R^N and $\alpha\mathbf{x}$ indicates the scalar product.

5. A subset $\mathbb{W}$ of a vector space $\mathbb{V}$ over field $\mathbb{F}$ is called a subspace of a vector space if $\mathbb{W}$ is a vector space over $\mathbb{F}$ with operations of vector addition and scalar multiplication defined in $\mathbb{V}$. In simple terms, a $\mathbb{W} \subseteq \mathbb{V}$ is a subspace if it satisfies all the conditions of a vector space with the vector addition and scalar multiplication operators remaining the same as that defined for $\mathbb{V}$.

6. Since the elements of $\mathbb{W}$ are also elements of $\mathbb{V}$ (which satisfy the conditions for a vector space), it is not necessary to check for all conditions required for a vector space. (Discussed in preclass lecture).

7. If $\mathbb{W}$ is a set of one or more vectors from a vector space $\mathbb{V}$ over $\mathbb{F}$, then $\mathbb{W}$ is a subspace of $\mathbb{V}$ if and only if the following conditions hold

   (a) for any $x, y \in \mathbb{W}$, $x + y \in \mathbb{W}$,

   (b) for any $\alpha \in \mathbb{F}$, and $x \in \mathbb{W}$, $\alpha \cdot x \in \mathbb{W}$,

   (c) the additive identity is an element of $\mathbb{W}$, and

   (d) for any $x, \in \mathbb{W}$, there is an element $y \in \mathbb{W}$ such that $x + y = 0$. (presence of additive inverse).

8. For example, the ordered set of points on the line (i.e., collection of vectors with each element representing the coordinate of a point through the line) through the origin in a two dimensional space $\mathcal{R}^2$ is a Subspace of $\mathcal{R}^2$.

9. Linear combination of a set of vectors: Let $\{\mathbf{v}_1, \cdots \mathbf{v}_m\} \in \mathcal{R}^N$ be a set of vectors. Then the linear combination of vectors is mathematically defined as

$$\sum_{i=1}^{m} \alpha_i \mathbf{v}_i \tag{5}$$

   The set of α_i's are called coefficients of linear combination.

10. Span of a set of vectors: The span of a set of vectors $\mathbb{V}$ is a collection (or set) of all linear combinations of the vectors $\mathbb{V}$ i.e.,

$$\text{Span}\{\mathbf{v}_1, \cdots \mathbf{v}_m\} = \left\{ \sum_{i=1}^{m} \alpha_i \mathbf{v}_i \ : \ \forall \alpha_i \in \mathcal{R} \right\} \tag{6}$$

Essentially, Span is a collection or set of vectors that can be obtained by taking a linear combination of the vectors in $\mathbb{V}$. Theoretically, the span of a given vector set could have an infinite collection of vectors.

EE2100: Matrix Analysis

Review Notes - 7

Topics covered :

1. Linear combination
2. Span
3. Linear independence

---

1. A note on notation: The key point of focus in this and a few lecture from now will be on studying aspects related to a set of vectors. Let $\{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$ be a set of vectors and we will denote the set of vectors with $\mathbb{U}$. The notation u_{ij} indicates the jth element of vector $\mathbf{u}_i$.

2. Linear combination of a set of vectors: Let $\{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$ be a set of vectors. Then the linear combination of vectors is mathematically defined as

$$\sum_{i=1}^{m} \alpha_i \mathbf{u}_i \tag{1}$$

The set of α_i's are called coefficients of linear combination.

3. Span of a set of vectors: The span of a set of vectors $\mathbb{U}$ is a collection (or set) of all linear combinations of the vectors $\mathbb{U}$ i.e.,

$$\textbf{Span}\{\vec{u}_1, \cdots \vec{u}_m\} = \left\{ \sum_{i=1}^{m} \alpha_i \mathbf{u}_i \ : \ \forall \alpha_i \in \mathcal{R} \right\} \tag{2}$$

Essentially, Span is a collection or set of vectors that can be obtained by taking a linear combination of the vectors in $\mathbb{U}$. Theoretically, the span of a given vector set could have an infinite collection of vectors.

4. The span of a set of vectors (in $\mathcal{R}^N$) is a subspace of $\mathcal{R}^N$. [Recollect the arguments covered in the class].

5. The idea of Span is at the heart of the fundamental objectives of the topics covered in this course. Some of them are

   (a) Objective 1: Given a subspace $\mathbb{W} \subseteq \mathcal{R}^N$, can we find a collection of vectors $\{\vec{u}_1, \cdots \vec{u}_m\}$ such that $\textbf{Span}\{\vec{u}_1, \cdots \vec{u}_m\} = \mathbb{W}$. Further, can we come up with a set of minimum number of vectors, say $\mathbb{U} = \{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$, such the the span of $\mathbb{U}$ contains all possible vectors in $\mathcal{R}^N$ (Another common way of stating the objective is "Can we from a set $\mathbb{U}$ with minimum number of vectors such that it spans entire $\mathcal{R}^N$?").

   The associated followup question that we will also be interested in is as follows. Given a set $\mathbb{U}$ ($\textbf{Span}(\mathbb{U} = \mathbb{W})$), and a vector $\mathbf{a} \in \mathbb{W}$, compute the coefficients of linear combination α_i such that $\mathbf{a} = \sum_i \alpha_i \mathbf{u}_i$. Further, what properties should the vectors in $\mathbb{U}$ satisfy for the coefficients α_i to be unique?

(b) Objective 2: Given a set of vectors $\{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$, can we figure out if $\mathbb{U}$ spans the subspace of our interest $\mathbb{W} \subseteq \mathcal{R}^N$?

The associated followup questions that we will also be interested in are as follows. If $\mathbb{U}$ spans entire $\mathbb{W} \subseteq \mathcal{R}^N$, how would we compute the coefficients of linear combination for a given vector $\mathbf{a} \in \mathbb{W}$. On the other hand, $\mathbb{U}$ does not span entire $\mathbb{W}$, given a vector $\mathbf{a} \notin \text{Span}(\mathbb{U})$, can we identify a vector $\mathbf{b} \in \text{Span}(\mathbb{U})$ such that $\|\mathbf{b} - \mathbf{a}\|_2$ is minimum.

6. Let $\mathbb{U} = \{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$. If any vector, say $\mathbf{u}_i$ can be expressed as a linear combination of other elements in $\mathbb{U}$, then $\textbf{Span}(\mathbb{U}) = \textbf{Span}(\mathbb{U} \backslash \mathbf{u}_i)$ (i.e., set $\mathbb{U}$ excluding $\mathbf{u}_i$) [Revisit the examples covered in the class].

7. Let $\mathbb{U} = \{\vec{u}_1, \cdots \vec{u}_m\} \in \mathcal{R}^N$. The vectors in set $\mathbb{U}$ are linearly independent if no vector in $\mathbb{U}$ can be expressed as a linear combination of other vectors in $\mathbb{U}$. Alternatively, the vectors in set $\mathbb{U}$ are linearly independent if the only linear combination that takes the linear combination of vectors in $\mathbb{U}$ to $\mathbf{0} \in \mathbf{R}^N$ is when $\alpha_i = 0 \ \forall i$ i.e., the summation $\sum_{i=1}^{m} \alpha_i \mathbf{u}_i = 0$ only when $\alpha_i = 0 \ \forall i$. A set of vectors that are not linearly independent are called linearly dependent.

8. Let $\mathbb{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\} \in \mathcal{R}^N$ be a collection of linearly independent vectors. Then any vector $\mathbf{y}$ in the span of $\mathbb{U}$ can be expressed uniquely in terms of the elements of $\mathbb{U}$.

Proof by contradiction: Consider $\mathbf{y} \in \text{Span}(\mathbb{U})$, where $\mathbb{U}$ is a collection of linear independent vectors. Let $\mathbf{y} \neq \vec{0}$ be represented in terms of the linear combination of elements in a non-unique way i.e., there exists sets of coefficients of linear combination, say, $(\alpha_1, \cdots, \alpha_m)$ and $(\beta_1, \cdots, \beta_m)$ such that

$$\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{u}_i = \sum_{i=1}^{m} \beta_i \mathbf{u}_i \implies \sum_{i=1}^{m} (\alpha_i - \beta_i) \mathbf{u}_i = 0 \tag{3}$$

Equation 3 is a contradiction the fact that $\mathbb{U}$ is a set of linearly independent vectors. Equivalently, (3) indicates that for the linear combination to be non-unique, the set of vectors in $\mathbb{U}$ must be linearly dependent.

9. Let $\mathbb{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\} \in \mathcal{R}^N$ be a collection of vectors in subspace $\mathbb{W}$. The set $\mathbb{U}$ is called a spanning set if and only if, the set $\mathbb{U}$ spans the entire subspace $\mathbb{W}$ or stated equivalently, the set $\mathbb{U}$ is called a spanning set if and only if every $\mathbf{w} \in \mathbb{W}$ can be expressed as a linear combination of vectors in $\mathbb{U}$.

10. A spanning set of subspace $\mathbb{W}$ that consists of a collection of linearly independent vectors is called the basis of subspace $\mathbb{W}$. Alternatively, the basis can also be defined as the minimum number of vectors that span a given subspace.

11. In this course, the collection of basis vectors of a subspace is denoted by $\mathbb{B}$ and the individual elements (or basis vectors) are denoted by $\mathbf{b}_i$ (i.e., the basis vector i).

EE2100: Matrix Analysis

Review Notes - 8

Topics covered :

1. Dimension of a Subspace
2. Some Theorems

---

1. Any set $\mathbb{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\} \in \mathcal{R}^N$ constitutes a spanning set of the subspace $\mathbb{W} \subseteq \mathcal{R}^N$ if and only if, the set $\mathbb{U}$ spans the entire subspace $\mathbb{W}$.

2. Definition of Basis of a Subspace: The basis (typically denoted by $\mathbb{B}$) of a subspace (say $\mathbb{W}$) can be defined in multiple ways. Two of the commonly adopted definitions (in this course) are

Def. (a) A spanning set of subspace $\mathbb{W}$ that consists of a collection of linearly independent vectors is called the basis of subspace $\mathbb{W}$.

Def. (b) The basis can also be defined as the minimum number of vectors that span a given subspace.

Note that, defining basis using the definition (b) requires that the set of the vectors in the basis be linearly independent. An easy way to establish the equivalence of the definitions is as follows.

Let $\mathbb{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ denote the basis of subspace which contains linearly independent vectors. Assume that $\mathbb{B}$ is non-minimum i.e., it is possible to span $\mathbb{W}$ with a set $\mathbb{B}_1 \subset \mathbb{B}$. However, this is not possible since any vector $\mathbf{b}_i \in \mathbb{B} \in \mathbb{W}$ cannot be expressed as a linear combination of other vectors in $\mathbb{B}_1$ (given $\mathbb{B}$ is a linearly independent set). Accordingly, any subset $\mathbb{B}_1 \subset \mathbf{B}$ cannot span $\mathbb{W}$ i.e., any $\mathbb{B}_1 \subset \mathbf{B}$ cannot form the Basis of $\mathbb{W}$.

3. Any vector in the subspace (say $\mathbf{w}_i \in \mathbb{W}$) can be represented uniquely as a linear combination of the basis vectors of the corresponding subspace. This comes as a natural consequence of the linearly independence and spanning properties of the basis vectors.

4. Any given subspace can have the more than a single Basis. However, the definition of basis as a minimum number of vectors that span the given subspace necessitates that all basis have the same number of basis vectors. The number of basis vectors in the basis of a subspace is called the dimension of the subspace.

5. Let $\mathbb{W} \subseteq \mathcal{R}^n$ be a subspace whose dimension is m i.e. $dim(\mathbb{W}) = m$. Then, any collection of m linearly independent vectors constitutes a basis. (Recollect the arguments covered in the class. We'll comeback to this proof a little later in the course.)

6. Let $\mathbb{W} \subseteq \mathcal{R}^n$ be a subspace whose dimension is m i.e. $dim(\mathbb{W}) = m$. Then, any collection $\mathbb{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_{m+1}\}$ is a linearly dependent set.

<u>Proof by contradiction</u>: Let $\mathbb{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_{m+1}\} \in \mathbb{W}$ be a collection of N-length Linearly independent vectors. Since $\mathbb{U}$ constitutes a LI set of $m+1$ vectors, one can from the basis of the subspace $\mathbb{W}$ using

m elements form $\mathbb{U}$. Let $\mathbb{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$. Since the basis spans entire $\mathbb{W}$, the vector $\mathbf{u}_{m+1} \in \mathbb{W}$ can be expressed as a linear combination of $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$, which contradicts the initial assumption that $\mathbb{U}$ is a collection of linearly independent vectors.

EE2100: Matrix Analysis

Review Notes - 9

Topics covered :

1. Representation of a vector in orthogonal basis

---

1. Representation of a vector in a given basis: Let $\mathbb{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_N\}$ denote the basis of a subspace $\mathbb{W}$ whose dimension is N. Since the basis spans $\mathbb{W}$, any vector $\mathbf{x} \in \mathbb{W}$ can be represented as a linear combination of the basis vectors i.e.,

$$\mathbf{x} = \sum_{i=1}^{N} \alpha_i \mathbf{b}_i \tag{1}$$

The coefficients of the linear combination [i.e., α_i's in (1)] that result in a linear combination of $\vec{x}$ in terms of basis $\mathbb{B}$ are called the coordinates of a vector $\mathbf{x}$ relative to basis $\mathbb{B}$. The collection (or ordered set) of all coordinates of a vector in basis $\mathbb{B}$ is called the coordinate vector of $\mathbf{x}$ relative to basis $\mathbb{B}$ and is denoted by $\mathbf{x}_B$. The coordinate vector of $\mathbf{x}$ relative to basis $\mathbb{B}$ is commonly referred to as representation of a vector in basis $\mathbf{B}$.

2. Let $\mathbb{W}$ be a subspace of dimension N (or for simplicity assume that $\mathbb{W} = \mathcal{R}^N$). Then the collection of vectors $\{\mathbf{e}_1, \cdots, \mathbf{e}_N\}$ constitute the basis of the subspace (Proof is quite trivial) and is commonly referred to as the standard basis. The coordinate vector of any vector $\mathbf{x}$ in the subspace is the vector $\mathbf{x}$ itself.

3. An other commonly explored basis is various fields is the basis of orthogonal vectors. Any collection of N-orthogonal vectors in a given subspace are linearly independent and hence can be a basis of N-dimensional subspace.

Proof of Linear independence: Let $\mathbb{U} = \{\mathbf{u}_1, \cdots \mathbf{u}_N\}$ be a collection of orthogonal vectors. The collection is linearly independent if and only if

$$\sum_{i=1}^{N} \alpha_i \mathbf{u}_i = \mathbf{0} \text{ when } \alpha_i = 0 \;\forall\; i \tag{2}$$

Taking the inner product of the linear combination in (2) with $\mathbf{u}_1$ results in

$$\alpha_1 \|\mathbf{u}_1\|_2^2 + \alpha_2 \underbrace{\mathbf{u}_1 \cdot \mathbf{u}_2}_{0} + \cdots + \alpha_N \underbrace{\mathbf{u}_1 \cdot \mathbf{u}_N}_{0} = \mathbf{u}_1 \cdot \mathbf{0} \implies \alpha_1 = 0 \tag{3}$$

In a similar manner, it is possible to show that the linear combination in (2) goes to zero only when $\alpha_i = 0 \;\forall\; i$, indicating that the collection of the vectors are linearly independent.

4. Representation of a vector in orthogonal basis: Let $\mathbb{B}$ be an orthogonal basis of an N-dimensional subspace $\mathbb{W}$. Any vector $\mathbf{x} \in \mathbb{W}$ can be written as a linear combination of the basis i.e.,

$$\mathbf{x} = \sum_{i=1}^{N} \alpha_i \mathbf{b}_i. \tag{4}$$

The components of the vector $\mathbf{x}$ in the basis B can be computed by using the orthogonality property of the basis vectors i.e.,

$$\mathbf{b}_k \cdot \mathbf{x} \ = \ \mathbf{b}_k \cdot \sum_{i=1}^{N} \alpha_i \mathbf{b}_i \implies \alpha_k = \frac{\mathbf{b}_k \cdot \mathbf{x}}{\|\mathbf{b}_k\|_2^2} \tag{5}$$

EE2100: Matrix Analysis

Review Notes - 10

Topics covered :

1. Grahm Schmidt Algorithm
2. Projection of a Vector onto Subspace

---

1. Grahm Schmidt Algorithm: Given a collection of vectors, Grahm Schmidt Algorithm can generate a collection of vectors that are orthogonal to each other. If the given collection of vectors (say m) are linearly independent, then the Grahm Schmidt algorithm will generate a set of orthogonal vectors of same size (i.e., m). If any of the vectors in the given collection are linearly dependent, then the Grahm Schmidt algorithm will result in a zero vector, which can in turn be used to detect the present of linearly dependent vectors.

2. **Problem definition**: Let $\{\mathbf{v}_1, \cdots \mathbf{v}_m\}$ of length n be the spanning set of a subspace $\mathbb{W}$. Generate an orthogonal basis, say $\{\mathbf{b}_1, \cdots \mathbf{b}_m\}$ for the subspace $\mathbb{W}$.

   One way to generate the orthogonal basis $\{\mathbf{b}_1, \cdots \mathbf{b}_m\}$ from the spanning set $\{\mathbf{v}_1, \cdots \mathbf{v}_m\}$ is by using the Grahm Schmidt algorithm. The Grahm Schmidt algorithm relies on the idea of projection. The basic idea is as follows. Generate the i^{th} vector such that it does not have component along any of the $i-1$ vectors in the set (look at the review lecture for finer details).

Algorithm 1 Grahm Schmidt Algorithm

```
b_1 = v_1
for i = 2 to m do
    b_i = v_i
    for j = 1 to i − 1 do
        b_i = b_i − Proj_{b_j} v_i
    end for
    if b_i = 0 then
        The set {v_1, ··· v_m} is a spanning set and not a basis. Reapply the Grahm Schmidt Algorithm on {v_1, ··· v_m} \ v_i
    end if
end for
```

3. Projection of a Vector onto Subspace: Consider a subspace $\mathbb{W}$ whose basis is $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$. Let $\mathbf{v}$ be a vector that is not in the subspace $\mathbb{W}$ (i.e., $\mathbf{v} \notin \mathbb{W}$). The vector (say $\mathbf{w} \in \mathbb{W}$) in the subspace with the shortest distance to $\mathbf{v}$ is referred to as the projection of $\mathbf{v}$ onto $\mathbb{W}$ (denoted by $\mathbf{Proj}_{\mathbb{W}}\mathbf{v}$).

4. Theorem: If $\mathbf{w}$ is the projection of $\mathbf{v}$ onto the subspace $\mathbb{W}$, then $\mathbf{v} - \mathbf{w} \perp \mathbb{W}$.

   <u>Proof:</u> Consider a vector $\mathbf{y} \in \mathbb{W}$. Since $\mathbf{w} \in \mathbb{W}$ is the projection of $\mathbf{v}$, the minimum value of $\|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|$ (or $\|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|^2$) occurs when $\beta = 0$. The value of $\|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|^2$ can be computed as

$$\begin{aligned} \|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|^2 &= (\mathbf{v} - \mathbf{w} + \beta\mathbf{y}) \cdot (\mathbf{v} - \mathbf{w} + \beta\mathbf{y}) \\ &= \|\mathbf{v} - \mathbf{w}\|^2 + 2\beta(\mathbf{v} - \mathbf{w}) \cdot \mathbf{y} + \beta^2\|\mathbf{y}\|^2 \end{aligned} \tag{1}$$

At the minimum value, (i.e., at $\beta = 0$) the derivative of $\|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|^2$ with respect to β should vanish, i.e.,

$$\begin{aligned} \frac{d}{d\beta}\left(\|\mathbf{v} - \mathbf{w} + \beta\mathbf{y}\|^2\right) &= 0 \\ 2(\mathbf{v} - \mathbf{w}) \cdot \mathbf{y} + 2\beta\|\mathbf{y}\|^2 &= 0 \\ (\mathbf{v} - \mathbf{w}) \cdot \mathbf{y} &= 0 \text{ since } \beta = 0 \text{ and } \|\mathbf{y}\|^2 > 0 \end{aligned} \tag{2}$$

Equation (2) indicates that the vector $\mathbf{v} - \mathbf{w}$ is orthogonal to any vector $\mathbf{y}$ in the subspace or equivalently, the vector $\mathbf{v} - \mathbf{w}$ is orthogonal to $\mathbb{W}$.

EE2100: Matrix Analysis

Review Notes - 11

Topics covered :

1. Projection of a Vector onto a Subspace
2. Introduction to $K-$means clustering

---

1. Theorem: If $\mathbf{w}$ is the projection of $\mathbf{v}$ onto the subspace $\mathbb{W}$, then $\mathbf{v} - \mathbf{w} \perp \mathbb{W}$.

2. Computing the projection of a vector onto the subspace: The projection of vector $\mathbf{v} \notin \mathbb{W}$ onto a subspace $\mathbf{W}$ with orthogonal basis $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ is given by

$$\mathbf{Proj}_{\mathbb{W}}\mathbf{v} = \mathbf{w} = \sum_{i=1}^{m} \mathbf{Proj}_{\mathbf{b}_i}\mathbf{v} \tag{1}$$

Proof: Let $\mathbf{w} \in \mathbb{W}$ be the projection of $\mathbf{v}$ onto the subspace $\mathbb{W}$. The vector $\mathbf{w}$ can be written in terms of orthogonal basis as

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{b}_i \tag{2}$$

The coefficients α_i can be computed by utilizing the fact that $\mathbf{v} - \mathbf{w} \perp \mathbb{W}$ i.e., $\mathbf{v} - \mathbf{w} \perp \mathbf{b}_j \ \forall j = 1, \cdots, m$ i.e.,

$$\begin{aligned} (\mathbf{v} - \mathbf{w}) \cdot \mathbf{b}_j &= 0 \\ \mathbf{w} \cdot \mathbf{b}_j &= \mathbf{v} \cdot \mathbf{b}_j \\ \alpha_j (\mathbf{b}_j \cdot \mathbf{b}_j) &= \mathbf{v} \cdot \mathbf{b}_j \\ \alpha_j &= \frac{\mathbf{v} \cdot \mathbf{b}_j}{\mathbf{v}_j \cdot \mathbf{b}_j} = \frac{1}{\|\mathbf{b}_j\|}\mathbf{v} \cdot \mathbf{e}_{\mathbf{b}_j} \end{aligned} \tag{3}$$

Substituting (3) in (2) gives

$$\mathbf{w} = \sum_{j=1}^{m} \frac{\mathbf{v} \cdot \mathbf{e}_{\mathbf{b}_j}}{\|\mathbf{b}_j\|}\mathbf{b}_j = \sum_{j=1}^{m} \mathbf{Proj}_{\mathbf{b}_j}\mathbf{v} \tag{4}$$

3. Grahm Schmidt Algorithm for Computing the Projection of a vector onto subspace: The projection of vector $\mathbf{v} \notin \mathbb{W}$ onto a subspace $\mathbf{W}$ with basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ can be computed by applying Grahm Schmidt algorithm on the set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}\}$.

Proof: The basis of the subspace and $\mathbf{v}$ form a collection of linearly independent vectors and hence the Grahm Schmidt algorithm will generate a collection of orthogonal vectors. Let $\{\mathbf{b}_1, \cdots, \mathbf{b}_{m+1}\}$ indicate the collection of orthogonal vectors generated by the Grahm Schmidt algorithm. The vector $\mathbf{b}_{m+1}$ is given by

$$\begin{aligned} \mathbf{b}_{m+1} &= \mathbf{v} - \mathbf{Proj}_{\mathbf{b}_1}\mathbf{b} - \cdots - \mathbf{Proj}_{\mathbf{b}_m}\mathbf{b} \\ &= \mathbf{v} - \sum_{i=1}^{m} \mathbf{Proj}_{\mathbf{b}_i}\mathbf{v} \end{aligned} \tag{5}$$

The vector $\mathbf{b}_{m+1}$ is orthogonal to every vector in the collection $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$. As a result, any vector generated from the linear combination of $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ (i.e., span($\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$)) is orthogonal to $\mathbf{b}_{m+1}$. Taking into account the fact that $\mathbf{Span}\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$=$\mathbf{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ (which is the subspace $\mathbb{W}$), it can be concluded that the vector $\mathbf{b}_{m+1}$ is orthogonal to subspace $\mathbb{W}$. The projection of $\mathbf{b}$ onto the subspace $\mathbb{W}$ is thus

$$\mathbf{Proj}_{\mathbb{W}}\mathbf{v} = \sum_{i=1}^{m} \mathbf{Proj}_{\mathbf{b}_i}\mathbf{v} \tag{6}$$

K-Means Clustering

Problem :- $X = \{\vec{x}_1, \dots \vec{x}_m\} \in \mathbb{R}^n$

$\Downarrow$

form K clusters.

① Centroid | group identifier :-

$$\vec{\mu}_k : \min \overbrace{\sum_{i \in X_k} \|\vec{x}_i - \vec{\mu}_k\|^2}.$$

$$\vec{\mu}_k = \frac{1}{|X_k|} \sum_{i \in X_k} \vec{x}_i$$

② grouping :-

$$x \in X_i \quad \text{where} \quad \arg\min_i (\|\vec{x} - \vec{\mu}_i\|)$$

① Start with Random $\vec{\mu}_i$ $\forall i = 1, \dots k$.

② Repeat till Convergence

$\downarrow$ based on J.

a) grouping

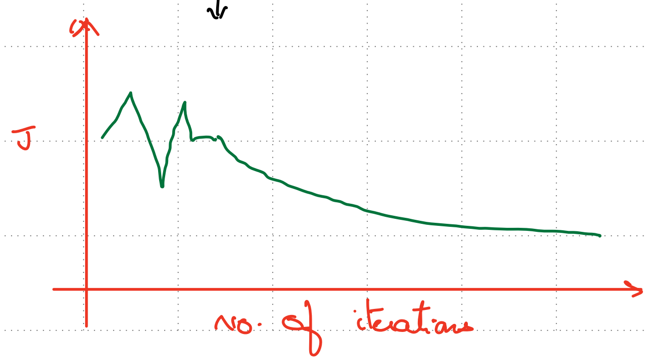
b) Centroid

if the elements in the group don't change $\Rightarrow$ $\vec{\mu}_i$ $\forall i = 1, \dots k$ does not change.

$\downarrow$

$$J_k = \sum_{i \in X_k} \|\vec{x}_i - \vec{\mu}_k\|^2$$

$$J = \sum_{k=1}^{K} J_k.$$



$K \in [a, b]$

$J_{K \in [a,b]}$

optional no. of clusters } J to be minimum

EE2100: Matrix Analysis

Review Notes - 13

Topics covered :

1. Introduction to matrices
2. Elementary matrix operations

---

1. A matrix can be thought of as a rectangular array of numbers or a mathematical representation of data in tabular form. Matrices are commonly written between a rectangular brackets. Matrix can also be thought of as a way of representing a collection of vectors which share the same properties or of same length. Matrices are denoted by upper case letters (A) or upper case bold letters ($\mathbf{A}$).

2. An important attribute of a matrix is the size or dimension of the matrix and is commonly written as $m \times n$, where m is the number of rows and n is the number of columns. The rows of a matrix are the number of horizontal arrays in the matrix and the number of columns are the number of vertical arrays in the matrix.

3. Rows are indexed from top to bottom while the columns are indexed from left to right. The elements (or) entries (or) coefficients of a matrix are the values in the array. If $\mathbf{A}$ is a matrix, then the entry in the ith row and jth column is commonly denoted by $A_{i,j}$. The entries of the matrix where the row index is equal to the column index are called the diagonal entries (i.e., A_{ii}).

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1i} & \cdots & A_{1j} & \cdots & A_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{i1} & A_{i2} & \cdots & A_{ii} & \cdots & A_{ij} & \cdots & A_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{j1} & A_{j2} & \cdots & A_{ji} & \cdots & A_{jj}; & \cdots & A_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mi} & \cdots & A_{mj} & \cdots & A_{mn} \end{pmatrix} \tag{1}$$

4. It is often convenient to analyze matrix as a collection of row and column vectors.

$$\mathbf{A} = \begin{array}{c c} \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1i} & \cdots & A_{1j} & \cdots & A_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{i1} & A_{i2} & \cdots & A_{ii} & \cdots & A_{ij} & \cdots & A_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{j1} & A_{j2} & \cdots & A_{ji} & \cdots & A_{jj}; & \cdots & A_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mi} & \cdots & A_{mj} & \cdots & A_{mn} \end{pmatrix} & \begin{array}{c} \\ \\ \vec{a}_i \\ \\ \\ \\ \\ \end{array} \\ \begin{array}{cccccccc} & & & & & \vec{b}_j & & \end{array} & \end{array} \tag{2}$$

For example, the matrix A in (2) can be represented as a collection of row vectors i.e.,

$$\mathbf{A} = \left\{\mathbf{a}_1^T, \mathbf{a}_2^T, \cdots, \mathbf{a}_i^T, \cdots \mathbf{a}_m^T\right\} \tag{3}$$

where the vector $\mathbf{a}_i$ (see Eq. 2) is a vector whose entries are the entries of the ith row. When representing matrix as a collection of row vectors, the length of each vector is n.

Alternatively, the matrix A in (2) can be represented as a collection of column vectors i.e.,

$$\mathbf{A} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_j, \cdots \mathbf{b}_n\} \tag{4}$$

where the vector $\mathbf{b}_j$ (see Eq. 2) is a vector whose entries are the entries of the jth column. When representing matrix as a collection of column vectors, the length of each vector is m.

5. Some of the matrices commonly encountered in the analysis are

   (a) Zero matrix (commonly denoted by $\mathbf{0}_{m\times n}$): A $m \times n$ matrix where each entry is 0.

   (b) Square matrix : A matrix where the number of rows is equal to number of columns.

   (c) Tall matrix : A matrix where the number of rows are greater then the number of columns.

   (d) Wide/fat matrix: A matrix where the number of columns are greater than the number of rows.

   (e) Identity matrix (commonly denoted by $\mathbf{I}_{m\times m}$): A $m \times m$ matrix where the diagonal entries are 1 and all the other elements are 0.

   $$\mathbf{I}_{m\times m} : I_{ij} = \begin{cases} 1 & \forall i = j \\ 0 & i \neq j \end{cases} \tag{5}$$

   (f) Diagonal matrix (commonly denoted by $\mathbf{D}_{m\times m}$): A $m \times m$ matrix where at least 1 diagonal entry is non-zero and all the other elements are 0.

   $$\mathbf{D}_{m\times m} : d_{ij} = \begin{cases} \neq 0 & \text{for at least one } i = j \\ 0 & i \neq j \end{cases} \tag{6}$$

   (g) Lower triangular matrix (commonly denoted by $\mathbf{L}_{m\times n}$): A $m \times n$ matrix where all entries above the diagonal are all 0.

   $$\mathbf{L}_{m\times n} : d_{ij} = \begin{cases} 0 & \forall\ i > j \\ x & i \geq j \end{cases} \quad \text{where } x \in \mathcal{R} \tag{7}$$

   (h) Upper triangular matrix (commonly denoted by $\mathbf{U}_{m\times n}$): A $m \times n$ matrix where all entries below the diagonal are all 0.

   $$\mathbf{U}_{m\times n} : d_{ij} = \begin{cases} 0 & \forall\ i < j \\ x & i \geq j \end{cases} \quad \text{where } x \in \mathcal{R} \tag{8}$$

6. Matrix addition: Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices of same size (say $m \times n$), then the matrix addition is defined

as

$$\mathbf{C} \ = \ \mathbf{A} + \mathbf{B} := C_{ij} = A_{ij} + B_{ij} \tag{9}$$

The matrix addition satisfies the commutative and associative property.

7. Scalar multiplication: Let $\mathbf{A}$ be an $m \times n$ and $\alpha \in \mathcal{R}$, then the scalar multiplication is defined as

$$\mathbf{C} \ = \ \alpha\mathbf{A} := C_{ij} = \alpha A_{ij} \tag{10}$$

8. Transpose of a matrix: Let $\mathbf{A}$ be an $m \times n$ matrix, then the transpose (denoted by $\mathbf{A}^T$) is an $n \times m$ matrix given by

$$\left(\mathbf{A}^T\right)_{ij} = \mathbf{A}_{ji} \tag{11}$$

A matrix whose transpose is equal itself (i.e. $\mathbf{A}^T = \mathbf{A}$) is called a Symmetric matrix and a matrix which satisfies the property $\mathbf{A}^T = -\mathbf{A}$ is called the Skew-Symmetric matrix.

9. If matrices are treated as objects where its entries come from a Field $\mathbb{F}$ with the addition and multiplication operators defined as (9) and (10) respectively, then the collection of all possible matrices of size $m \times n$ constitutes a vector space over the field. These vector spaces are often denoted as $\mathbb{F}^{m \times n}$.

EE2100: Matrix Analysis

Review Notes - 14

Topics covered :

1. Matrix Vector Product
2. Matrix as a Transformation

---

1. A mathematical way to denote a row vector with m entries (say $\mathbf{a}$) is $\mathbf{a} \in \mathcal{R}^{1\times m}$.

2. **Matrix vector product**: Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{x} \in \mathcal{R}^n$, then the product of matrix $\mathbf{Ax}$ (recollect the arguments about compatibility) is a vector of length m (say $\mathbf{y} \in \mathcal{R}^m$) whose entries are given by

$$y_i = \sum_{i=1}^{n} A_{ik}x_k \tag{1}$$

If $\mathbf{a}_i \in \mathcal{R}^n$ denotes a vector such that its j^{th} entry is $A_{i,j}$ (i.e., the elements of i^{th} row of $\mathbf{A}$ are the elements of $\mathbf{a}_i$), then, the ith entry of the matrix vector product can be obtained as.

$$y_i = \mathbf{a}_i \cdot \mathbf{x} \text{ or } \mathbf{a}_i^T\mathbf{x} \tag{2}$$

In terms of column vectors (say b_i which indicates a vector whose entries are the entries of the ith column of $\mathbf{A}$), the vector $\mathbf{y}$ can be written as (recollect the arguments covered in the class)

$$\mathbf{y} = \sum_{i=1}^{n} x_i\mathbf{b}_i \tag{3}$$

The result of the matrix vector product is a vector that can be seen as a linear combination of column vectors of $\mathbf{A}$ and the coefficients of linear combinations are the entries of $\mathbf{x}$.

Examples:

- Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{e}_i \in \mathcal{R}^n$, then, the result of $\mathbf{Ae}_i$ is the i^{th} column of $\mathbf{A}$.
- Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{x} \in \mathcal{R}^n$ such that

$$x_i = \begin{cases} \alpha & i = p \\ \beta & i = q \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The result of $\mathbf{Ax}$ is $\alpha\mathbf{a}_p + \beta\mathbf{a}_q$, where $\mathbf{a}_p$ and $\mathbf{a}_q$ are the column vectors corresponding to columns p and q of $\mathbf{A}$ respectively.

3. **Matrix vector product - Another form**: Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{x} \in \mathcal{R}^{m}$, then the product $\mathbf{x}^T\mathbf{A}$ (recollect the arguments about compatibility) is a vector of length n (say $\mathbf{y} \in \mathcal{R}^n$) whose entries are given by

$$y_i = \sum_{i=1}^{m} x_k A_{ki} \tag{5}$$

Instead if $\mathbf{x}$ is a row vector i.e., $\mathbf{x} \in \mathcal{R}^{1\times m}$, then the product $\mathbf{xA}$ is also defined in the same manner i.e.,

$$y_i = \sum_{i=1}^{m} x_k A_{ki} \tag{6}$$

In terms of column vectors (say b_i which indicates a vector whose entries are the entries of the ith column of $\mathbf{A}$), then the entries of vector $\mathbf{y}$ can be written as

$$y_i = \mathbf{x}^T\mathbf{b}_i \tag{7}$$

If $\mathbf{a}_i \in \mathcal{R}^n$ denotes a vector such that its j^{th} entry is $A_{i,j}$ (i.e., the elements of i^{th} row of $\mathbf{A}$ are the elements of $\mathbf{a}_i$), then,

$$\mathbf{y} = \sum_{i=1}^{m} x_i \mathbf{a}_i^T \tag{8}$$

The result of the matrix vector product is a vector that can be seen as a linear combination of transpose of row vectors of $\mathbf{A}$.

Examples:

- Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{e}_i \in \mathcal{R}^m$, then, the result of $\mathbf{e}_i^T\mathbf{A}$ is the i^{th} row of $\mathbf{A}$.
- Let $\mathbf{A} \in \mathcal{R}^{m\times n}$ and $\mathbf{x} \in \mathcal{R}^m$ such that

$$x_i = \begin{cases} \alpha & i = p \\ \beta & i = q \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

  The result of $\mathbf{x}^T\mathbf{A}$ is $\alpha\mathbf{a}_p^T + \beta\mathbf{a}_q^T$, where $\mathbf{a}_p \in \mathcal{R}^n$ and $\mathbf{a}_q \in \mathcal{R}^n$ are the transpose of row vectors corresponding to columns p and q of $\mathbf{A}$, respectively.

4. **Function of Vectors** (Introduction to the idea of matrix as a transformation): Let $\mathbf{x} \in \mathcal{R}^n$ be a vector in subspace $\mathbb{W}$. Let $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ denote a function that operates on a vector in $\mathbb{W}$. Some examples of functions are

   (a) The L_2 norm of a vector is a function that can be denoted as $T(\mathbf{x}) : \mathcal{R}^n \rightarrow \mathcal{R}$ where $T(\mathbf{x}) = \sqrt{\sum_{i=1}^{n} x_i^2}$.

   (b) A hypothetical function that operates on a function $\mathbf{x}$ and returns a vector of length 3 whose entries are the maximum, minimum and average value of $\mathbf{x}$ respectively. The function can be denoted as $T(\mathbf{x}) : \mathcal{R}^n \rightarrow \mathcal{R}^3$.

The subspace over which the function is defined (or) the subspace of vectors on which the function operates (i.e., takes as inputs) is called the domain. The range is a collection of outputs for every vector in the domain. It is a common practice in linear algebra to use the term Transformation instead of function (reasons will be discussed in the next lecture).

5. **Linear function/Transformation**: A Transformation $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is a linear transformation if and only if it satisfies the following properties.
   (a) For any $\mathbf{x},\ \mathbf{y} \in \mathcal{R}^n$, $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$, and
   (b) For any $\mathbf{x} \in \mathcal{R}^n$ and $\alpha \in \mathcal{R}$, $T(\alpha\mathbf{x}) = \alpha T(\mathbf{x})$.
6. Let $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ be a linear transformation. Then the result of the transformation when applied to a vector $\mathbf{x}$ can be obtained using a matrix-vector product $\mathbf{Ax}$. The matrix $\mathbf{A}$ is often referred to as the transformation matrix corresponding to transformation T..
7. **Deriving the transformation matrix**: Let $\mathbf{x}$ be a vector of a subspace whose basis is $\{\mathbf{b}_1, \cdots, \mathbf{b}_n\}$. The vector $\mathbf{x}$ in terms of the basis can be written as

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{b}_i \tag{10}$$

Let T be a linear transformation. The result of a linear transformation operating on $\mathbf{x}$ can be given as

$$T(\mathbf{x}) = T\left(\sum_{i=1}^{n} \alpha_i \mathbf{b}_i\right) = \sum_{i=1}^{n} \alpha_i T\left(\mathbf{b}_i\right) \tag{11}$$

Let $\mathbf{a}_i$ be the result of the transformation applied to basis vector $\mathbf{b}_i$. Accordingly,

$$T(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i = \mathbf{A}\boldsymbol{\alpha} \tag{12}$$

Equation 12 indicates that the transformation applied on $\mathbf{x}$ can be obtained as a matrix vector product.

EE2100: Matrix Analysis

Review Notes - 15

Topics covered :

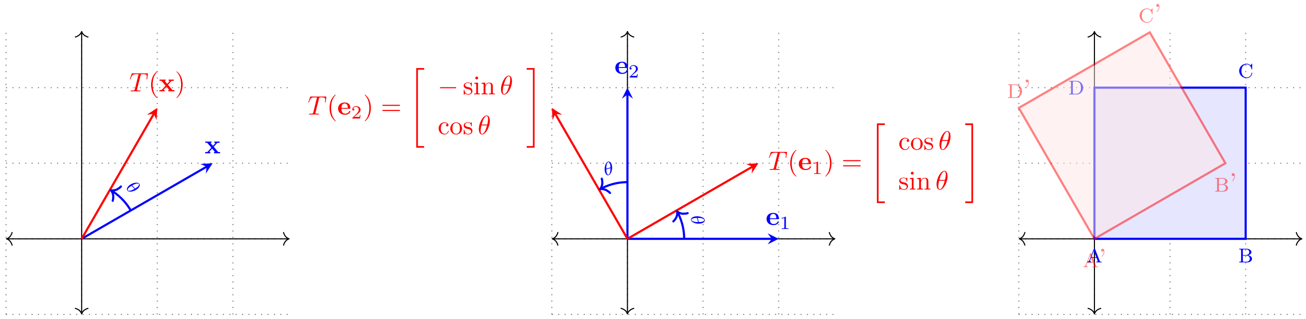
1. Matrix of a Linear Transformation

---

1. For a vector x represented in the standard basis, the linear transformation applied on the vector can be computed as

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} \text{ where } \mathbf{A} = [T(\mathbf{e}_1), \cdots, T(\mathbf{e}_n)] \tag{1}$$

2. Let $T : \mathcal{R}^2 \to \mathcal{R}^2$ denote a linear transformation that rotates a vector in $\mathcal{R}^2$ by a fixed angle θ. The result of such a transformation (i.e., rotation by $\frac{\pi}{6}$ in counter clockwise direction) when applied to a vector and object (input shown in blue and output in red) in $\mathcal{R}^2$ is shown in Fig. 1. The transformation matrix



(a) Vector (b) Basis vectors (c) Shape

Fig. 1: Illustration of rotation

corresponding to the rotation operation (see Fig. 1(c)) is given by

$$\mathbf{A} = [T(\mathbf{e}_1), T(\mathbf{e}_2)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{2}$$

3. Let $T : \mathcal{R}^2 \to \mathcal{R}^2$ denote a linear transformation that projects a vector in $\mathcal{R}^2$ onto a given vector (say $\mathbf{v}$). Let $\mathbf{e}_v = [e_{v1}, e_{v2}]^T$ denote the direction/unit vector of $\mathbf{v}$. The result of such a transformation when applied to a vector and object in $\mathcal{R}^2$ is shown in Fig. 2 (input shown in blue and output in red). The transformation matrix corresponding to the reflection operation (see Fig. 2(c)) is given by

$$\mathbf{A} = [T(\mathbf{e}_1), T(\mathbf{e}_2)] = \begin{bmatrix} e_{v1}e_{v1} & e_{v2}e_{v1} \\ e_{v1}e_{v2} & e_{v2}e_{v2} \end{bmatrix} \tag{3}$$

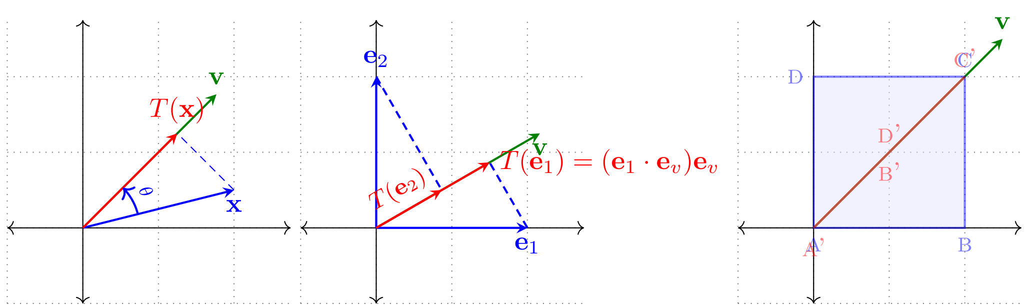


Fig. 2: Illustration of Projection

EE2100: Matrix Analysis

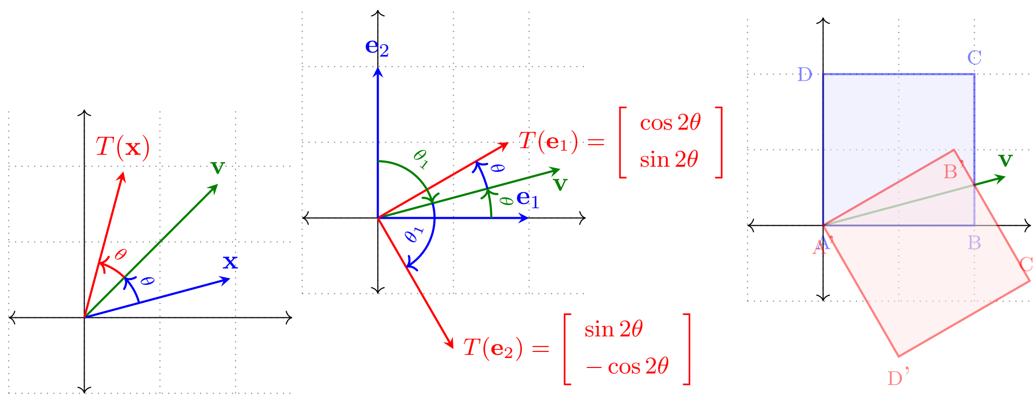
Review Notes - 16

Topics covered :

1. Transformation Matrices
2. The idea of inverse of a Matrix

---

1. Let $T : \mathcal{R}^2 \rightarrow \mathcal{R}^2$ denote a linear transformation that reflects a vector in $\mathcal{R}^2$ with respect to a given vector (say $\mathbf{v}$). The result of such a transformation when applied to a vector and object (input shown in blue and output in red) in $\mathcal{R}^2$ is shown in Fig. 1. The transformation matrix corresponding to the reflection



(a) Vector (b) Basis vectors (c) Shape

Fig. 1: Illustration of reflection

operation (see Fig. 1(c)) is given by

$$\mathbf{A} = [T(\mathbf{e}_1), T(\mathbf{e}_2)] = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \tag{1}$$

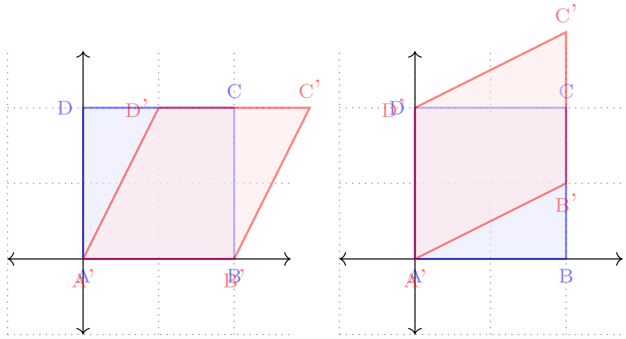
2. In general, any matrix can be considered as a linear transformation. The transformation matrices in $\mathcal{R}^2$ of the form

$$\mathbf{A}_x = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \text{ or } \mathbf{A}_y = \begin{bmatrix} 1 & 0 \\ y & 1 \end{bmatrix} \tag{2}$$

are called shear transformations.

3. Inverse of a matrix - Intuition: Let $\mathbf{A} \in \mathcal{R}^{n \times n}$ be a real square matrix. The matrix $\mathbf{A}$ can be thought of as a collection of column vectors $\mathbf{a}_i$ what are the output of some linear transformation (say $T : \mathcal{R}^n \rightarrow \mathcal{R}^n$) applied to the standard basis vectors i.e.,

$$\mathbf{A} \;=\; [T\,(\mathbf{e}_1)\;\; T\,(\mathbf{e}_2)\;\; \cdots\;\; T\,(\mathbf{e}_n)] \tag{3}$$



(a) Transformation $\mathbf{A}_x$ (b) Transformation $\mathbf{A}_y$

Fig. 2: Illustration of Shear Transformation

The inverse transformation (say T_1, which can be shown to be linear) can be thought of as a transformation that operates on the column vectors of $\mathbf{A}$ and result in the standard basis i.e., if $T(\mathbf{e}_i) = \mathbf{a}_i$, then the inverse transformation (say denoted by T_1) should satisfy $T_1(\mathbf{a}_i) = \mathbf{e}_i$. Accordingly, the inverse transformation must satisfy the following equations

$$T_1(\mathbf{a}_i) = \mathbf{e}_i \ \forall i \leq n \tag{4}$$

The transformation matrix corresponding to Linear transformation T_1 (say $\mathbf{B}$) is called the inverse of the matrix i.e.,

$$\mathbf{B} = [T_1(\mathbf{e}_1) \ T_1(\mathbf{e}_2) \ \cdots \ T_1(\mathbf{e}_n)] \tag{5}$$

4. Will be covered in the next lecture: The set of equations given by (4) can be alternately represented as

$$[\mathbf{T}_1(\mathbf{a}_1) \ \mathbf{T}_1(\mathbf{a}_2) \cdots \ \mathbf{T}_1(\mathbf{a}_n)] = [\mathbf{e}_1 \ \mathbf{e}_2 \cdots \ \mathbf{e}_n] \tag{6}$$

An alternate way to represent (6) is

$$\mathbf{BA} = \mathbf{I} \tag{7}$$

Equation 7 indicates that the $\mathbf{B}$ is an inverse of a matrix $\mathbf{A}$ if $\mathbf{BA} = \mathbf{I}$.

EE2100: Matrix Analysis

Review Notes - 17

Topics covered :

1. Matrix Multiplication

---

1. The multiplication operation on two matrices is defined if and only of the number of columns of the first operand is equal to the number of rows of the second operand. For example, if $\mathbf{A} \in \mathcal{R}^{m \times n}$ and $\mathbf{B} \in \mathcal{R}^{p \times q}$ denote two matrices, then $\mathbf{AB}$ is defined if and only if $n = p$ and $\mathbf{BA}$ is only defined if and only if $q = m$. Notice that, the conditions for $\mathbf{AB}$ and $\mathbf{BA}$ to be defined are different and hence, the fact that $\mathbf{AB}$ is defined does guarantee that $\mathbf{BA}$ is also defined.

2. The multiplication operation between two matrices $\mathbf{A} \in \mathcal{R}^{m \times n}$ and $\mathbf{B} \in \mathcal{R}^{n \times p}$ results in a matrix $\mathbf{C} \in \mathcal{R}^{m \times p}$ whose entries are given by

$$C_{ij}(\text{entry corresponding to } i^{th} \text{ row and } j^{th} \text{ column of } \mathbf{C}) = \mathbf{a}_i \cdot \mathbf{b}_j = a_i^T \mathbf{b}_j \tag{1}$$

where $\mathbf{a}_i$ is a vector whose entries are the entries corresponding to row i of matrix $\mathbf{A}$ and $\mathbf{b}_j$ is the column vector corresponding to column j of matrix $\mathbf{B}$.

3. **Multiplication with special kind of matrices**: Let $\mathbf{A} \in \mathcal{R}^{m \times n}$ be a matrix of real entries. The operation $\mathbf{BA}$ where $\mathbf{B} \in \mathcal{R}^{p \times m}$ is often referred to as pre-multiplying the matrix $\mathbf{A}$ with matrix $\mathbf{B}$. Similarly, the operation $\mathbf{AC}$ where $\mathbf{C} \in \mathcal{R}^{n \times p}$ is often referred to as post-multiplying the matrix $\mathbf{A}$ with matrix $\mathbf{C}$.

   - **Pre-multiplying with a diagonal matrix**: Let $\mathbf{D} \in \mathcal{R}^{m \times m}$ denote a diagonal matrix with D_{ii} indicating the diagonal entry corresponding to row i. The result of Pre-multiplying $\mathbf{A}$ with $\mathbf{D}$ i.e., $\mathbf{DA}$ is a matrix whose rows are scaled versions of rows of matrix $\mathbf{A}$. The scaling factor corresponding to each row is D_{ii} i.e.,

$$\begin{bmatrix} D_{11} & 0 & \cdots & 0 \\ 0 & D_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D_{mm} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} D_{11}A_{11} & D_{11}A_{12} & \cdots & \cdots & D_{11}A_{1n} \\ D_{22}A_{21} & D_{22}A_{22} & \cdots & \cdots & D_{22}A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{mm}A_{m1} & D_{mm}A_{m2} & \cdots & \cdots & D_{mm}A_{mn} \end{bmatrix} \tag{2}$$

   - **Post multiplying with a diagonal matrix**: Let $\mathbf{D} \in \mathcal{R}^{n \times n}$ denote a diagonal matrix with D_{ii} indicating the diagonal entry corresponding to row i. The result of Post-multiplying $\mathbf{A}$ with $\mathbf{D}$ i.e., $\mathbf{AD}$ is a matrix whose columns are scaled versions of columns of matrix $\mathbf{A}$. The scaling factor corresponding to each column is D_{ii} i.e.,

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & \cdots & A_{mn} \end{bmatrix} \begin{bmatrix} D_{11} & 0 & \cdots & \cdots & 0 \\ 0 & D_{22} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & D_{nn} \end{bmatrix} = \begin{bmatrix} D_{11}A_{11} & D_{22}A_{12} & \cdots & \cdots & D_{nn}A_{1n} \\ D_{11}A_{21} & D_{22}A_{22} & \cdots & \cdots & D_{nn}A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{11}A_{m1} & D_{22}A_{m2} & \cdots & \cdots & D_{nn}A_{mn} \end{bmatrix} \tag{3}$$

- **Pre-multiplying with a special kind of lower triangular matrix** [To be covered in Tutorial 5]: Let $\mathbf{L}_p \in \mathcal{R}^{n\times n}$ denote a lower triangular matrix whose entries are given by

$$(L_p)_{ii} = \begin{cases} 1 & \text{for } i \le p \\ \alpha & \text{for } i > p \end{cases} \quad \text{and } (L_p)_{ji} = \begin{cases} 0 & \text{for } j < i \\ -\beta_j & \text{for } j > p \text{ and } i = p \\ 0 & \text{otherwise} \end{cases} \qquad \text{[see (6), as an example]} \tag{4}$$

Pre-multiplying a matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ with $\mathbf{L_p}$, results in elementary row operations of the following form.

$$\mathcal{R}_j : \alpha\mathcal{R}_j - \beta_j\mathcal{R}_p \quad \forall j > p \text{ i.e.,} \tag{5}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \vdots & 0 \\ 0 & 0 & \cdots & -\beta_{p+1} & \alpha & \cdots & 0 \\ 0 & 0 & \cdots & -\beta_{p+2} & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\beta_n & 0 & \cdots & \alpha \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} & A_{1,p+1} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2p} & A_{2,p+1} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} & A_{p,p+1} & \vdots & A_{p,n} \\ A_{p+1,1} & A_{p+1,2} & \cdots & A_{p+1,p} & A_{p+1,p+1} & \vdots & A_{p+1,n} \\ A_{p+2,1} & A_{p+2,2} & \cdots & A_{p+2,p} & A_{p+2,p+1} & \vdots & A_{p+2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{np} & A_{n,p+1} & \cdots & A_{nn} \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} & A_{1,p+1} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2p} & A_{2,p+1} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} & A_{p,p+1} & \vdots & A_{p,n} \\ \alpha A_{p+1,1} - \beta_{p+1}A_{p1} & \alpha A_{p+1,2} - \beta_{p+1}A_{p2} & \cdots & \alpha A_{p+1,p} - \beta_{p+1}A_{pp} & \alpha A_{p+1,p+1} - \beta_{p+1}A_{p,p+1} & \vdots & \alpha A_{p+1,n} - \beta_{p+1}A_{pn} \\ \alpha A_{p+2,1} - \beta_{p+2}A_{p1} & \alpha A_{p+2,2} - \beta_{p+2}A_{p2} & \cdots & \alpha A_{p+2,p} - \beta_{p+2}A_{pp} & \alpha A_{p+2,p+1} - \beta_{p+2}A_{p,p+1} & \vdots & \alpha A_{p+2,n} - \beta_{p+2}A_{pn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha A_{n,1} - \beta_n A_{p1} & \alpha A_{n,2} - \beta_n A_{p2} & \cdots & \alpha A_{n,p} - \beta_n A_{pp} & \alpha A_{n,p+1} - \beta_n A_{p,p+1} & \vdots & \alpha A_{n,n} - \beta_n A_{pn} \end{bmatrix} \tag{6}$$

EE2100: Matrix Analysis

Review Notes - 18

Topics covered :

1. Solutions to systems of Linear equations

---

1. Consider a system of m linear equations with n unknowns (see (1), where i^{th} unknown is denoted by x_i).

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= b_m \end{aligned} \tag{1}$$

The system of linear equations can be represented as

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}} \tag{2}$$

where the matrix consists of ordered set of coefficients of the linear equation and is of size $m \times n$ (often denoted as $\mathbf{A} \in \mathcal{R}^{m \times n}$). The ordered set of unknowns constitute the vector $\mathbf{x} \in \mathcal{R}^n$ and the ordered set of values/constants constitute the vector $\mathbf{b} \in \mathcal{R}^m$.

2. If the matrix is treated as a collection of column vectors $\{\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n\}$ where $\mathbf{a}_i \in \mathcal{R}^m$, the set of linear equations can be represented as

$$\sum_{i=1}^{n} x_i \mathbf{a}_i = \mathbf{b} \tag{3}$$

3. Equation (3) indicates that the process of determining the values of unknowns (x_i) is the process of determining the coefficients that result in a linear combination (of column vectors, i.e., $\mathbf{a}_i$) of $\mathbf{b}$. This way of treating a matrix, i.e., as a collection of column vectors, gives handy insights into the nature of the solution for a given system of linear equations.

4. If we treat the matrix $\mathbf{A}$ as a collection of column vectors, then, the following inferences can be made about the solution to the given system of linear equations.

   - The system of linear equations will have no solution if $\mathbf{b}$ is not in the span of the column vectors i.e., $\mathbf{b} \notin \text{Span}\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$.

- The system of linear equations will have at least one solution (covers both the possibilities, i.e., unique solution or infinitely many solutions) if $\mathbf{b}$ is in the span of the column vectors i.e., $\mathbf{b} \in \text{Span}\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$.
- If the column vectors are linearly independent, then any vector in $\mathbf{b} \in \text{Span}\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$ can be represented uniquely as a linear combination of $\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$. In terms of solution of system of linear equations, an unique solution exists if $\mathbf{b} \in \text{Span}\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$ and $\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$ are linearly independent.
- If the column vectors are linearly dependent, then there is at least one non zero collection of coefficients that take the linear combination of column vectors to $\mathbf{0}$. In terms of matrix, this implies that, there is at least one vector (say $\mathbf{x}_z$) that satisfies $\mathbf{A}\mathbf{x}_z = \mathbf{0}$.
- Let the column vectors of $\mathbf{A}$ be linearly dependent with $\mathbf{b} \in \text{Span}\{\mathbf{a}_1 \cdots, \mathbf{a}_n\}$. Let $\mathbf{x}_p$ denote one solution to the set of linear equations i.e., $\mathbf{A}\mathbf{x}_p = \mathbf{b}$. Further, let $\mathbf{x}_z$ denote the vector that satisfies $\mathbf{A}\mathbf{x}_z = \mathbf{0}$. Then, $\mathbf{x}_p + \alpha\mathbf{x}_z$, where $\alpha \in \mathcal{R}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. Stated alternately, the system of linear equations has infinitely many solutions.

To summarize, the span of the column vectors of a matrix and the collection of coefficients that result in linear combination of $\mathbf{0}$ play a crucial role in determining the nature of solutions for $\mathbf{A}\mathbf{x} = \mathbf{b}$. These two are effectively captured by the fundamental subspaces of a matrix.

5. **Four fundamental subspaces of a matrix**: Let $\mathbf{A}$ denote the matrix under consider. The four fundamental subspaces are
   - The **Column space** of a matrix is the span of the column vectors of the matrix. It is often denoted by $\mathbf{Col}(\mathbf{A})$.
   - The **Null space** of a matrix is the set of vectors that satisfy $\mathbf{A}\mathbf{x} = \mathbf{0}$ (i.e., the collection of coefficients that results in a linear combination of column vectors to be $\mathbf{0}$). It is often denoted by $\mathbf{Null}(\mathbf{A})$. One vector that is always an element of the null space of any matrix is the $\mathbf{0}$ (which may not be of interest). Furthermore, if $\mathbf{x}_z$ is an element $\mathbf{Null}(\mathbf{A})$, the $\alpha\mathbf{x}_z \in \mathbf{Null}(\mathbf{A})\ \forall\alpha \in \mathcal{R}$. In this course, the Null space of a matrix is denoted as a span of linearly independent vectors $\mathbf{x}_z$ such that $\mathbf{A}\mathbf{x}_z = \mathbf{0}$ i.e.m, $\mathbf{Null}(\mathbf{A}) = \mathbf{Span}\{\text{Linearly independent } \mathbf{x}_z \neq \mathbf{0} : \mathbf{A}\mathbf{x}_z = \mathbf{0}\}$.
   - The **Row space** of a matrix is the span of the row vectors of the matrix. It is often denoted by $\text{Row}(\mathbf{A})$. The row space is also the column space of the transpose of the matrix.
   - The **Left Null space** of a matrix is the set of vectors that satisfy $\mathbf{A}^T\mathbf{x} = \mathbf{0}$

Analyzing the nature of solution/s using the fundamental subspaces

(a) Consider a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{A} \in \mathcal{R}^{n\times n}$, $\mathbf{x}, \mathbf{b} \in \mathcal{R}^n$. Comment on the nature of the solution/s if the column vectors are the column vectors of $\mathbf{A}$ are linearly independent.

Sol. Since, the column vectors of $\mathbf{A}$ (n in total) are linearly independent, they constitute the basis of $\mathcal{R}^n$. Hence, it is possible to realize every $\mathbf{b} \in \mathcal{R}^n$ uniquely as a linear combination of the column vectors of $\mathbf{A}$. Hence, $\forall\mathbf{b} \in \mathcal{R}^n$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution and $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

Rank-Nullity Theorem

Let $\{\vec{a}_1, \ldots, \vec{a}_n\} \in \mathbb{R}^m$ be spanning set of W.

r are Linearly independent $\Rightarrow \{\underbrace{\vec{a}_1, \ldots \vec{a}_r}_{LI}, \underbrace{\vec{a}_{r+1}, \ldots \vec{a}_n}_{LD}\}$

$$\text{Basis of } W = \{\vec{a}_1, \ldots \vec{a}_r\} \Rightarrow \vec{a}_j = \sum_{i=1}^{r} \alpha_{j,i} \vec{a}_i \quad j = r+1, \ldots n. \longrightarrow (1)$$

$\dim(W) = r$

$\exists$ only $(n-r)$ Linearly independent Co-ordinate Vectors $\vec{z}_i = \begin{bmatrix} z_{i,1} \\ \vdots \\ z_{i,n} \end{bmatrix} \in \mathbb{R}^n$. Such that $\sum_{j=1}^{n} z_{i,j} \vec{a}_j = \vec{0} \longrightarrow (2)$

$(\vec{z}_1, \ldots \vec{z}_{n-r}) \in \mathbb{R}^n$ Such that $\sum_{j=1}^{n} z_{i,j} \vec{a}_j = \vec{0}$

from (1) $\Rightarrow \vec{a}_{r+1} = \sum_{i=1}^{r} \alpha_{r+1,i} \vec{a}_i \Rightarrow \vec{a}_{r+1} - \sum_{i=1}^{r} \alpha_{r+1,i} \vec{a}_i = \vec{0}$

$$\Downarrow$$

$$\vec{z}_1 = \begin{bmatrix} -\alpha_{r+1,1} \\ \vdots \\ -\alpha_{r+1,r} \\ \hline 1 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^n \text{ takes the LC of } \{\vec{a}_1, \ldots \vec{a}_n\} = \vec{0}$$

Rank-Nullity Theorem

$$\vec{a}_{r+2} = \sum_{i=1}^{r} \alpha_{r+2,i}\, \vec{a}_i \Rightarrow \vec{a}_{r+2} - \sum_{i=1}^{r} \alpha_{r+2,i}\, \vec{a}_i = \vec{0}$$

$$\Downarrow$$

$$\vec{z}_2 = \begin{bmatrix} -\alpha_{r+2,1} \\ \vdots \\ -\alpha_{r+2,r} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n \text{ takes the LC of } \{\vec{a}_1, \ldots \vec{a}_n\} = \vec{0}$$

We can build LI vectors $\{\vec{z}_1, \ldots \vec{z}_{n-r}\}$ such that $\sum_{j=1}^{n} z_{i,j}\, \vec{a}_j = \vec{0}$

Part - B:- If $\vec{z} \notin \{\vec{z}_1, \ldots \vec{z}_{n-r}\}$ & still satisfies $\sum_{j=1}^{n} z_j\, \vec{a}_j = \vec{0}$, then $\vec{z} \in \text{Span}(\vec{z}_1, \ldots \vec{z}_{n-r}) \in \mathbb{R}^n$.

Rank-Nullity Theorem

$A \in \mathbb{R}^{m\times n} = [\vec{a}_1, \ldots\ldots, \vec{a}_n]$, with $\vec{a}_i \in \mathbb{R}^m$.

$\underbrace{\vec{a}_1, \ldots}_{r \text{ are LI.}}$

$\exists\ (n-r)$ LI vectors $\vec{z}_i$: $\underbrace{\sum_{j=1}^{n} z_{i,j}\vec{a}_j}_{} = \vec{0} \longrightarrow$ (1)

$\Downarrow$

$\exists\ (n-r)$ LI vectors $\vec{z}$: $A\vec{z} = \vec{0} \longrightarrow$ (2)

$\Downarrow$

Null space of $A = \text{Span}\{\vec{z}_1, \ldots \vec{z}_{n-r}\}$

Nullity (A) = No. of Linearly independent Vectors in Null (A)

$= n - r.$

Rank (A) = No. of Linearly independent Column Vectors in A.

$= r = \dim(\text{Span}(\vec{a}_1, \ldots \vec{a}_n)) = \dim(\text{Col}(A))$

$\rightarrow$ **Rank(A) + Nullity(A) = n.** $\leftarrow$

Let $A \in \mathbb{R}^{n\times n}$

$\Downarrow$

if $\vec{a}_i$'s of A are LI,

$\Downarrow$

Rank(A) = n.

Nullity(A) = 0.

$\Downarrow$

$\forall \vec{b} \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has

a unique solution.

Rank-Nullity Theorem

Let $A = [\underbrace{\vec{a}_1, \dots \vec{a}_r}_{r}, \underbrace{\vec{a}_{r+1}, \dots \vec{a}_n}_{(n-r)}] \in \mathbb{R}^{m \times n}$ Col(A) $\in \mathbb{R}^m$

Null (A) $\in \mathbb{R}^n$.

$$\text{Rank}(A) + \underbrace{\text{Nullity}(A)}_{} = n$$

Shown so far :- $(\vec{z}_1, \dots \vec{z}_{n-r}) \in \mathbb{R}^n$ & are LI such that $A\vec{z}_i = \vec{0}$

To be shown :- Any $\vec{v} : A\vec{v} = \vec{0}$ must be a LC of $(\vec{z}_1, \dots \vec{z}_{n-r})$

Proof :- $A = [\overbrace{\vec{a}_1, \dots \vec{a}_r}^{A_1 \in \mathbb{R}^{m \times r}}, \overbrace{\vec{a}_{r+1}, \dots \vec{a}_n}^{A_2 \in \mathbb{R}^{m \times (n-r)}}]$

$$\underbrace{\vec{a}_{r+1} = \overbrace{\sum_{i=1}^{r} \alpha_{r+1,i}}\, \vec{a}_i}_{} \Rightarrow \vec{y}_1 = \begin{bmatrix} \alpha_{r+1,1} \\ \vdots \\ \alpha_{r+1,r} \end{bmatrix} \in \mathbb{R}^r \Rightarrow \vec{a}_{r+1} = A_1 \vec{y}_1$$

$$\vdots$$

$$\vec{a}_n = A_1 \vec{y}_{n-r}$$

$$A_2 = [A_1\vec{y}_1, \dots A_1\vec{y}_{n-r}] = A_1 Y \quad \text{where } Y = [\vec{y}_1, \dots \vec{y}_{n-r}] \in \mathbb{R}^{r \times (n-r)}$$

Rank-Nullity Theorem

$$A_1 Y = \left(\mathbb{R}^{m\times r}\right)\left(\mathbb{R}^{r\times(n-r)}\right)$$

$$A = [A_1 \quad A_1 Y] \longrightarrow (1)$$

Let $\vec{v}$: $A\vec{v} = \vec{0} \Rightarrow [A_1 \quad A_1 Y]\,\vec{v} = \vec{0}$ $\qquad \vec{v} = \begin{bmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{bmatrix}$, $\vec{v}_1 \in \mathbb{R}^r$, $\vec{v}_2 \in \mathbb{R}^{n-r}$. $\longrightarrow (2)$

$$[A_1 \quad A_1 Y]\begin{bmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{bmatrix} = \vec{0} \in \mathbb{R}^m.$$

$$\underset{m\times r}{A_1}\vec{v}_1 + A_1 Y \vec{v}_2 = \vec{0}$$

$$A_1(\underbrace{\vec{v}_1 + Y\vec{v}_2}_{\vec{x}}) = \vec{0} \Rightarrow \vec{v}_1 + Y\vec{v}_2 = \vec{0} \Rightarrow \vec{v}_1 = -Y\vec{v}_2. \longrightarrow (3)$$

Any $\vec{v}$: $A\vec{v} = \vec{0}$ is of the form $\begin{bmatrix} -Y\vec{v}_2 \\ \hline \vec{v}_2 \end{bmatrix} = \begin{bmatrix} -Y \\ \hline I \end{bmatrix}\vec{v}_2$, $\begin{bmatrix} -Y \\ \hline I \end{bmatrix} \in \mathbb{R}^{n\times(n-r)}$, Y: $r\times n-r$, $I \in \mathbb{R}^{(n-r)\times(n-r)}$

$$= \begin{array}{c} r\{ \\ \\ \end{array}\begin{bmatrix} -\vec{y}_1, & -\vec{y}_2, & \cdots & -\vec{y}_{n-r} \\ \hline 1 & 0 & & 0 \\ 0 & 1 & & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & 1 \end{bmatrix}$$

$$= [\vec{z}_1, \cdots\cdots \vec{z}_{n-r}]$$

Any $\vec{v}$: $A\vec{v} = 0$ is of the form $\underbrace{[\vec{z}_1, \cdots \vec{z}_{n-r}]}_{\mathbb{R}^{n\times n-r}}\,\vec{v}_2$, $\vec{v}_2 \to n-r$

$$= \sum_{i=1}^{n-r} v_{2,i}\,\vec{z}_i$$

EE2100: Matrix Analysis

Review Notes - 20

Topics covered :

1. Gaussian Elimination

---

1. Consider a system of m linear equations with n unknowns (see (1), where i^{th} unknown is denoted by x_i).

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= b_m \end{aligned} \tag{1}$$

The system of linear equations can be represented as

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}} \tag{2}$$

where the matrix consists of ordered set of coefficients of the linear equation and is of size $m \times n$ (often denoted as $\mathbf{A} \in \mathcal{R}^{m\times n}$). The ordered set of unknowns constitute the vector $\mathbf{x} \in \mathcal{R}^n$ and the ordered set of values/constants constitute the vector $\mathbf{b} \in \mathcal{R}^m$.

2. The commonly used approach to solve a system of linear equations is the Gaussian Elimination. The method of Gaussian Elimination relies on elementary row operations, i.e., the set of operations that do not effect the solution. The three elementary row operations are:

   - Interchanging the rows of the matrix **A** (i.e., changing the order in which the linear equations are handled) and the corresponding entries in **b**.
   - Scaling the row/s of the matrix **A** and the corresponding entry/entries in **b**.
   - Modifying the row/s of the matrix (and the corresponding entries in **b**) by taking linear combination with other row/s. This is the most commonly used elementary operation in the forward elimination phase of the Gaussian Elimination and is typically represented as $R_i : \alpha R_i - \beta R_j$. The expression $R_i : \alpha R_i - \beta R_j$ implies that Row i of **A** and **b** are modified as a linear combination of rows i and j, where the coefficients of linear combination are α and $-\beta$ respectively. It is important to keep in mind that the linear combination operation must also be applied for elements in **b**.

3. **Solving $\mathbf{Ux} = \mathbf{b}$ (Idea of back substitution)**: Consider a system of linear equations which correspond to the from $\mathbf{Ux} = \mathbf{b}$, where $\mathbf{U}$ is an upper triangular square matrix of size $n \times n$, i.e.,

$$\begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1,n-1} & U_{1n} \\ 0 & A_{22} & \cdots & U_{2n} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & U_{n-1,n-1} & U_{n-1,n} \\ 0 & 0 & \cdots & 0 & U_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \tag{3}$$

The value of x_n can be computed by writing the linear equation corresponding to the last row as

$$U_{nn}x_n = b_n \implies x_n = \frac{1}{U_{nn}}[b_n] \tag{4}$$

Since the value of x_n is computed, the value of x_{n-1} can be computed from the linear equation corresponding to row $n-1$ as

$$U_{n-1,n-1}x_{n-1} + U_{n-1,n}x_n = b_{n-1} \implies x_{n-1} = \frac{1}{U_{n-1,n-1}}[b_{n-1} - U_{n-1,n}x_n] \tag{5}$$

Subsequently, since the value of x_n and x_{n-1} is computed, the value of x_{n-2} can be computed from the linear equation corresponding to row $n-2$ as

$$U_{n-2,n-2}x_{n-2}+U_{n-2,n-1}x_{n-1}+U_{n-2,n}x_n = b_{n-2} \implies x_{n-2} = \frac{1}{U_{n-2,n-2}}[b_{n-2} - U_{n-2,n-1}x_{n-1} - U_{n-2,n}x_n] \tag{6}$$

The process can be extended to compute the values of all the unknown variables. In general, value of variable x_j can be computed as

$$x_j = \frac{1}{U_{jj}}\left[b_j - \sum_{i=j+1}^{n} U_{ji}x_i\right] \tag{7}$$

This approach of solving a system of linear equations of the form $\mathbf{Ux} = \mathbf{b}$ is commonly referred to as back substitution.

4. **Gaussian Elimination**: The approach of Gaussian Elimination used for solving systems of linear equations has two key phases. The first phase, commonly referred to as the Forward Elimination phase reduces a given matrix to an upper triangular matrix using elementary row operations. The first phase, commonly referred to as the Back substitution computes the value of unknown variables.

5. In the forward elimination phase, starting from the first row till row $n-1$, we apply elementary row operations of the form $R_i : \alpha R_i - \beta R_j$ to make all the entries below the diagonal to 0 (since the aim is to modify the matrix as an Upper triangular matrix). The elementary row operation needed to make the entry A_{ji} (where $j > i$) below the diagonal entry A_{ii} to 0 is given by $R_j : A_{ii}R_j - A_{ji}R_i$. It is important to keep in mind that the linear combination operation must also be applied for elements in $\mathbf{b}$.

6. The steps involved in the Gaussian Elimination are summarized in Algorithm 1.

Algorithm 1 Gaussian Elimination

procedure GE($\mathbf{A}, \mathbf{b}$) ▷ The inputs are the matrix $\mathbf{A}$ and vector $\mathbf{b}$
    for $i = 1$ to $n - 1$ **do** ▷ Forward Elimination
        for $j = i + 1$ to n **do** ▷ To make $A_{ji} = 0 \forall j > i$
            Apply the elementary row operation $R_j : A_{ii}R_j - A_{ji}R_i$
        end for
    end for
    $\mathbf{U} \rightarrow$modified matrix at the end of FE.
    for $j = n$ to 1 **do** ▷ Back Substitution
        Compute x_j as $x_j = \frac{1}{U_{jj}} \left[b_j - \sum_{i=j+1}^{n} U_{ji}x_i \right]$.
    end for
end procedure

Example of Gaussian Elimination

1. Consider a system of linear equations given by

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{8}$$

Compute the values of unknown $\mathbf{x}$ using Gaussian Elimination.

Sol. The first step in the forward elimination phase is to make the entries A_{21} and A_{31} to 0. The elementary row operation and the linear equations after performing the corresponding elementary row operations are

$$\left.\begin{array}{c} \underbrace{\begin{array}{ll} R_2 & : 2R_2 - R_1 \\ R_3 & : 2R_3 - R_1 \end{array}}_{\text{Elementary row operations}} \end{array}\right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{9}$$

The second step in the forward elimination phase is to make the entry A_{32} to 0.

$$\left. R_3 \quad : 3R_3 - 5R_2 \right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}. \tag{10}$$

At this stage, the values of unknown variables can be computed using back substitution as

$$\begin{aligned} 4x_3 = -2 &\implies x_3 = -\frac{1}{2} \\ 3x_2 + x_3 = 1 &\implies x_2 = \frac{1}{2} \\ 2x_1 + x_2 + x_3 = 1 &\implies x_1 = \frac{1}{2} \end{aligned} \tag{11}$$

EE2100: Matrix Analysis

Review Notes - 21

Topics covered :

1. Identifying nature of solution/s from Gaussian Elimination
2. Computing Null space of a Matrix

---

1. **Detecting Infinitely many solutions**: Consider a system of linear equations given by

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}. \tag{1}$$

A close look at the column vectors of the matrix indicates that they are linearly dependent i.e., $\mathbf{a}_3 = \mathbf{a}_1 - \mathbf{a}_2$. Further, the matrix has two linearly independent columns and hence the rank is 2. By rank nullity theorem, one can conclude that the matrix has one linearly independent non-zero vector that satisfies $\mathbf{Ax} = 0$. We will now attempt to solve the given system of linear equations using Gaussian Elimination.

Sol. The first step in the forward elimination phase is to make the entries A_{21} and A_{31} to 0. The elementary row operation and the linear equations after performing the corresponding elementary row operations are

$$\left.\underbrace{\begin{matrix} R_2 & : 2R_2 - R_1 \\ R_3 & : 2R_3 - R_1 \end{matrix}}_{\text{Elementary row operations}}\right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 5 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}. \tag{2}$$

The second step in the forward elimination phase is to make the entry A_{32} to 0.

$$\left. R_3 \quad : 3R_3 - 5R_2 \right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix}. \tag{3}$$

The linear equation corresponding to the last row is of the form $0x = 0$, which indicates that the variable x_3 can take any value and hence the system of equations have infinitely many solutions. One way to represent all the solutions of the given system of linear equations is by using parametric representation. In this case, let $x_3 = \alpha$, then using back substitution, the parametric variation of other unknown variables can be computed as

$$\begin{aligned} 3x_2 - 3x_3 = 3 &\implies x_2 = 1 + \alpha \\ 2x_1 + x_2 + x_3 = 1 &\implies x_1 = 1 - \alpha \end{aligned} \tag{4}$$

The non-zero linearly independent vector in the null space of a matrix can be computed by solving $\mathbf{Ax} = \mathbf{0}$, i.e.,

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5}$$

The first step in the forward elimination phase is to make the entries A_{21} and A_{31} to 0. The elementary row operation and the linear equations after performing the corresponding elementary row operations are

$$\left.\underbrace{\begin{matrix} R_2 & : 2R_2 - R_1 \\ R_3 & : 2R_3 - R_1 \end{matrix}}_{\text{Elementary row operations}}\right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 5 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6}$$

The second step in the forward elimination phase is to make the entry A_{32} to 0.

$$R_3 \quad : 3R_3 - 5R_2 \Big\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{7}$$

Let $x_3 = \alpha$, then using back substitution, the parametric variation of other unknown variables can be computed as

$$\begin{aligned} 3x_2 - 3x_3 = 0 &\implies x_2 = \alpha \\ 2x_1 + x_2 + x_3 = 0 &\implies x_1 = -\alpha \end{aligned} \tag{8}$$

2. **Detecting presence of No solution**: Consider a system of linear equations given by

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{9}$$

Sol. The first step in the forward elimination phase is to make the entries A_{21} and A_{31} to 0. The elementary row operation and the linear equations after performing the corresponding elementary row operations are

$$\left.\underbrace{\begin{matrix} R_2 & : 2R_2 - R_1 \\ R_3 & : 2R_3 - R_1 \end{matrix}}_{\text{Elementary row operations}}\right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 5 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{10}$$

The second step in the forward elimination phase is to make the entry A_{32} to 0.

$$\left.\begin{array}{ll} R_3 & : 3R_3 - 5R_2 \end{array}\right\} \implies \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}. \tag{11}$$

The linear equation corresponding to the last row is of the form $0x = v$ (where $v \neq 0$), which indicates that the equation is never satisfied irrespective of what ever values x_3 takes. Hence the system of linear equations do not have any solution.

3. To summarize, the occurrence of a zero row during the process of Gaussian Elimination is an indication that the system of linear equations have either no solution (if the corresponding entry in **b** is non-zero) or infinitely many solutions (if the corresponding entry in **b** is also zero).

.

$A = LU$

EE2100: Matrix Theory

Lecture 22: LU Decomposition

Topics :

1. LU Decomposition

LU Decomposition - Motivation

To solve $A\vec{x} = \vec{b}$ (Gaussian elimination)
- → Forward elimination ($U\vec{x} = \vec{3}$)
- → Back substitution

Scenario where $A\vec{x} = \vec{b}$ multiple no. of times

only $\vec{b}$ varies each time

DE: $\underbrace{a_n \frac{d^n x}{dt^2} + \cdots + a_1 \frac{dx}{dt}}_{A\vec{x} = \vec{b}} = f(x, t)$ (Temp)

$x(t)$

best way: $A = LU$ → do it once. $\Rightarrow (LU)\vec{x} = \vec{b}$

If we can compute inverse of L: $L^{-1}(LU)\vec{x} = L^{-1}\vec{b} \Rightarrow U\vec{x} = L^{-1}\vec{b} \to$ ②

① $(L_{n-1} \cdots L_1) A = U$

$FA = U$ where $F = L_{n-1} \cdots L_1$

a) F is lower triangular
b) inverse of F exists &

$F^{-1} = L$ (lower Δ^{lar})

To show

$\Downarrow$

$L(FA) = LU$.

$A = LU$

$\Rightarrow L^{-1} = L_{n-1} \cdots L_1$

$A\vec{x} = \vec{b}$ for 'm' different $\vec{b} \in \mathbb{R}^n$.

GE	LU
a) 'm' FE	a) LU decomposition & L^{-1} computation
b) 'm' BS	+
$m \times O(FE) + m \times O(BS)$	b) 'm' $O(L^{-1}\vec{b})$
	+
	c) 'm' $O(B.S)$

LU Decomposition - Example

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$R_2 : 2R_2 - R_1$

$R_3 : 2R_3 - R_1$

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

$R_3 : 3R_3 - 5R_2$

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \qquad \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

<u>Observations</u>:-

a) Set of elementary row operations during the FE phase are the same.

→ b) at the end of FE phase, $U\vec{x} = \vec{z}_1$ & $U\vec{x} = \vec{z}_2$ where 'U' is same. '$\vec{z}$' is the result of FE applied to $\vec{b}$

c) elementary row operations in FE phase can be rep as

$$(L_{n-1} \cdots L_2 L_1) A \vec{x} = (L_{n-1} \cdots L_1)\vec{b} \longrightarrow (1)$$

$$(L_{n-1} \cdots L_1) A = U$$

LU Decomposition - Generalization

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow A\vec{x} = \vec{b}$$

$R_2 : 2R_2 - R_1$

$R_3 : 2R_3 - R_1$

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Leftrightarrow L_1 A\vec{x} = L_1 \vec{b} \quad \text{where } L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

$R_3 : 3R_3 - 5R_2$

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \Leftrightarrow L_2(L_1 A\vec{x}) = L_2 L_1 \vec{b} \quad \text{when } L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 3 \end{bmatrix}$$

$L_2 L_1 A = U \longrightarrow (3)$

$L_2^{-1}(L_2 L_1)A = L_2^{-1} U = \quad L_1 A = L_2^{-1} U \longrightarrow (4)$

$\underbrace{L_1^{-1}}(L_1 A) = L_1^{-1} L_2^{-1} U \Rightarrow A = \underbrace{L_1^{-1} L_2^{-1}} U$

$A = LU \quad \& \quad L = L_1^{-1} L_2^{-1}$

Key: $\underbrace{L_2 L_1 A}\vec{x} = L_2 L_1 \vec{b}$

To show L_2^{-1} & L_1^{-1} exists & are lower Δ^{lar}

& $L_1^{-1} L_2^{-1}$ is also lower Δ^{lar}

LU Decomposition - Generalization

$A' \leftarrow$ $A'' \leftarrow$ $\leftarrow L_p \leftarrow \tilde{L}_p^{-1}$

Key :- given $A \in \mathbb{R}^{n \times n}$

$(L_{n-1} \cdots L_1) A = U$

$L_{n-1} \cdots L_1 = F$

a) F is also lower Δlar

b) inverse of F exists & is also lower Δlar $(F^{-1} = L)$

$L\,FA = LU$

$A = LU$

$(L_1^{-1} \cdots L_{n-1}^{-1})(L_{n-1} \cdots L_1) A = (L_1^{-1} \cdots L_{n-1}^{-1}) U$

a) $L_1^{-1} \cdots L_n^{-1}$ exists

b) $L_1^{-1} \cdots L_{n-1}^{-1}$ is also lower Δlar

$A = LU$ where $L = (L_1^{-1} \cdots L_{n-1}^{-1})$

Key Point : $A = LU$

but for solving $A\vec{x} = \vec{b}$ using LU, it is essential to compute L^{-1}

$L^{-1} = \underbrace{L_{n-1} \cdots L_1}$

L_1 is a matrix such that $A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}$

$L_1 A$ will be having the first column of the form $\alpha \vec{e}_1$

$$L_1 = \begin{bmatrix} 1 & \cdots & & 0 \\ -A_{21} & A_{11} & \cdots & 0 \\ \vdots & & \ddots & 0 \\ & 0 & & \\ -A_{n1} & & & A_{11} \end{bmatrix}$$

$L_1 A$ to have $A'_{21} = \cdots = A'_{n1} = 0$

$R_2 : A_{11} R_2 - A_{21} R_1$

$\vdots$

$R_n : A_{11} R_n - A_{n1} R_1$

$$L_1 A = A' = \begin{bmatrix} A'_{11} & \cdots & A'_{1n} \\ 0 & A'_{22} \cdots & A'_{2n} \\ \vdots & A'_{32} \cdots & A'_{3n} \\ 0 & & A'_{nn} \end{bmatrix}$$

$R_3 : A'_{22} R_3 - A'_{32} R_2$

$R_4 :$

$\vdots$

$R_n : A'_{22} R_n - A'_{n2} R_2$

$$L_2 = \begin{bmatrix} 1 & \cdots & & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & -A'_{32} & A'_{22} & 0 \\ \vdots & \vdots & \ddots & \\ 0 & -A'_{n2} & 0 & A'_{22} \end{bmatrix}$$

$L_p \leftarrow$ p^{th} step of FE

$(P = 1, \ldots, n-1)$

.

EE2100: Matrix Theory

Lecture 23: LU Decomposition

Topics :

1. $\mathbf{L}_p$ Matrix and its Inverse

LU Decomposition

$LU \cdot$ $\qquad L_{P} \cdot$

$A = LU$

$$\underline{L_{n-1} \cdots\cdots L_1} A = U \rightarrow ① \Rightarrow A = \underbrace{(L_1^{-1} \cdots\cdots L_{n-1}^{-1})}_{L \cdot} U \rightarrow (*)$$

A	$(A)_{i,j}$	$R_j \leftarrow$ Row 'j' of $A \in \mathbb{R}^{n\times n}$
$A_1 = L_1 A \in \mathbb{R}^{n\times n}$	$(A_1)_{i,j}$	$(R_1)_j \leftarrow$ Row 'j' of $A_1 \in \mathbb{R}^{n\times n}$
$A_2 = L_2 L_1 A$	$(A_2)_{i,j}$	$(R_2)_j \leftarrow$ Row 'j' of $A_2 \in \mathbb{R}^{n\times n}$
$\vdots$	$\vdots$	
$A_{n-1} = L_{n-1} \cdots L_1 A \cdot$	$(A_{n-1})_{i,j}$	$(R_{n-1})_j \leftarrow$ Row j of $A_{n-1} \in \mathbb{R}^{n\times n} \cdot$

$A\vec{x} = \vec{b}$ using LU

$L_1 A\vec{x} = L_1 \vec{b}$

$\vdots$

$U\vec{x} = \underbrace{(L_{n-1} \cdots L_1)}_{L^{-1}} \vec{b}$

$LU\vec{x} = \vec{b}$

$U\vec{x} = L^{-1}\vec{b}$

$$\underbrace{\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -(A)_{2,1} & (A)_{1,1} & \cdots & \cdots & 0 \\ \vdots & & \ddots & & 0 \\ \vdots & 0 & & \ddots & \\ -(A)_{n,1} & & & & (A)_{1,1} \end{bmatrix}}_{\tilde{L}_1 \in \mathbb{R}^{n\times n}} \underbrace{\begin{bmatrix} (A)_{1,1} & \cdots & \cdots & (A)_{1,n} \\ \vdots \; 0 & & & \\ (A)_{p,1} & \cdots & \cdots & (A)_{p,n} \\ \vdots & & & \\ (A)_{n,1} & \cdots & \cdots & (A)_{n,n} \end{bmatrix}}_{A \in \mathbb{R}^{n\times n}}$$

$R_j : (A)_{1,1} R_j - (A)_{j,1} R_1$

$\forall j > 1$

LU Decomposition

$$\underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \\ 0 & -(A_1)_{3,2}/(A_1)_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & -(A_1)_{n,2}/(A_1)_{2,2} & & \end{bmatrix}}_{L_2 \in \mathbb{R}^{n\times n}} \rightarrow \begin{bmatrix} (A_1)_{1,1} & (A_1)_{1,2} & \cdots & (A_1)_{1,n} \\ 0 & (A_1)_{2,2} & \cdots & (A_1)_{2,n} \\ \vdots & \vdots & & \vdots \\ 0 & (A_1)_{n,2} & & (A_1)_{n,n} \end{bmatrix}$$

$(R_1)_1 = (R)_1$

0

$(R_1)_j = (A_1)_{2,2} R_j - (A_1)_{j,2} (R)_2 \quad \forall j > 2.$

To find $L_p \in \mathbb{R}^{n\times n} \Rightarrow L_p A_{p-1} = A_p.$

$$(L_p)_{i,i} = \begin{cases} 1 & , i = 1, \ldots p \\ (A_{p-1})_{p,p} & , i = p+1, \ldots n. \end{cases}$$

$$(L_p)_{i,j} = \begin{cases} 0 & j > i \\ -(A_{p-1})_{i,j} & \text{when } j = p \text{ \& } i > j \\ 0 & \text{else.} \end{cases}$$

LU Decomposition

$L_1 \in \mathbb{R}^{n\times n} \rightarrow$ operates on rows of $A \longrightarrow$ rows of matrix A_1

$$(R)_1 = (R_1)_1 \longrightarrow 1$$

$$(A)_{1,1}(R)_j - (A)_{j,1} R_1 = (R_1)_j \quad \forall\, j>1 \longrightarrow 2$$

$L_1^{-1} \in \mathbb{R}^{n\times n} \rightarrow$ operates on rows of $A_1 \longrightarrow$ rows of matrix A.

a) Scale the first row of A_1 by 1. $\longrightarrow (R)_1$ of A.

b) $$(R)_j = \frac{1}{(A)_{1,1}}\left[(R_1)_j + (A)_{j,1}\underbrace{(R_1)_1}_{=(R)_1}\right]$$

$$(R)_j = \frac{1}{(A)_{1,1}}\left[(R_1)_j + (A)_{j,1}(R_1)_1\right] \quad \forall\, j>1.$$

$$L_1^{-1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \frac{(A)_{2,1}}{(A)_{1,1}} & \frac{1}{(A)_{1,1}} & & 0 & \\ \vdots & & \ddots & & \\ \vdots & 0 & & \ddots & \\ \frac{(A)_{n,1}}{(A)_{1,1}} & & & & \frac{1}{(A)_{1,1}} \end{bmatrix}$$

LU Decomposition

$$L_p^{-1} = \begin{cases} 1 & \text{for } i = 1, \dots, p \\ \dfrac{1}{(A_{p-1})_{p,p}} & i = p+1, \dots, n. \end{cases}$$

$$(L_p^{-1})_{i,j} = \begin{cases} 0 & j > i \\ \dfrac{(A_{p-1})_{i,j}}{(A_{p-1})_{p,p}} & \text{when } j = p \text{ \& } i > j \\ 0 & \text{else.} \end{cases}$$

To show :- Product of two lower Δ^{lar} matrices is also lower Δ^{lar}.

Proof :- A & $B \in \mathbb{R}^{n \times n}$ are lower Δ^{lar}

$C = AB$.

$C_{ij} = (i^{th}$ row of $A) \cdot (j^{th}$ Column of $B)$

To show that C is lower Δ^{lar}, $C_{ij} = 0 \;\forall\; j > i$

LU Decomposition

$$
\begin{bmatrix}
A_{11} & 0 & \cdots & \cdots & 0 \\
A_{21} & A_{22} & \cdots & \cdots & 0 \\
\vdots & & & & \vdots \\
A_{i1} & A_{i2} & \cdots & A_{ii} \cdots & 0 \\
\vdots & & & & \vdots \\
A_{j1} & A_{j2} & \cdots & A_{ji} \cdots A_{jj} & 0 \\
\vdots & & & & \vdots \\
A_{n1} & \cdots & \cdots & \cdots & A_{nn}
\end{bmatrix}
\begin{array}{c}
\begin{array}{ccccc} \phantom{B_{11}} & & i^{th} & j^{th} & \phantom{B_{nn}} \\ & & \downarrow & \downarrow & \end{array} \\
\begin{bmatrix}
B_{11} & \cdots & 0 & \cdots 0 \cdots & 0 \\
B_{21} & & 0 & 0 & \vdots \\
\vdots & & \vdots & \vdots & \vdots \\
B_{j1} & & B_{ji} & B_{jj} & \vdots \\
\vdots & & \vdots & \vdots & \vdots \\
B_{n1} & & B_{ni} & B_{nj} & B_{nn}
\end{bmatrix}
\end{array}
$$

i

$j > i$

$$
C_{ij} = A_{i1}\cancel{B_{j1}}^{\,0} + A_{i2}\cancel{B_{j2}}^{\,0} + \cdots + A_{ii}\cancel{B_{ji}}^{\,0} + \quad = 0 \qquad \forall j > i
$$