

LINEAR ALGEBRA -II

B V Rajarama Bhat

Indian Statistical Institute, Bangalore

References

- ▶ Basic references:

References

- ▶ Basic references:
- ▶ Linear Algebra, A. Ramachandra Rao and P. Bhimasankaram.

References

- ▶ Basic references:
- ▶ Linear Algebra, A. Ramachandra Rao and P. Bhimasankaram.
- ▶ Algebra, Michael Artin.

References

- ▶ Basic references:
- ▶ Linear Algebra, A. Ramachandra Rao and P. Bhimasankaram.
- ▶ Algebra, Michael Artin.
- ▶ Linear Algebra, Henry Helson.

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle
- ▶ Question : Consider the initial configuration as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{bmatrix}$$

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle
- ▶ Question : Consider the initial configuration as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{bmatrix}$$

- ▶ Can we re-arrange it to the natural order by moves permitted by the game?

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle
- ▶ Question : Consider the initial configuration as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{bmatrix}$$

- ▶ Can we re-arrange it to the natural order by moves permitted by the game?
- ▶ Some rich man offered a lot of money for people to come up with a solution.

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle
- ▶ Question : Consider the initial configuration as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{bmatrix}$$

- ▶ Can we re-arrange it to the natural order by moves permitted by the game?
- ▶ Some rich man offered a lot of money for people to come up with a solution.
- ▶ But there was no solution!

Lecture 1: Permutations

- ▶ Here is a well-known game, called fifteen puzzle:
- ▶ <https://lorecioni.github.io/fifteen-puzzle-game/>
- ▶ Information on the Wikipedia :
https://en.wikipedia.org/wiki/15_puzzle
- ▶ Question : Consider the initial configuration as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{bmatrix}$$

- ▶ Can we re-arrange it to the natural order by moves permitted by the game?
- ▶ Some rich man offered a lot of money for people to come up with a solution.
- ▶ But there was no solution!
- ▶ The Rubik's cube is a toy very you see a lot of 'permutations' in action.

Definition

- **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .

Definition

- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.2:** Take $S = \{s_1, s_2, \dots, s_n\}$.

Definition

- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.2:** Take $S = \{s_1, s_2, \dots, s_n\}$.
- ▶ Define $\sigma_1 : S \rightarrow S$ by

$$\sigma_1(s_j) = \begin{cases} s_{j+1} & \text{if } 1 \leq j < n \\ s_1 & \text{if } j = n. \end{cases}$$

Definition

- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.2:** Take $S = \{s_1, s_2, \dots, s_n\}$.
- ▶ Define $\sigma_1 : S \rightarrow S$ by

$$\sigma_1(s_j) = \begin{cases} s_{j+1} & \text{if } 1 \leq j < n \\ s_1 & \text{if } j = n. \end{cases}$$

- ▶ Assume $n \geq 3$. Define $\sigma_2 : S \rightarrow S$ by

$$\sigma_2(s_j) = \begin{cases} s_3 & \text{if } j = 1 \\ s_1 & \text{if } j = 3 \\ s_j & \text{otherwise.} \end{cases}$$

Definition

- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.2:** Take $S = \{s_1, s_2, \dots, s_n\}$.
- ▶ Define $\sigma_1 : S \rightarrow S$ by

$$\sigma_1(s_j) = \begin{cases} s_{j+1} & \text{if } 1 \leq j < n \\ s_1 & \text{if } j = n. \end{cases}$$

- ▶ Assume $n \geq 3$. Define $\sigma_2 : S \rightarrow S$ by

$$\sigma_2(s_j) = \begin{cases} s_3 & \text{if } j = 1 \\ s_1 & \text{if } j = 3 \\ s_j & \text{otherwise.} \end{cases}$$

- ▶ Then σ_1, σ_2 are permutations.

Notation

- ▶ One way to display the permutation is to write it down explicitly:

Notation

- ▶ One way to display the permutation is to write it down explicitly:
- ▶ We may show σ_1 above by:

$$\begin{pmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_1 \end{pmatrix}$$

Notation

- ▶ One way to display the permutation is to write it down explicitly:
- ▶ We may show σ_1 above by:

$$\begin{pmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_1 \end{pmatrix}$$

- ▶ Similarly σ_2 is displayed as:

$$\begin{pmatrix} s_1 & s_2 & s_3 & s_4 & \dots & s_n \\ s_3 & s_2 & s_1 & s_4 & \dots & s_n \end{pmatrix}$$

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .
- ▶ If σ, τ are elements of G , then $\sigma \circ \tau$ is also an element of G and the operation $\circ$ has the following properties:

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .
- ▶ If σ, τ are elements of G , then $\sigma \circ \tau$ is also an element of G and the operation $\circ$ has the following properties:
- ▶ (i) **Associativity:** $\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$ for all $\sigma_1, \sigma_2, \sigma_3$ in G .

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .
- ▶ If σ, τ are elements of G , then $\sigma \circ \tau$ is also an element of G and the operation $\circ$ has the following properties:
- ▶ (i) **Associativity:** $\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$ for all $\sigma_1, \sigma_2, \sigma_3$ in G .
- ▶ (ii) **Existence of identity:** There exists $\iota \in G$ such that $\iota \circ \sigma = \sigma \circ \iota = \sigma$ for all $\sigma \in G$.

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .
- ▶ If σ, τ are elements of G , then $\sigma \circ \tau$ is also an element of G and the operation $\circ$ has the following properties:
- ▶ (i) **Associativity:** $\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$ for all $\sigma_1, \sigma_2, \sigma_3$ in G .
- ▶ (ii) **Existence of identity:** There exists $\iota \in G$ such that $\iota \circ \sigma = \sigma \circ \iota = \sigma$ for all $\sigma \in G$.
- ▶ (iii) **Existence of inverse:** For $\sigma \in G$, there exists σ^{-1} in G such that $\sigma^{-1} \circ \sigma = \sigma \circ \sigma^{-1} = \iota$.

Group property

- ▶ **Theorem 1.3:** Let S be a finite set. Let G be the collection of all permutations of S .
- ▶ If σ, τ are elements of G , then $\sigma \circ \tau$ is also an element of G and the operation $\circ$ has the following properties:
- ▶ (i) **Associativity:** $\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3$ for all $\sigma_1, \sigma_2, \sigma_3$ in G .
- ▶ (ii) **Existence of identity:** There exists $\iota \in G$ such that $\iota \circ \sigma = \sigma \circ \iota = \sigma$ for all $\sigma \in G$.
- ▶ (iii) **Existence of inverse:** For $\sigma \in G$, there exists σ^{-1} in G such that $\sigma^{-1} \circ \sigma = \sigma \circ \sigma^{-1} = \iota$.
- ▶ **Proof.** Take ι as the identity map and then properties (i) to (iii) should be clear.

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

cycles

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.

- ▶ In the following we always take the finite set S under consideration as $\{1, 2, \dots, n\}$ for some fixed $n \in \mathbb{N}$.
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.
- ▶ For distinct $k_1, k_2, \dots, k_r$ in $\{1, 2, \dots, n\}$ (with $r \in \mathbb{N}$) we denote the cycle $k_1 \mapsto k_2 \mapsto \dots \mapsto k_r \mapsto k_1$ simply as $(k_1, k_2, \dots, k_r)$.

Cycle decomposition of permutations

- ▶ For any permutation σ , σ^2 denotes $\sigma \circ \sigma$ and more generally for any natural number r , $\sigma^r = \sigma \circ \sigma \circ \cdots \circ \sigma$ (r times).

Cycle decomposition of permutations

- ▶ For any permutation σ , σ^2 denotes $\sigma \circ \sigma$ and more generally for any natural number r , $\sigma^r = \sigma \circ \sigma \circ \cdots \circ \sigma$ (r times).
- ▶ **Lemma 1.5:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then for any $k_1 \in S$, there exists $r \in \mathbb{N}$ such that $\sigma^r(k_1) = k_1$.

Cycle decomposition of permutations

- ▶ For any permutation σ , σ^2 denotes $\sigma \circ \sigma$ and more generally for any natural number r , $\sigma^r = \sigma \circ \sigma \circ \cdots \circ \sigma$ (r times).
- ▶ **Lemma 1.5:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then for any $k_1 \in S$, there exists $r \in \mathbb{N}$ such that $\sigma^r(k_1) = k_1$.
- ▶ **Proof.** Take $k_2 = \sigma(k_1)$. If $k_2 = k_1$, we can take $r = 1$, and we are done.

Cycle decomposition of permutations

- ▶ For any permutation σ , σ^2 denotes $\sigma \circ \sigma$ and more generally for any natural number r , $\sigma^r = \sigma \circ \sigma \circ \cdots \circ \sigma$ (r times).
- ▶ **Lemma 1.5:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then for any $k_1 \in S$, there exists $r \in \mathbb{N}$ such that $\sigma^r(k_1) = k_1$.
- ▶ **Proof.** Take $k_2 = \sigma(k_1)$. If $k_2 = k_1$, we can take $r = 1$, and we are done.
- ▶ If $k_2 \neq k_1$, take $k_3 = \sigma(k_2)$. If $k_3 = k_1$, we can take $r = 2$ and we are done.

Cycle decomposition of permutations

- ▶ For any permutation σ , σ^2 denotes $\sigma \circ \sigma$ and more generally for any natural number r , $\sigma^r = \sigma \circ \sigma \circ \cdots \circ \sigma$ (r times).
- ▶ **Lemma 1.5:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then for any $k_1 \in S$, there exists $r \in \mathbb{N}$ such that $\sigma^r(k_1) = k_1$.
- ▶ **Proof.** Take $k_2 = \sigma(k_1)$. If $k_2 = k_1$, we can take $r = 1$, and we are done.
- ▶ If $k_2 \neq k_1$, take $k_3 = \sigma(k_2)$. If $k_3 = k_1$, we can take $r = 2$ and we are done.
- ▶ $k_3 = k_2$ is not possible, as this would mean that $k_2 = \sigma(k_1) = \sigma(k_2)$ and contradicting injectivity of σ .

Continuation

- ▶ Continuing this way, by induction if we get distinct $k_1, k_2, \dots, k_s$ with $k_1 \dashrightarrow k_2 \dashrightarrow \dots \dashrightarrow k_s$, we take $k_{s+1} = \sigma(k_s)$. If $k_{s+1} = k_1$, we can take $r = s$ and we are done.

Continuation

- ▶ Continuing this way, by induction if we get distinct $k_1, k_2, \dots, k_s$ with $k_1 \dashrightarrow k_2 \dashrightarrow \dots \dashrightarrow k_s$, we take $k_{s+1} = \sigma(k_s)$. If $k_{s+1} = k_1$, we can take $r = s$ and we are done.
- ▶ For any $2 \leq t \leq s$, $k_{s+1} = k_t = \sigma(k_{t-1})$ is not possible due to injectivity of σ .

Continuation

- ▶ Continuing this way, by induction if we get distinct $k_1, k_2, \dots, k_s$ with $k_1 \dashrightarrow k_2 \dashrightarrow \dots \dashrightarrow k_s$, we take $k_{s+1} = \sigma(k_s)$. If $k_{s+1} = k_1$, we can take $r = s$ and we are done.
- ▶ For any $2 \leq t \leq s$, $k_{s+1} = k_t = \sigma(k_{t-1})$ is not possible due to injectivity of σ .
- ▶ If $k_1, \dots, k_{s+1}$ are all distinct, we can continue the induction process.

Continuation

- ▶ Continuing this way, by induction if we get distinct $k_1, k_2, \dots, k_s$ with $k_1 \dashrightarrow k_2 \dashrightarrow \dots \dashrightarrow k_s$, we take $k_{s+1} = \sigma(k_s)$. If $k_{s+1} = k_1$, we can take $r = s$ and we are done.
- ▶ For any $2 \leq t \leq s$, $k_{s+1} = k_t = \sigma(k_{t-1})$ is not possible due to injectivity of σ .
- ▶ If $k_1, \dots, k_{s+1}$ are all distinct, we can continue the induction process.
- ▶ This process has to terminate after some steps as the set S is finite. In fact S has exactly n distinct elements and so we will have $r \leq n$.

Continuation

- ▶ Continuing this way, by induction if we get distinct $k_1, k_2, \dots, k_s$ with $k_1 \dashrightarrow k_2 \dashrightarrow \dots \dashrightarrow k_s$, we take $k_{s+1} = \sigma(k_s)$. If $k_{s+1} = k_1$, we can take $r = s$ and we are done.
- ▶ For any $2 \leq t \leq s$, $k_{s+1} = k_t = \sigma(k_{t-1})$ is not possible due to injectivity of σ .
- ▶ If $k_1, \dots, k_{s+1}$ are all distinct, we can continue the induction process.
- ▶ This process has to terminate after some steps as the set S is finite. In fact S has exactly n distinct elements and so we will have $r \leq n$.
- ▶ **Exercise 1.6:** Show that there exists some $t \in \mathbb{N}$ such that $\sigma^t(j) = j$ for all $j \in S$.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.
- ▶ We may write down the permutation by listing the cycles it has.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.
- ▶ We may write down the permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.
- ▶ We may write down the permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.
- ▶ We may also write it as $(3, 7, 1)(4)(2, 5), (6)$ or as $(7, 1, 3)(5, 2)(6)(4)$. In other words, in what order we write these cycles does not matter.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.
- ▶ We may write down the permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.
- ▶ We may also write it as $(3, 7, 1)(4)(2, 5), (6)$ or as $(7, 1, 3)(5, 2)(6)(4)$. In other words, in what order we write these cycles does not matter.
- ▶ Some authors do not write down 1-cycles at all. It is understood that elements of S which are not written down form 1-cycles.

Products of cycles

- ▶ We have seen that a permutation may have several cycles.
- ▶ We may write down the permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.
- ▶ We may also write it as $(3, 7, 1)(4)(2, 5), (6)$ or as $(7, 1, 3)(5, 2)(6)(4)$. In other words, in what order we write these cycles does not matter.
- ▶ Some authors do not write down 1-cycles at all. It is understood that elements of S which are not written down form 1-cycles.
- ▶ With this notation this permutation is simply $(1, 3, 7), (2, 5)$ or $(3, 7, 1)(5, 2)$ etc.

Products of cycles continued

- ▶ Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.

Products of cycles continued

- ▶ Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.
- ▶ Then $(k_1, k_2, \dots, k_r)$ denotes the permutation:

$k_1 \mapsto k_2 \mapsto k_3 \mapsto \dots \mapsto k_r \mapsto k_1, j \mapsto j$ otherwise.

Products of cycles continued

- ▶ Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.
- ▶ Then $(k_1, k_2, \dots, k_r)$ denotes the permutation:

$k_1 \mapsto k_2 \mapsto k_3 \mapsto \dots \mapsto k_r \mapsto k_1, j \mapsto j$ otherwise.

Products of cycles continued

- ▶ Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.
- ▶ Then $(k_1, k_2, \dots, k_r)$ denotes the permutation:

$k_1 \mapsto k_2 \mapsto k_3 \mapsto \dots \mapsto k_r \mapsto k_1, j \mapsto j$ otherwise.

Products of cycles continued

- ▶ Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.
- ▶ Then $(k_1, k_2, \dots, k_r)$ denotes the permutation:

$$k_1 \mapsto k_2 \mapsto k_3 \mapsto \dots \mapsto k_r \mapsto k_1, j \mapsto j \text{ otherwise.}$$

- ▶ More generally if $k_{11}, k_{12}, \dots, k_{1r_1}, k_{21}, k_{22}, \dots, k_{2r_2}, k_{31}, k_{32}, \dots, k_{3r_3}, \dots, k_{m1}, k_{m2}, \dots, k_{mr_m}$ are distinct elements of S , then

$$(k_{11}, k_{12}, \dots, k_{1r_1})(k_{21}, k_{22}, \dots, k_{2r_2}) \cdots (k_{m1}, k_{m2}, \dots, k_{mr_m})$$

is a 'product' of cycles, with

$$\sigma(k_{11}) = k_{12}, \sigma(k_{12}) = k_{13}, \dots, \sigma(k_{1r_1}) = k_{11},$$

$$\sigma(k_{21}) = k_{22}, \dots, \sigma(k_{2r_2}) = k_{21}, \dots,$$

$$\sigma(k_{m1}) = k_{m2} \dots, \sigma(k_{mr_m}) = k_{m1}, \sigma(j) = j, \text{ otherwise.}$$

Products of cycles theorem

- **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.
- ▶ If $S = \{k_1, k_2, \dots, k_r\}$, we are done. If not, take any $j \in \{k_1, k_2, \dots, k_r\}^c$, and we can get a cycle $(j_1, j_2, \dots)$ with elements distinct from $\{k_1, k_2, \dots, k_r\}$. Continuing this way, we can exhaust whole of S , as S is a finite set.

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.
- ▶ If $S = \{k_1, k_2, \dots, k_r\}$, we are done. If not, take any $j \in \{k_1, k_2, \dots, k_r\}^c$, and we can get a cycle $(j_1, j_2, \dots)$ with elements distinct from $\{k_1, k_2, \dots, k_r\}$. Continuing this way, we can exhaust whole of S , as S is a finite set.
- ▶ Clearly then, the permutation σ is a product of cycles. The uniqueness should also be clear as these cycles determine σ .

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.
- ▶ If $S = \{k_1, k_2, \dots, k_r\}$, we are done. If not, take any $j \in \{k_1, k_2, \dots, k_r\}^c$, and we can get a cycle $(j_1, j_2, \dots)$ with elements distinct from $\{k_1, k_2, \dots, k_r\}$. Continuing this way, we can exhaust whole of S , as S is a finite set.
- ▶ Clearly then, the permutation σ is a product of cycles. The uniqueness should also be clear as these cycles determine σ .
- ▶ For i, j in S , write $i \sim j$ if $j = \sigma^r(i)$ for some $r \in \{0, 1, \dots\}$.

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.
- ▶ If $S = \{k_1, k_2, \dots, k_r\}$, we are done. If not, take any $j \in \{k_1, k_2, \dots, k_r\}^c$, and we can get a cycle $(j_1, j_2, \dots)$ with elements distinct from $\{k_1, k_2, \dots, k_r\}$. Continuing this way, we can exhaust whole of S , as S is a finite set.
- ▶ Clearly then, the permutation σ is a product of cycles. The uniqueness should also be clear as these cycles determine σ .
- ▶ For i, j in S , write $i \sim j$ if $j = \sigma^r(i)$ for some $r \in \{0, 1, \dots\}$.
- ▶ Then $\sim$ is an equivalence relation. ([Exercise](#)).

Products of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ **Proof.** Take $k_1 = 1$. Then by Lemma 1.3, and its proof, σ has a cycle $(k_1, k_2, \dots, k_r)$ with $k_1 = 1$.
- ▶ If $S = \{k_1, k_2, \dots, k_r\}$, we are done. If not, take any $j \in \{k_1, k_2, \dots, k_r\}^c$, and we can get a cycle $(j_1, j_2, \dots)$ with elements distinct from $\{k_1, k_2, \dots, k_r\}$. Continuing this way, we can exhaust whole of S , as S is a finite set.
- ▶ Clearly then, the permutation σ is a product of cycles. The uniqueness should also be clear as these cycles determine σ .
- ▶ For i, j in S , write $i \sim j$ if $j = \sigma^r(i)$ for some $r \in \{0, 1, \dots\}$.
- ▶ Then $\sim$ is an equivalence relation. ([Exercise](#)).
- ▶ It may be seen that i, j are in the same cycle if and only if $i \sim j$. In other words, the equivalence classes form different cycles of the permutation.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.3,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.3,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.3,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.3,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

- Therefore the signature of a cycle is defined as
 $(k_1, k_2, \dots, k_r) = (-1)^{n-(1+(n-r))} = (-1)^{r-1}$.

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .

Continuation

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .
- ▶ Permutations with signature $(+1)$ are known as **even** permutations and those with signature (-1) are known as **odd** permutations.

Continuation

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .
- ▶ Permutations with signature $(+1)$ are known as **even** permutations and those with signature (-1) are known as **odd** permutations.
- ▶ **END OF LECTURE 1.**

Lecture 2: Permutations -2

► Recall:

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.

Lecture 2: Permutations -2

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.
- ▶ For distinct $k_1, k_2, \dots, k_r$ in $\{1, 2, \dots, n\}$ (with $r \in \mathbb{N}$) we denote the cycle $k_1 \mapsto k_2 \mapsto \dots \mapsto k_r \mapsto k_1$ simply as $(k_1, k_2, \dots, k_r)$.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ We may write down a permutation by listing the cycles it has.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ We may write down a permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

- Therefore the signature of a cycle is defined as
 $(k_1, k_2, \dots, k_r) = (-1)^{n-(1+(n-r))} = (-1)^{r-1}$.

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .

Continuation

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .
- ▶ Permutations with signature $(+1)$ are known as **even** permutations and those with signature (-1) are known as **odd** permutations.

Permutations as products of transpositions

- ▶ Let $\sigma = (k_1, k_2, \dots, k_r)$ be a cycle on $S = \{1, 2, \dots, n\}$. This means that $k_1, k_2, \dots, k_r$ are distinct elements in S and we are looking at the permutation:

$$k_1 \mapsto k_2 \mapsto \dots \mapsto k_{r-1} \mapsto k_r \mapsto k_1.$$

Permutations as products of transpositions

- ▶ Let $\sigma = (k_1, k_2, \dots, k_r)$ be a cycle on $S = \{1, 2, \dots, n\}$. This means that $k_1, k_2, \dots, k_r$ are distinct elements in S and we are looking at the permutation:

$$k_1 \mapsto k_2 \mapsto \dots \mapsto k_{r-1} \mapsto k_r \mapsto k_1.$$

- ▶ To arrive at this permutation, we may first interchange k_1, k_2 and then interchange k_1, k_3 , and then k_1, k_4 and so on and finally k_1 and k_r .

Permutations as products of transpositions

- ▶ Let $\sigma = (k_1, k_2, \dots, k_r)$ be a cycle on $S = \{1, 2, \dots, n\}$. This means that $k_1, k_2, \dots, k_r$ are distinct elements in S and we are looking at the permutation:

$$k_1 \mapsto k_2 \mapsto \dots \mapsto k_{r-1} \mapsto k_r \mapsto k_1.$$

- ▶ To arrive at this permutation, we may first interchange k_1, k_2 and then interchange k_1, k_3 , and then k_1, k_4 and so on and finally k_1 and k_r .
- ▶ In other words, if $\tau_{i,j}$ is the transposition between i and j , then

$$\sigma = \tau_{k_1, k_r} \circ \dots \circ \tau_{k_1, k_3} \circ \tau_{k_1, k_2}.$$

Permutations as products of transpositions

- ▶ Let $\sigma = (k_1, k_2, \dots, k_r)$ be a cycle on $S = \{1, 2, \dots, n\}$. This means that $k_1, k_2, \dots, k_r$ are distinct elements in S and we are looking at the permutation:

$$k_1 \mapsto k_2 \mapsto \dots \mapsto k_{r-1} \mapsto k_r \mapsto k_1.$$

- ▶ To arrive at this permutation, we may first interchange k_1, k_2 and then interchange k_1, k_3 , and then k_1, k_4 and so on and finally k_1 and k_r .
- ▶ In other words, if $\tau_{i,j}$ is the transposition between i and j , then

$$\sigma = \tau_{k_1, k_r} \circ \dots \circ \tau_{k_1, k_3} \circ \tau_{k_1, k_2}.$$

- ▶ Since every permutation is a product of disjoint cycles it follows that every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

Permutations as products of transpositions

- ▶ Let $\sigma = (k_1, k_2, \dots, k_r)$ be a cycle on $S = \{1, 2, \dots, n\}$. This means that $k_1, k_2, \dots, k_r$ are distinct elements in S and we are looking at the permutation:

$$k_1 \mapsto k_2 \mapsto \dots \mapsto k_{r-1} \mapsto k_r \mapsto k_1.$$

- ▶ To arrive at this permutation, we may first interchange k_1, k_2 and then interchange k_1, k_3 , and then k_1, k_4 and so on and finally k_1 and k_r .
- ▶ In other words, if $\tau_{i,j}$ is the transposition between i and j , then

$$\sigma = \tau_{k_1, k_r} \circ \dots \circ \tau_{k_1, k_3} \circ \tau_{k_1, k_2}.$$

- ▶ Since every permutation is a product of disjoint cycles it follows that every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

- ▶ Note that this factorization is not unique.

Homomorphism property of the signature

- **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

Homomorphism property of the signature

- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Proof.** We first prove the theorem when τ is a transposition. So let τ be transposition of two distinct indices in S .

Homomorphism property of the signature

- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Proof.** We first prove the theorem when τ is a transposition. So let τ be transposition of two distinct indices in S .
- ▶ Let

$$\sigma = (k_{11}, k_{12}, \dots, k_{1r_1})(k_{21}, k_{22}, \dots, k_{2r_2}) \cdots (k_{p1}, k_{p2}, \dots, k_{pr_p})$$

be the cycle decomposition of σ .

Homomorphism property of the signature

- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Proof.** We first prove the theorem when τ is a transposition. So let τ be transposition of two distinct indices in S .
- ▶ Let

$$\sigma = (k_{11}, k_{12}, \dots, k_{1r_1})(k_{21}, k_{22}, \dots, k_{2r_2}) \cdots (k_{p1}, k_{p2}, \dots, k_{pr_p})$$

be the cycle decomposition of σ .

- ▶ By the definition of ϵ , $\epsilon(\sigma) = (-1)^{n-p}$ and we have also seen that $\epsilon(\tau) = -1$.

Homomorphism property of the signature

- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Proof.** We first prove the theorem when τ is a transposition. So let τ be transposition of two distinct indices in S .
- ▶ Let

$$\sigma = (k_{11}, k_{12}, \dots, k_{1r_1})(k_{21}, k_{22}, \dots, k_{2r_2}) \cdots (k_{p1}, k_{p2}, \dots, k_{pr_p})$$

be the cycle decomposition of σ .

- ▶ By the definition of ϵ , $\epsilon(\sigma) = (-1)^{n-p}$ and we have also seen that $\epsilon(\tau) = -1$.
- ▶ Therefore our aim is to show that $\epsilon(\tau \circ \sigma) = (-1)^{n-p+1}$.

Continuation

- ▶ Now there are two possibilities, the transposition τ could be of two indices of same cycle of σ or could be of different cycles of σ .

Continuation

- ▶ Now there are two possibilities, the transposition τ could be of two indices of same cycle of σ or could be of different cycles of σ .
- ▶ Case (i): Suppose τ is transposition of k_{1i} and k_{1j} .

Continuation

- ▶ Now there are two possibilities, the transposition τ could be of two indices of same cycle of σ or could be of different cycles of σ .
- ▶ Case (i): Suppose τ is transposition of k_{1i} and k_{1j} .
- ▶ With out loss of generality we may take $i < j$.

Continuation

- ▶ Now there are two possibilities, the transposition τ could be of two indices of same cycle of σ or could be of different cycles of σ .
- ▶ Case (i): Suppose τ is transposition of k_{1i} and k_{1j} .
- ▶ With out loss of generality we may take $i < j$.
- ▶ Then $\tau \circ \sigma$ acting on $\{k_{11}, k_{12}, \dots, k_{1r_1}\}$ has two cycles, namely

$$(k_{11}, \dots, k_{1(i-1)}, k_{1j}, k_{1(j+1)}, \dots, k_{1r_1})$$

and

$$(k_{1i}, k_{1(i+1)}, \dots, k_{1(j-1)}).$$

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .
- ▶ Then $\tau \circ \sigma$ acting on these two cycles gives a single cycle:

$$(k_{11}, \dots, k_{1(i-1)}, k_{2j}, k_{2(j+1)}, k_{2r_2}, k_{21},$$

$$\dots, k_{2(j-1)}, k_{1i}, k_{1(i+1)}, \dots, k_{1r_1}).$$

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .
- ▶ Then $\tau \circ \sigma$ acting on these two cycles gives a single cycle:

$$(k_{11}, \dots, k_{1(i-1)}, k_{2j}, k_{2(j+1)}, k_{2r_2}, k_{21},$$

$$\dots, k_{2(j-1)}, k_{1i}, k_{1(i+1)}, \dots, k_{1r_1}).$$

- ▶ In other words in both cases the number of cycles changes by 1 (either +1 or -1).

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .
- ▶ Then $\tau \circ \sigma$ acting on these two cycles gives a single cycle:

$$(k_{11}, \dots, k_{1(i-1)}, k_{2j}, k_{2(j+1)}, k_{2r_2}, k_{21}, \\ \dots, k_{2(j-1)}, k_{1i}, k_{1(i+1)}, \dots, k_{1r_1}).$$

- ▶ In other words in both cases the number of cycles changes by 1 (either +1 or -1).
- ▶ **Caution:** We have written the proof above essentially assuming that $r_1, r_2 \geq 2$, as we have indices such as $(i-1), (j-1)$ etc. You should verify that the result is true for all r_1, r_2 .

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .
- ▶ Then $\tau \circ \sigma$ acting on these two cycles gives a single cycle:

$$(k_{11}, \dots, k_{1(i-1)}, k_{2j}, k_{2(j+1)}, k_{2r_2}, k_{21}, \\ \dots, k_{2(j-1)}, k_{1i}, k_{1(i+1)}, \dots, k_{1r_1}).$$

- ▶ In other words in both cases the number of cycles changes by 1 (either +1 or -1).
- ▶ **Caution:** We have written the proof above essentially assuming that $r_1, r_2 \geq 2$, as we have indices such as $(i-1), (j-1)$ etc. You should verify that the result is true for all r_1, r_2 .
- ▶ This proves that $\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma)$ whenever τ is a transposition.

Continuation

- ▶ Case (ii): Suppose τ is transposition of k_{1i} and k_{2j} .
- ▶ Then $\tau \circ \sigma$ acting on these two cycles gives a single cycle:

$$(k_{11}, \dots, k_{1(i-1)}, k_{2j}, k_{2(j+1)}, k_{2r_2}, k_{21}, \\ \dots, k_{2(j-1)}, k_{1i}, k_{1(i+1)}, \dots, k_{1r_1}).$$

- ▶ In other words in both cases the number of cycles changes by 1 (either +1 or -1).
- ▶ **Caution:** We have written the proof above essentially assuming that $r_1, r_2 \geq 2$, as we have indices such as $(i-1), (j-1)$ etc. You should verify that the result is true for all r_1, r_2 .
- ▶ This proves that $\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma)$ whenever τ is a transposition.
- ▶ Since every permutation is a product of transpositions, by mathematical induction we get $\epsilon(\tau \circ \sigma)$ for every τ, σ .

Consequences

- **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

Consequences

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

- ▶ **Proof.** This is clear from the previous theorem and mathematical induction.

Consequences

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

- ▶ **Proof.** This is clear from the previous theorem and mathematical induction.
- ▶ **Corollary 2.3:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_l$, where $\tau_1, \tau_2, \dots, \tau_k, \sigma_1, \sigma_2, \dots, \sigma_l$ are transpositions, then $k - l$ is even. In particular, k is odd/even if and only if l is odd/even.

Consequences

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

- ▶ **Proof.** This is clear from the previous theorem and mathematical induction.
- ▶ **Corollary 2.3:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_l$, where $\tau_1, \tau_2, \dots, \tau_k, \sigma_1, \sigma_2, \dots, \sigma_l$ are transpositions, then $k - l$ is even. In particular, k is odd/even if and only if l is odd/even.
- ▶ Recall that a permutation σ is called even if $\epsilon(\sigma) = 1$ and is called odd if $\epsilon(\sigma) = -1$. These results tell us that composition of two permutations is even if and only if either both of them are even or both of them are odd.

Consequences

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

- ▶ **Proof.** This is clear from the previous theorem and mathematical induction.
- ▶ **Corollary 2.3:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_l$, where $\tau_1, \tau_2, \dots, \tau_k, \sigma_1, \sigma_2, \dots, \sigma_l$ are transpositions, then $k - l$ is even. In particular, k is odd/even if and only if l is odd/even.
- ▶ Recall that a permutation σ is called even if $\epsilon(\sigma) = 1$ and is called odd if $\epsilon(\sigma) = -1$. These results tell us that composition of two permutations is even if and only if either both of them are even or both of them are odd.
- ▶ In the Fifteen Puzzle game, one can see that moves of the game do not change the signature (It is an invariant.). This proves why it is not possible to reach the natural permutation on starting from (15 14)

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,



$$(P^\sigma x)_i = \sum_j p_{ij}^\sigma x_j = x_{\sigma^{-1}(i)}.$$

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,



$$(P^\sigma x)_i = \sum_j p_{ij}^\sigma x_j = x_{\sigma^{-1}(i)}.$$

- Note that $P^\sigma e_j = e_{\sigma(j)}$. Therefore P^σ just permutes the basis elements $e_1, e_2, \dots, e_n$, sending e_j to $e_{\sigma(j)}$. Hence for any two permutations σ, τ , $P^{\tau \circ \sigma} = P^\tau \cdot P^\sigma$.

Stochastic and doubly stochastic matrices

- **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be stochastic if

Stochastic and doubly stochastic matrices

- ▶ **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be **stochastic** if
- ▶ (i) $d_{ij} \geq 0$ for all $1 \leq i, j \leq n$ (entries are non-negative);

Stochastic and doubly stochastic matrices

- ▶ **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be **stochastic** if
 - ▶ (i) $d_{ij} \geq 0$ for all $1 \leq i, j \leq n$ (entries are non-negative);
 - ▶ (ii) $\sum_{j=1}^n d_{ij} = 1$ for every i (row sums of the matrix are equal to 1.)

Stochastic and doubly stochastic matrices

- ▶ **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be **stochastic** if
 - ▶ (i) $d_{ij} \geq 0$ for all $1 \leq i, j \leq n$ (entries are non-negative);
 - ▶ (ii) $\sum_{j=1}^n d_{ij} = 1$ for every i (row sums of the matrix are equal to 1.)
- ▶ A stochastic matrix is said to be **doubly stochastic** if it also satisfies

Stochastic and doubly stochastic matrices

- ▶ **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be **stochastic** if
 - ▶ (i) $d_{ij} \geq 0$ for all $1 \leq i, j \leq n$ (entries are non-negative);
 - ▶ (ii) $\sum_{j=1}^n d_{ij} = 1$ for every i (row sums of the matrix are equal to 1.)
- ▶ A stochastic matrix is said to be **doubly stochastic** if it also satisfies
 - ▶ (iii) $\sum_{i=1}^n d_{ij} = 1$ (column sums are also equal to 1.)

Stochastic and doubly stochastic matrices

- ▶ **Definition 2.5:** A matrix $D = [d_{ij}]_{1 \leq i, j \leq n}$ is said to be **stochastic** if
 - ▶ (i) $d_{ij} \geq 0$ for all $1 \leq i, j \leq n$ (entries are non-negative);
 - ▶ (ii) $\sum_{j=1}^n d_{ij} = 1$ for every i (row sums of the matrix are equal to 1.)
- ▶ A stochastic matrix is said to be **doubly stochastic** if it also satisfies
 - ▶ (iii) $\sum_{i=1}^n d_{ij} = 1$ (column sums are also equal to 1.)
- ▶ Clearly all permutation matrices are doubly stochastic matrices.

Extreme points of convex sets

- ▶ Let $\mathcal{D}$ be the set of all $n \times n$ doubly stochastic matrices.

Extreme points of convex sets

- ▶ Let $\mathcal{D}$ be the set of all $n \times n$ doubly stochastic matrices.
- ▶ Then $\mathcal{D}$ is a **convex** set, that is for $D, E \in \mathcal{D}$,

$$pD + (1 - p)E \in \mathcal{D}, \quad \forall 0 \leq p \leq 1,$$

Extreme points of convex sets

- ▶ Let $\mathcal{D}$ be the set of all $n \times n$ doubly stochastic matrices.
- ▶ Then $\mathcal{D}$ is a **convex** set, that is for $D, E \in \mathcal{D}$,

$$pD + (1 - p)E \in \mathcal{D}, \quad \forall 0 \leq p \leq 1,$$

- ▶ that is, the line segment joining D, E is contained in $\mathcal{D}$.

Extreme points of convex sets

- ▶ Let $\mathcal{D}$ be the set of all $n \times n$ doubly stochastic matrices.
- ▶ Then $\mathcal{D}$ is a **convex** set, that is for $D, E \in \mathcal{D}$,

$$pD + (1 - p)E \in \mathcal{D}, \quad \forall 0 \leq p \leq 1,$$

- ▶ that is, the line segment joining D, E is contained in $\mathcal{D}$.
- ▶ A matrix $F \in \mathcal{D}$ is said to be an **extreme point** if

$$F = pD + (1 - p)E$$

with $0 < p < 1$, implies $D = E = F$.

Extreme points of convex sets

- ▶ Let $\mathcal{D}$ be the set of all $n \times n$ doubly stochastic matrices.
- ▶ Then $\mathcal{D}$ is a **convex** set, that is for $D, E \in \mathcal{D}$,

$$pD + (1 - p)E \in \mathcal{D}, \quad \forall 0 \leq p \leq 1,$$

- ▶ that is, the line segment joining D, E is contained in $\mathcal{D}$.
- ▶ A matrix $F \in \mathcal{D}$ is said to be an **extreme point** if

$$F = pD + (1 - p)E$$

with $0 < p < 1$, implies $D = E = F$.

- ▶ **Simple Exercise:** Permutation matrices are extreme points of $\mathcal{D}$.

Birkhoff-von Neumann theorem

- ▶ **Theorem 2.6:** A doubly stochastic matrix is an extreme point of the convex set of doubly stochastic matrices if and only if it is a permutation matrix.

Birkhoff-von Neumann theorem

- ▶ **Theorem 2.6:** A doubly stochastic matrix is an extreme point of the convex set of doubly stochastic matrices if and only if it is a permutation matrix.
- ▶ **Proof.** Difficult exercise. (Omitted).

Birkhoff-von Neumann theorem

- ▶ **Theorem 2.6:** A doubly stochastic matrix is an extreme point of the convex set of doubly stochastic matrices if and only if it is a permutation matrix.
- ▶ **Proof.** Difficult exercise. (Omitted).
- ▶ **END OF LECTURE 2.**

Lecture 3: Leibniz formula for determinants

► Recall:

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.

Lecture 3: Leibniz formula for determinants

- ▶ Recall:
- ▶ **Definition 1.1:** Let S be a finite set. Then a bijective function $\sigma : S \rightarrow S$ is said to be a **permutation** of S .
- ▶ **Example 1.4:** Suppose $S = \{1, 2, \dots, 7\}$. Consider the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 4 & 2 & 6 & 1 \end{pmatrix}$$

- ▶ We see $1 \mapsto 3 \mapsto 7 \mapsto 1$. This we call as a **cycle**. It is a cycle of **length** 3.
- ▶ This permutation also has $2 \mapsto 5 \mapsto 2$, a cycle of length 2.
- ▶ It also has $4 \mapsto 4$ and $6 \mapsto 6$, cycles of length 1.
- ▶ For distinct $k_1, k_2, \dots, k_r$ in $\{1, 2, \dots, n\}$ (with $r \in \mathbb{N}$) we denote the cycle $k_1 \mapsto k_2 \mapsto \dots \mapsto k_r \mapsto k_1$ simply as $(k_1, k_2, \dots, k_r)$.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ We may write down a permutation by listing the cycles it has.

Product of cycles theorem

- ▶ **Theorem 1.7:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ is a permutation of S . Then S decomposes uniquely as a product of cycles.
- ▶ We may write down a permutation by listing the cycles it has.
- ▶ For instance, the permutation of Example 1.4, is written as $(1, 3, 7)(2, 5)(4)(6)$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

Signature of a permutation

- **Definition 1.8:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and let σ be a permutation of S . Then the **signature** of σ is defined as the number

$$\epsilon(\sigma) = (-1)^{n-p}$$

where p is the number of cycles (including cycles of length 1) in the cycle decomposition of σ .

- For instance for the permutation σ of Example 1.4,
 $\epsilon(\sigma) = (-1)^{7-4} = (-1)^3 = -1$.
- Note that the signature of identity permutation is always 1.
- A cycle $(k_1, k_2, \dots, k_r)$ can be identified with the permutation σ defined by

$$\sigma(k_1) = k_2, \sigma(k_2) = k_3, \dots, \sigma(k_r) = k_1$$

and $\sigma(j) = j$ for $j \notin \{k_1, k_2, \dots, k_r\}$.

- Therefore the signature of a cycle is defined as
 $(k_1, k_2, \dots, k_r) = (-1)^{n-(1+(n-r))} = (-1)^{r-1}$.

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .

Continuation

- ▶ Cycles of length two are known as transpositions. We see that transpositions have signature (-1) .
- ▶ Permutations with signature $(+1)$ are known as **even** permutations and those with signature (-1) are known as **odd** permutations.

Permutations as products of transpositions

- ▶ Every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

Permutations as products of transpositions

- ▶ Every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

- ▶ Note that this factorization is not unique.

Permutations as products of transpositions

- ▶ Every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

- ▶ Note that this factorization is not unique.
- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

Permutations as products of transpositions

- ▶ Every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

- ▶ Note that this factorization is not unique.
- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \dots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

Permutations as products of transpositions

- ▶ Every permutation is a product of transpositions. In other words, given any permutation σ there exist transpositions $\tau_1, \tau_2, \dots, \tau_k$ (for some $k \in \{0, 1, \dots\}$) such that

$$\sigma = \tau_k \circ \tau_{k-1} \circ \dots \circ \tau_2 \circ \tau_1.$$

- ▶ Note that this factorization is not unique.
- ▶ **Theorem 2.1:** Let $S = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose σ, τ are two permutations of S . Then

$$\epsilon(\tau \circ \sigma) = \epsilon(\tau) \cdot \epsilon(\sigma).$$

- ▶ **Corollary 2.2:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \dots \circ \tau_k$, where $\tau_1, \dots, \tau_k$ are transpositions then

$$\epsilon(\tau) = (-1)^k.$$

- ▶ **Corollary 2.3:** If a permutation $\tau = \tau_1 \circ \tau_2 \circ \dots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \dots \circ \sigma_l$, where $\tau_1, \tau_2, \dots, \tau_k, \sigma_1, \sigma_2, \dots, \sigma_l$ are transpositions, then $k - l$ is even. In particular, k is odd/even if and only if l is odd/even.

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,



$$(P^\sigma x)_i = \sum_j p_{ij}^\sigma x_j = x_{\sigma^{-1}(i)}.$$

Permutation matrices

- **Definition 2.4:** Fix $n \in \mathbb{N}$ and let σ be a permutation of $\{1, 2, \dots, n\}$. Then the $n \times n$ matrix P^σ defined by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

is called the **permutation** matrix associated with the permutation σ . Note that every row or column of P^σ has exactly one non-zero entry which is 1.

- We also consider the matrix P^σ as the linear transformation $x \mapsto P^\sigma x$ on $\mathbb{R}^n$. More explicitly, if $x \in \mathbb{R}^n$ has the expansion $x = \sum_{j=1}^n x_j e_j$ in the standard basis $\{e_1, e_2, \dots, e_n\}$,



$$(P^\sigma x)_i = \sum_j p_{ij}^\sigma x_j = x_{\sigma^{-1}(i)}.$$

- Note that $P^\sigma e_j = e_{\sigma(j)}$. Therefore P^σ just permutes the basis elements $e_1, e_2, \dots, e_n$, sending e_j to $e_{\sigma(j)}$. Hence for any two permutations σ, τ , $P^{\tau \circ \sigma} = P^\tau \cdot P^\sigma$.

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .
- ▶ Now determinants of square matrices were defined inductively as follows:

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .
- ▶ Now determinants of square matrices were defined inductively as follows:
- ▶ If $n = 1$ and $A = [a_{11}]$, then $\det(A) = a_{11}$.

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .
- ▶ Now determinants of square matrices were defined inductively as follows:
- ▶ If $n = 1$ and $A = [a_{11}]$, then $\det(A) = a_{11}$.
- ▶ For $n \geq 2$,

$$\det(A) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) + \dots + (-1)^{n-1} a_{n1} \det(A_{n1}).$$

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .
- ▶ Now determinants of square matrices were defined inductively as follows:
- ▶ If $n = 1$ and $A = [a_{11}]$, then $\det(A) = a_{11}$.
- ▶ For $n \geq 2$,

$$\det(A) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) + \dots + (-1)^{n-1} a_{n1} \det(A_{n1}).$$

- ▶ This is known as Laplace formula/expansion for the determinant.

Determinants

- ▶ Notation: Let A be an $n \times n$ matrix. Then for any $1 \leq i, j \leq n$, the matrix formed by dropping i -th row and j -th column is known as (i, j) -th minor of A and is denoted by A_{ij} .
- ▶ Now determinants of square matrices were defined inductively as follows:
- ▶ If $n = 1$ and $A = [a_{11}]$, then $\det(A) = a_{11}$.
- ▶ For $n \geq 2$,

$$\det(A) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) + \dots + (-1)^{n-1} a_{n1} \det(A_{n1}).$$

- ▶ This is known as Laplace formula/expansion for the determinant.
- ▶ Here we have written the expansion using the first column. But we could have used any row or column.

Permutations and the determinant

- ▶ Let us look at the determinant for 2×2 and 3×3 matrices.

Permutations and the determinant

- ▶ Let us look at the determinant for 2×2 and 3×3 matrices.
- ▶ We have,

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$

Permutations and the determinant

- ▶ Let us look at the determinant for 2×2 and 3×3 matrices.
- ▶ We have,

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$



$$\begin{aligned} & \det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\ &\quad - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Permutations and the determinant

- ▶ Let us look at the determinant for 2×2 and 3×3 matrices.
- ▶ We have,

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$



$$\begin{aligned} & \det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\ &\quad - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

- ▶ You may note that there are $3! = 6$ terms here and each term is of the form $(\pm)a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}$ for some permutation σ of $\{1, 2, 3\}$.

Permutations and the determinant

- ▶ Let us look at the determinant for 2×2 and 3×3 matrices.
- ▶ We have,

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$



$$\begin{aligned} & \det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\ &\quad - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

- ▶ You may note that there are $3! = 6$ terms here and each term is of the form $(\pm)a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}$ for some permutation σ of $\{1, 2, 3\}$.
- ▶ This suggests the following theorem.

Leibniz formula for determinants

- ▶ In the following S_n denotes the set of all permutations of the set $S = \{1, 2, \dots, n\}$.

Leibniz formula for determinants

- ▶ In the following S_n denotes the set of all permutations of the set $S = \{1, 2, \dots, n\}$.
- ▶ **Theorem 3.1 (Leibniz formula):** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be an $n \times n$ matrix. Then

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Leibniz formula for determinants

- ▶ In the following S_n denotes the set of all permutations of the set $S = \{1, 2, \dots, n\}$.
- ▶ **Theorem 3.1 (Leibniz formula):** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be an $n \times n$ matrix. Then

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ To prove this theorem we use the following characterization of the determinant proved in Semester -I.

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:
- ▶ (i) $f(I) = 1$;

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:
- ▶ (i) $f(I) = 1$;
- ▶ (ii) f is linear in each row (keeping other rows fixed).

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:
- ▶ (i) $f(I) = 1$;
- ▶ (ii) f is linear in each row (keeping other rows fixed).
- ▶ (iii) If two adjacent rows of A are equal then $f(A) = 0$.

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:
 - ▶ (i) $f(I) = 1$;
 - ▶ (ii) f is linear in each row (keeping other rows fixed).
 - ▶ (iii) If two adjacent rows of A are equal then $f(A) = 0$.
- ▶ Then $f(A) = \det(A)$ for every $A \in M_n(\mathbb{R})$.

A characterization of the determinant

- ▶ Let $M_n(\mathbb{R})$ denote the vector space of $n \times n$ real matrices.
- ▶ **Theorem 1.?:** Let $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ be a function satisfying:
 - ▶ (i) $f(I) = 1$;
 - ▶ (ii) f is linear in each row (keeping other rows fixed).
 - ▶ (iii) If two adjacent rows of A are equal then $f(A) = 0$.
- ▶ Then $f(A) = \det(A)$ for every $A \in M_n(\mathbb{R})$.
- ▶ The determinant satisfies (i) to (iii) and the word 'adjacent' in (iii) can be dropped. The property (ii) is known as 'multi-linearity'.

Proof of Leibniz formula

► Define $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ by

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Proof of Liebniz formula

- ▶ Define $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ by

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ We wish to show that f satisfies conditions (i) to (iii) mentioned above.

Proof of Leibniz formula

- ▶ Define $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ by

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ We wish to show that f satisfies conditions (i) to (iii) mentioned above.
- ▶ If $A = I$, then, $a_{ij} = 0$ if $i \neq j$, hence

$$a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$$

unless σ is the identity permutation.

Proof of Leibniz formula

- ▶ Define $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ by

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ We wish to show that f satisfies conditions (i) to (iii) mentioned above.
- ▶ If $A = I$, then, $a_{ij} = 0$ if $i \neq j$, hence

$$a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$$

unless σ is the identity permutation.

- ▶ If σ is the identity permutation $\epsilon(\sigma) = 1$. Hence $f(I) = 1 \cdot 1 \cdots 1 = 1$.

Multi-linearity

- ▶ Consider i -th row of A . Suppose $a_{ij} = sb_{ij} + tc_{ij}$ for some real numbers s, t , b_{ij}, c_{ij} , $1 \leq j \leq n$.

Multi-linearity

- ▶ Consider i -th row of A . Suppose $a_{ij} = sb_{ij} + tc_{ij}$ for some real numbers s, t , b_{ij}, c_{ij} , $1 \leq j \leq n$.
- ▶ Now

$$\begin{aligned} f(A) &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots (sb_{i\sigma(i)} + tc_{i\sigma(i)}) \cdots a_{n\sigma(n)} \end{aligned}$$

Multi-linearity

- ▶ Consider i -th row of A . Suppose $a_{ij} = sb_{ij} + tc_{ij}$ for some real numbers s, t , b_{ij}, c_{ij} , $1 \leq j \leq n$.
- ▶ Now

$$\begin{aligned} f(A) &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots (sb_{i\sigma(i)} + tc_{i\sigma(i)}) \cdots a_{n\sigma(n)} \end{aligned}$$

- ▶ Then it is clear that f satisfies (ii).

Continuation

- ▶ Now suppose i -th row and j -th row of A are equal for some $1 \leq i < j \leq n$, that is $a_{ik} = a_{jk}$ for all $1 \leq k \leq n$.

Continuation

- ▶ Now suppose i -th row and j -th row of A are equal for some $1 \leq i < j \leq n$, that is $a_{ik} = a_{jk}$ for all $1 \leq k \leq n$.
- ▶ Let τ be the transposition of i and j .

Continuation

- ▶ Now suppose i -th row and j -th row of A are equal for some $1 \leq i < j \leq n$, that is $a_{ik} = a_{jk}$ for all $1 \leq k \leq n$.
- ▶ Let τ be the transposition of i and j .
- ▶ To begin with we observe that

$$\{\sigma \circ \tau : \sigma \in S_n\} = S_n.$$

Continuation

- ▶ Now suppose i -th row and j -th row of A are equal for some $1 \leq i < j \leq n$, that is $a_{ik} = a_{jk}$ for all $1 \leq k \leq n$.
- ▶ Let τ be the transposition of i and j .
- ▶ To begin with we observe that

$$\{\sigma \circ \tau : \sigma \in S_n\} = S_n.$$

- ▶ Since for any σ in S_n , $\sigma \circ \tau \in S_n$,

$$\{\sigma \circ \tau : \sigma \in S_n\} \subseteq S_n$$

is obvious. Now consider any permutation η in S_n . We can write η as $\sigma \circ \tau$, where $\sigma = \eta \circ (\tau)^{-1}$. This shows,

$$S_n \subseteq \{\sigma \circ \tau : \sigma \in S_n\}.$$

Continuation

► Therefore,

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma \circ \tau) a_{1\sigma \circ \tau(1)} a_{2\sigma \circ \tau(2)} \cdots a_{i\sigma \circ \tau(i)} \cdots a_{j\sigma \circ \tau(j)} \cdots a_{n\sigma \circ \tau(n)}.$$

Continuation

- Therefore,

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma \circ \tau) a_{1\sigma \circ \tau(1)} a_{2\sigma \circ \tau(2)} \cdots a_{i\sigma \circ \tau(i)} \cdots a_{j\sigma \circ \tau(j)} \cdots a_{n\sigma \circ \tau(n)}.$$

- We have $\epsilon(\sigma \circ \tau) = \epsilon(\sigma) \cdot \epsilon(\tau) = -\epsilon(\sigma)$.

Continuation

- Therefore,

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma \circ \tau) a_{1\sigma \circ \tau(1)} a_{2\sigma \circ \tau(2)} \cdots a_{i\sigma \circ \tau(i)} \cdots a_{j\sigma \circ \tau(j)} \cdots a_{n\sigma \circ \tau(n)}.$$

- We have $\epsilon(\sigma \circ \tau) = \epsilon(\sigma) \cdot \epsilon(\tau) = -\epsilon(\sigma)$.
- Also, $\tau(i) = j, \tau(j) = i$ and $\tau(k) = k$ for $k \neq i, j$. Moreover, since i -th row and j -th row of A are equal,

$$\begin{aligned} f(A) &= - \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{i\sigma(j)} \cdots a_{j\sigma(i)} \cdots a_{n\sigma(n)} \\ &= - \sum_{n \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{j\sigma(j)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} \\ &= -f(A). \end{aligned}$$

Continuation

- Therefore,

$$f(A) = \sum_{\sigma \in S_n} \epsilon(\sigma \circ \tau) a_{1\sigma \circ \tau(1)} a_{2\sigma \circ \tau(2)} \cdots a_{i\sigma \circ \tau(i)} \cdots a_{j\sigma \circ \tau(j)} \cdots a_{n\sigma \circ \tau(n)}.$$

- We have $\epsilon(\sigma \circ \tau) = \epsilon(\sigma) \cdot \epsilon(\tau) = -\epsilon(\sigma)$.
- Also, $\tau(i) = j, \tau(j) = i$ and $\tau(k) = k$ for $k \neq i, j$. Moreover, since i -th row and j -th row of A are equal,

$$\begin{aligned} f(A) &= - \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{i\sigma(j)} \cdots a_{j\sigma(i)} \cdots a_{n\sigma(n)} \\ &= - \sum_{n \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{j\sigma(j)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} \\ &= -f(A). \end{aligned}$$

- Therefore, $2f(A) = 0$ or $f(A) = 0$. This proves (iii) and hence $f(A) = \det(A)$. ■.

Determinant of permutation matrices

- ▶ Recall that for a permutation $\sigma \in S_n$, we have defined the associated 'permutation matrix' P^σ by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

Determinant of permutation matrices

- ▶ Recall that for a permutation $\sigma \in S_n$, we have defined the associated 'permutation matrix' P^σ by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ **Theorem 3.2:** For $\sigma \in S_n$,

$$\det(P^\sigma) = \epsilon(\sigma).$$

Determinant of permutation matrices

- ▶ Recall that for a permutation $\sigma \in S_n$, we have defined the associated 'permutation matrix' P^σ by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ **Theorem 3.2:** For $\sigma \in S_n$,

$$\det(P^\sigma) = \epsilon(\sigma).$$

- ▶ **Proof:** We have

$$\det(P^\sigma) = \sum_{\eta \in S^n} \epsilon(\eta) P_{1\eta(1)}^\sigma P_{2\eta(2)}^\sigma \cdots P_{n\eta(n)}^\sigma.$$

Determinant of permutation matrices

- ▶ Recall that for a permutation $\sigma \in S_n$, we have defined the associated 'permutation matrix' P^σ by

$$p_{ij}^\sigma = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ **Theorem 3.2:** For $\sigma \in S_n$,

$$\det(P^\sigma) = \epsilon(\sigma).$$

- ▶ **Proof:** We have

$$\det(P^\sigma) = \sum_{\eta \in S^n} \epsilon(\eta) P_{1\eta(1)}^\sigma P_{2\eta(2)}^\sigma \cdots P_{n\eta(n)}^\sigma.$$

- ▶ We see that for a term in this sum to be non-trivial we need $j = \sigma(\eta(j))$ for every j , or $\eta = (\sigma)^{-1}$.

Continuation

► Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

Continuation

- ▶ Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

- ▶ Recall, $\epsilon(\sigma) \cdot \epsilon(\sigma^{-1}) = \epsilon(\iota) = 1$.

Continuation

- ▶ Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

- ▶ Recall, $\epsilon(\sigma) \cdot \epsilon(\sigma^{-1}) = \epsilon(\iota) = 1$.
- ▶ Hence $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ for every σ . ■

Continuation

- ▶ Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

- ▶ Recall, $\epsilon(\sigma) \cdot \epsilon(\sigma^{-1}) = \epsilon(\iota) = 1$.
- ▶ Hence $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ for every σ . ■
- ▶ **Definition 3.3:** For a square matrix $A = [a_{ij}]_{1 \leq i, j \leq n}$, the **permanent** of A is defined as:

$$\text{per}(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Continuation

- ▶ Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

- ▶ Recall, $\epsilon(\sigma) \cdot \epsilon(\sigma^{-1}) = \epsilon(\iota) = 1$.
- ▶ Hence $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ for every σ . ■
- ▶ **Definition 3.3:** For a square matrix $A = [a_{ij}]_{1 \leq i, j \leq n}$, the **permanent** of A is defined as:

$$\text{per}(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ The permanent of A does appear in some areas of mathematics. However, it is not as useful as the determinant and in general it is more difficult to compute.

Continuation

- ▶ Therefore

$$\det(P^\sigma) = \epsilon(\sigma^{-1})$$

- ▶ Recall, $\epsilon(\sigma) \cdot \epsilon(\sigma^{-1}) = \epsilon(\iota) = 1$.
- ▶ Hence $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ for every σ . ■
- ▶ **Definition 3.3:** For a square matrix $A = [a_{ij}]_{1 \leq i, j \leq n}$, the **permanent** of A is defined as:

$$\text{per}(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ The permanent of A does appear in some areas of mathematics. However, it is not as useful as the determinant and in general it is more difficult to compute.
- ▶ **An interesting problem:** Show that for any $n \times n$ doubly stochastic matrix D ,

$$\text{per}(D) \geq \frac{n!}{n^n}.$$

In other words, the permanent on doubly stochastic matrices

Lecture 4: Determinants of partitioned matrices

- ▶ Reference for 'Doubly stochastic matrices':

Lecture 4: Determinants of partitioned matrices

- ▶ Reference for 'Doubly stochastic matrices':
- ▶ (i) Non-negative matrices and applications, R B Bapat and T E S Raghavan

Lecture 4: Determinants of partitioned matrices

- ▶ Reference for 'Doubly stochastic matrices':
- ▶ (i) Non-negative matrices and applications, R B Bapat and T E S Raghavan
- ▶ (ii) Books on 'Markov Chains'. (For stochastic matrices).

Upper and lower triangular matrices

- Definition 4.1 : A matrix $A = [a_{ij}]$ is said to be **upper triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq j < i \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 4.1** : A matrix $A = [a_{ij}]$ is said to be **upper triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq j < i \leq n.$$

- ▶ A matrix $A = [a_{ij}]$ is said to be **lower triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq i < j \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 4.1** : A matrix $A = [a_{ij}]$ is said to be **upper triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq j < i \leq n.$$

- ▶ A matrix $A = [a_{ij}]$ is said to be **lower triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq i < j \leq n.$$

Upper and lower triangular matrices

- **Definition 4.1** : A matrix $A = [a_{ij}]$ is said to be **upper triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq j < i \leq n.$$

- A matrix $A = [a_{ij}]$ is said to be **lower triangular** if

$$a_{ij} = 0, \quad \forall 1 \leq i < j \leq n.$$

- So if A is upper triangular, then it has the form:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.
- ▶ Then $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$ unless $i \leq \sigma(i)$ for every i .

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.
- ▶ Then $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$ unless $i \leq \sigma(i)$ for every i .
- ▶ But the only permutation σ which satisfies $i \leq \sigma(i)$ for all i , is the identity permutation (Why?). Now the result follows.

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.
- ▶ Then $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$ unless $i \leq \sigma(i)$ for every i .
- ▶ But the only permutation σ which satisfies $i \leq \sigma(i)$ for all i , is the identity permutation (Why?). Now the result follows.
- ▶ Alternatively, we may expand the determinant of A using first column and use induction.

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Liebnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.
- ▶ Then $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$ unless $i \leq \sigma(i)$ for every i .
- ▶ But the only permutation σ which satisfies $i \leq \sigma(i)$ for all i , is the identity permutation (Why?). Now the result follows.
- ▶ Alternatively, we may expand the determinant of A using first column and use induction.

Determinants of upper/lower triangular matrices

- ▶ **Theorem 4.2:** If a matrix $A = [a_{ij}]$ is upper triangular or lower triangular then the determinant of A is the product of its diagonal entries:

$$\det(A) = a_{11}a_{22} \cdots a_{nn}.$$

- ▶ **Proof.** We have Leibnitz formula:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

- ▶ Now assume that A is upper triangular. Then $a_{ij} = 0$ for $i > j$.
- ▶ Then $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = 0$ unless $i \leq \sigma(i)$ for every i .
- ▶ But the only permutation σ which satisfies $i \leq \sigma(i)$ for all i , is the identity permutation (Why?). Now the result follows.
- ▶ Alternatively, we may expand the determinant of A using first column and use induction.
- ▶ A similar proof works for lower triangular matrices through expansion using first row. ■

Partitioned vectors

- Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

Partitioned vectors

- Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

- We can view of the first m -coordinates of z as forming a vector in $\mathbb{R}^m$ and the remaining n -coordinates as forming a vector in $\mathbb{R}^n$.

Partitioned vectors

- Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

- We can view of the first m -coordinates of z as forming a vector in $\mathbb{R}^m$ and the remaining n -coordinates as forming a vector in $\mathbb{R}^n$.
- So we write

$$z = \begin{pmatrix} x \\ y \end{pmatrix}$$

where

$$x = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix}, \quad y = \begin{pmatrix} z_{m+1} \\ z_{m+2} \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

- ▶ Conversely, given any $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, we get a vector $z \in \mathbb{R}^{m+n}$ as

$$z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Continuation

- ▶ Conversely, given any $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, we get a vector $z \in \mathbb{R}^{m+n}$ as

$$z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

- ▶ So in a way, we can think of $\mathbb{R}^{m+n}$ as constructed out of $\mathbb{R}^m$ and $\mathbb{R}^n$. We say that $\mathbb{R}^{m+n}$ is direct sum of $\mathbb{R}^m$ and $\mathbb{R}^n$.

Partitioned matrices or block matrices

- ▶ Now consider a matrix $P = [p_{ij}]_{1 \leq i, j \leq (m+n)}$ considered as a linear map on $\mathbb{R}^{m+n}$.

Partitioned matrices or block matrices

- ▶ Now consider a matrix $P = [p_{ij}]_{1 \leq i, j \leq (m+n)}$ considered as a linear map on $\mathbb{R}^{m+n}$.
- ▶ We partition P as

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where $A_{m \times m}$, $B_{m \times n}$, $C_{n \times m}$, $D_{n \times n}$ are given by

$$A = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}, \quad B = \begin{bmatrix} p_{1(m+1)} & \cdots & p_{1(m+n)} \\ \vdots & \ddots & \vdots \\ p_{m(m+1)} & \cdots & p_{m(m+n)} \end{bmatrix}.$$

Continuation



$$C = \begin{bmatrix} p_{(m+1)1} & \cdots & p_{(m+1)m} \\ \vdots & \ddots & \vdots \\ p_{(m+n)1} & \cdots & p_{(m+n)(m)} \end{bmatrix},$$

Continuation

▶

$$C = \begin{bmatrix} p_{(m+1)1} & \cdots & p_{(m+1)m} \\ \vdots & \ddots & \vdots \\ p_{(m+n)1} & \cdots & p_{(m+n)(m)} \end{bmatrix},$$

▶

$$D = \begin{bmatrix} p_{(m+1)(m+1)} & \cdots & p_{(m+1)(m+n)} \\ \vdots & \ddots & \vdots \\ p_{(m+n)(m+1)} & \cdots & p_{(m+n)(m+n)} \end{bmatrix}$$

The action of partitioned matrices on vectors

- ▶ Under notation as above, with

$$Pz = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ax + By \\ Cx + Dy \end{pmatrix}.$$

The action of partitioned matrices on vectors

- Under notation as above, with

$$Pz = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ax + By \\ Cx + Dy \end{pmatrix}.$$

- Note that $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $B : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $C : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $D : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

Multiplication of partitioned matrices

- ▶ **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

- ▶ In other words, the multiplication is like the usual matrix multiplication.

Multiplication of partitioned matrices

- ▶ **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

- ▶ In other words, the multiplication is like the usual matrix multiplication.
- ▶ **Proof.** The proof is by direct multiplication.

Continuation

- For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

Continuation

- For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■
- ▶ More generally, if $P = [A_{ij}]$, $Q = [B_{kl}]$ are partitioned matrices, with matching orders, then PQ is a partitioned matrix $[C_{ij}]$ with

$$C_{ij} = \sum_k A_{ik} B_{kj}.$$

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■
- ▶ More generally, if $P = [A_{ij}]$, $Q = [B_{kl}]$ are partitioned matrices, with matching orders, then PQ is a partitioned matrix $[C_{ij}]$ with

$$C_{ij} = \sum_k A_{ik} B_{kj}.$$

- ▶ Here, for the matrix multiplication to be meaningful, it is necessary that for fixed i, k, j , if the order of A_{ik} is $a \times b$ then the order of B_{kj} should be $b \times c$ for some c . This is what we mean by 'matching orders'.

Transpose of block matrices

- ▶ With notation as before, for

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

Transpose of block matrices

- ▶ With notation as before, for

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$



$$P^t = \begin{bmatrix} A^t & C^t \\ B^t & D^t \end{bmatrix},$$

Transpose of block matrices

- ▶ With notation as before, for

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$



$$P^t = \begin{bmatrix} A^t & C^t \\ B^t & D^t \end{bmatrix},$$

- ▶ This is easy to see by direct verification.

Transpose of block matrices

- ▶ With notation as before, for

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$



$$P^t = \begin{bmatrix} A^t & C^t \\ B^t & D^t \end{bmatrix},$$

- ▶ This is easy to see by direct verification.
- ▶ More generally, if we have a partitioned matrix

$$P = [A_{ij}]$$

then

$$P^t = [(A_{ji})^t].$$

Block upper triangular matrices

- ▶ Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Block upper triangular matrices

- ▶ Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

- ▶ Then P is said to be block upper triangular if $C = 0$, that is, C is the zero matrix.

Block upper triangular matrices

- ▶ Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

- ▶ Then P is said to be block upper triangular if $C = 0$, that is, C is the zero matrix.
- ▶ Note that a block upper triangular matrix need not be upper triangular.

Block upper triangular matrices

- ▶ Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

- ▶ Then P is said to be block upper triangular if $C = 0$, that is, C is the zero matrix.
- ▶ Note that a block upper triangular matrix need not be upper triangular.
- ▶ Whether a given matrix is block upper triangular or not depends on the chosen sizes of blocks.

Block upper triangular matrices

- ▶ Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

- ▶ Then P is said to be block upper triangular if $C = 0$, that is, C is the zero matrix.
- ▶ Note that a block upper triangular matrix need not be upper triangular.
- ▶ Whether a given matrix is block upper triangular or not depends on the chosen sizes of blocks.
- ▶ In a similar way one can define block lower triangular matrices.

Determinants of block upper triangular matrices

- **Theorem 4.4:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then

$$\det(P) = \det(A) \cdot \det(D).$$

Determinants of block upper triangular matrices

- **Theorem 4.4:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then

$$\det(P) = \det(A) \cdot \det(D).$$

- **Proof:** Suppose P is of size $(m+n) \times (m+n)$ and A, B, D are respectively of sizes $m \times m$, $m \times n$ and $n \times n$.

Determinants of block upper triangular matrices

- ▶ **Theorem 4.4:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then

$$\det(P) = \det(A) \cdot \det(D).$$

- ▶ **Proof:** Suppose P is of size $(m+n) \times (m+n)$ and A, B, D are respectively of sizes $m \times m$, $m \times n$ and $n \times n$.
- ▶ Now

$$\det(P) = \sum_{\sigma \in S_{m+n}} \epsilon(\sigma) P_{1\sigma(1)} P_{2\sigma(2)} \cdots P_{(m+n)\sigma(m+n)}.$$

Determinants of block upper triangular matrices

- ▶ **Theorem 4.4:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then

$$\det(P) = \det(A) \cdot \det(D).$$

- ▶ **Proof:** Suppose P is of size $(m+n) \times (m+n)$ and A, B, D are respectively of sizes $m \times m$, $m \times n$ and $n \times n$.
- ▶ Now

$$\det(P) = \sum_{\sigma \in S_{m+n}} \epsilon(\sigma) P_{1\sigma(1)} P_{2\sigma(2)} \cdots P_{(m+n)\sigma(m+n)}.$$

- ▶ Now $C = 0$, means that $P_{j\sigma(j)} = 0$ if $(j, \sigma(j))$ are such that $(m+1) \leq j \leq (m+n)$ and $1 \leq \sigma(j) \leq m$.

Continuation

- ▶ In other words the term in the expansion of the determinant of P becomes 0 if σ maps $\{m+1, m+2, \dots, (m+n)\}$ to $\{1, 2, \dots, m\}$.

Continuation

- ▶ In other words the term in the expansion of the determinant of P becomes 0 if σ maps $\{m+1, m+2, \dots, (m+n)\}$ to $\{1, 2, \dots, m\}$.
- ▶ Therefore for the term to be non-zero, it is necessary that σ maps $\{m+1, \dots, (m+n)\}$ to itself. Consequently it also maps $\{1, 2, \dots, m\}$ to itself.

Continuation

- ▶ In other words the term in the expansion of the determinant of P becomes 0 if σ maps $\{m+1, m+2, \dots, (m+n)\}$ to $\{1, 2, \dots, m\}$.
- ▶ Therefore for the term to be non-zero, it is necessary that σ maps $\{m+1, \dots, (m+n)\}$ to itself. Consequently it also maps $\{1, 2, \dots, m\}$ to itself.
- ▶ Such permutations are precisely permutations of the form $\tau \circ \eta$ where τ is permutation of $\{1, 2, \dots, m\}$ considered as a permutation of $\{1, 2, \dots, (m+n)\}$ by taking $\tau(j) = j$ for $j \in \{m+1, \dots, (m+n)\}$ and η is a permutation of $\{m+1, \dots, m+n\}$ extended to $\{1, \dots, (m+n)\}$ by taking $\eta(j) = j$ for $1 \leq j \leq m$.

Continuation

- ▶ In other words the term in the expansion of the determinant of P becomes 0 if σ maps $\{m+1, m+2, \dots, (m+n)\}$ to $\{1, 2, \dots, m\}$.
- ▶ Therefore for the term to be non-zero, it is necessary that σ maps $\{m+1, \dots, (m+n)\}$ to itself. Consequently it also maps $\{1, 2, \dots, m\}$ to itself.
- ▶ Such permutations are precisely permutations of the form $\tau \circ \eta$ where τ is permutation of $\{1, 2, \dots, m\}$ considered as a permutation of $\{1, 2, \dots, (m+n)\}$ by taking $\tau(j) = j$ for $j \in \{m+1, \dots, (m+n)\}$ and η is a permutation of $\{m+1, \dots, m+n\}$ extended to $\{1, \dots, (m+n)\}$ by taking $\eta(j) = j$ for $1 \leq j \leq m$.
- ▶ Note that the signature of a permutation does not change by considering such extensions.

Continuation

- ▶ Then it is clear that,

$$\begin{aligned} & \det(P) \\ = & \sum_{\tau, \eta} \epsilon(\tau) \cdot \epsilon(\eta) p_{1\tau(1)} \cdots p_{m\tau(m)} \cdot p_{m+1\eta(m+1)} \cdots p_{m+n\eta(m+n)} \\ = & \det(A) \cdot \det(D). \blacksquare \end{aligned}$$

Continuation

- ▶ Then it is clear that,

$$\begin{aligned} & \det(P) \\ &= \sum_{\tau, \eta} \epsilon(\tau) \cdot \epsilon(\eta) p_{1\tau(1)} \cdots p_{m\tau(m)} \cdot p_{m+1\eta(m+1)} \cdots p_{m+n\eta(m+n)} \\ &= \det(A) \cdot \det(D). \blacksquare \end{aligned}$$

- ▶ Now by mathematical induction the determinant of a block upper triangular matrices (with square blocks on the diagonal) is the product of the determinants of diagonal blocks. That is,

$$\det \left(\begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1n} \\ 0 & P_{22} & P_{23} & \cdots & P_{2n} \\ 0 & 0 & P_{33} & \cdots & P_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & P_{nn} \end{bmatrix} \right) = \det(P_{11}) \cdots \det(P_{nn}).$$

if $P_{11}, P_{22}, \dots, P_{nn}$ are square blocks.

Inverses of 2×2 upper triangular matrices.

- **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

Inverses of 2×2 upper triangular matrices.

- **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

- From the formula $\det(P) = \det(A) \cdot \det(D)$, we know that if P is invertible, then $\det(A)$ and $\det(D)$ are non-zero and hence A, D are invertible.

Inverses of 2×2 upper triangular matrices.

- ▶ **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices and $C = 0$. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

- ▶ From the formula $\det(P) = \det(A) \cdot \det(D)$, we know that if P is invertible, then $\det(A)$ and $\det(D)$ are non-zero and hence A, D are invertible.
- ▶ The formula for the inverse can be confirmed by verifying:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

A special case

- Corollary 4.6: For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

A special case

- **Corollary 4.6:** For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

- **Proof:** The result is clear for $n = 0, 1$. Now verify the formula for $n \in \mathbb{N}$ by induction. Taking inverses we have the result for all $n \in \mathbb{Z}$.

A special case

- **Corollary 4.6:** For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

- **Proof:** The result is clear for $n = 0, 1$. Now verify the formula for $n \in \mathbb{N}$ by induction. Taking inverses we have the result for all $n \in \mathbb{Z}$.
- This is actually a consequence of

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & B + C \\ 0 & I \end{bmatrix}.$$

The matrix product becomes simple addition here.

A special case

- ▶ **Corollary 4.6:** For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

- ▶ **Proof:** The result is clear for $n = 0, 1$. Now verify the formula for $n \in \mathbb{N}$ by induction. Taking inverses we have the result for all $n \in \mathbb{Z}$.
- ▶ This is actually a consequence of

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & B + C \\ 0 & I \end{bmatrix}.$$

The matrix product becomes simple addition here.

- ▶ **END OF LECTURE 4.**

Lecture 5: Determinant and inverses of block matrices

- We recall: Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

Lecture 5: Determinant and inverses of block matrices

- We recall: Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

Lecture 5: Determinant and inverses of block matrices

- ▶ We recall: Fix $m, n \in \mathbb{N}$. Consider a vector $z \in \mathbb{R}^{m+n}$:

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

- ▶ We can view of the first m -coordinates of z as forming a vector in $\mathbb{R}^m$ and the remaining n -coordinates as forming a vector in $\mathbb{R}^n$.
- ▶ So we write

$$z = \begin{pmatrix} x \\ y \end{pmatrix}$$

where

$$x = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix}, \quad y = \begin{pmatrix} z_{m+1} \\ z_{m+2} \\ \vdots \\ z_{m+n} \end{pmatrix}.$$

- ▶ Conversely, given any $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, we get a vector $z \in \mathbb{R}^{m+n}$ as

$$z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Continuation

- ▶ Conversely, given any $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, we get a vector $z \in \mathbb{R}^{m+n}$ as

$$z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

- ▶ So in a way, we can think of $\mathbb{R}^{m+n}$ as constructed out of $\mathbb{R}^m$ and $\mathbb{R}^n$. We say that $\mathbb{R}^{m+n}$ is direct sum of $\mathbb{R}^m$ and $\mathbb{R}^n$.

Partitioned matrices or block matrices

- ▶ Now consider a matrix $P = [p_{ij}]_{1 \leq i, j \leq (m+n)}$ considered as a linear map on $\mathbb{R}^{m+n}$.

Partitioned matrices or block matrices

- ▶ Now consider a matrix $P = [p_{ij}]_{1 \leq i, j \leq (m+n)}$ considered as a linear map on $\mathbb{R}^{m+n}$.
- ▶ We partition P as

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where $A_{m \times m}$, $B_{m \times n}$, $C_{n \times m}$, $D_{n \times n}$ are given by

$$A = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}, \quad B = \begin{bmatrix} p_{1(m+1)} & \cdots & p_{1(m+n)} \\ \vdots & \ddots & \vdots \\ p_{m(m+1)} & \cdots & p_{m(m+n)} \end{bmatrix}.$$

Continuation



$$C = \begin{bmatrix} p_{(m+1)1} & \cdots & p_{(m+1)m} \\ \vdots & \ddots & \vdots \\ p_{(m+n)1} & \cdots & p_{(m+n)(m)} \end{bmatrix},$$

Continuation

▶

$$C = \begin{bmatrix} p_{(m+1)1} & \cdots & p_{(m+1)m} \\ \vdots & \ddots & \vdots \\ p_{(m+n)1} & \cdots & p_{(m+n)(m)} \end{bmatrix},$$

▶

$$D = \begin{bmatrix} p_{(m+1)(m+1)} & \cdots & p_{(m+1)(m+n)} \\ \vdots & \ddots & \vdots \\ p_{(m+n)(m+1)} & \cdots & p_{(m+n)(m+n)} \end{bmatrix}$$

The action of partitioned matrices on vectors

- ▶ Under notation as above, with

$$Pz = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ax + By \\ Cx + Dy \end{pmatrix}.$$

The action of partitioned matrices on vectors

- Under notation as above, with

$$Pz = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ax + By \\ Cx + Dy \end{pmatrix}.$$

- Note that $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $B : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $C : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $D : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then

Multiplication of partitioned matrices

- **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

Multiplication of partitioned matrices

- ▶ **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

- ▶ In other words, the multiplication is like the usual matrix multiplication.

Multiplication of partitioned matrices

- ▶ **Theorem 4.3:** Consider two partitioned matrices

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

and



$$Q = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

with matching sizes. Then



$$PQ = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix},$$

- ▶ In other words, the multiplication is like the usual matrix multiplication.
- ▶ **Proof.** The proof is by direct multiplication.

Continuation

- For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

Continuation

- For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■
- ▶ More generally, if $P = [A_{ij}]$, $Q = [B_{kl}]$ are partitioned matrices, with matching orders, then PQ is a partitioned matrix $[C_{ij}]$ with

$$C_{ij} = \sum_k A_{ik} B_{kj}.$$

Continuation

- ▶ For instance, for $1 \leq i, j \leq m$,

$$\begin{aligned}(PQ)_{ij} &= \sum_{k=1}^{m+n} p_{ik} q_{kj} = \sum_{k=1}^m p_{ik} q_{kj} + \sum_{k=m+1}^{m+n} p_{ik} q_{kj} \\ &= (AE)_{ij} + (BG)_{ij}\end{aligned}$$

- ▶ Similar computations work for other coordinates. ■
- ▶ More generally, if $P = [A_{ij}]$, $Q = [B_{kl}]$ are partitioned matrices, with matching orders, then PQ is a partitioned matrix $[C_{ij}]$ with

$$C_{ij} = \sum_k A_{ik} B_{kj}.$$

- ▶ Here, for the matrix multiplication to be meaningful, it is necessary that for fixed i, k, j , if the order of A_{ik} is $a \times b$ then the order of B_{kj} should be $b \times c$ for some c . This is what we mean by 'matching orders'.

Determinants of block upper triangular matrices

- **Theorem 4.4:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices. Then

$$\det(P) = \det(A) \cdot \det(D).$$

Inverses of 2×2 upper triangular matrices.

► **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

Inverses of 2×2 upper triangular matrices.

- **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

- From the formula $\det(P) = \det(A) \cdot \det(D)$, we know that if P is invertible, then $\det(A)$ and $\det(D)$ are non-zero and hence A, D are invertible.

Inverses of 2×2 upper triangular matrices.

- ▶ **Theorem 4.5:** Consider a block upper triangular matrix

$$P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where A, D are square matrices. Then P is invertible if and only if A and D are invertible and in such a case,

$$P^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

- ▶ From the formula $\det(P) = \det(A) \cdot \det(D)$, we know that if P is invertible, then $\det(A)$ and $\det(D)$ are non-zero and hence A, D are invertible.
- ▶ The formula for the inverse can be confirmed by verifying:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

A special case

► Corollary 4.6: For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

A special case

- **Corollary 4.6:** For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

- **Proof:** The result is clear for $n = 0, 1$. Now verify the formula for $n \in \mathbb{N}$ by induction. Taking inverses we have the result for all $n \in \mathbb{Z}$.

A special case

- Corollary 4.6: For any matrix B ,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^n = \begin{bmatrix} I & nB \\ 0 & I \end{bmatrix}$$

for every $n \in \mathbb{Z}$.

- **Proof:** The result is clear for $n = 0, 1$. Now verify the formula for $n \in \mathbb{N}$ by induction. Taking inverses we have the result for all $n \in \mathbb{Z}$.
- This is actually a consequence of

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & B + C \\ 0 & I \end{bmatrix}.$$

The matrix product becomes simple addition here.

A factorization theorem for 2×2 block matrices

- Theorem 5.1: Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

A factorization theorem for 2×2 block matrices

- **Theorem 5.1:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- (i) If D is invertible, then

$$P = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}.$$

A factorization theorem for 2×2 block matrices

- **Theorem 5.1:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- (i) If D is invertible, then

$$P = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}.$$

- (ii) If A is invertible, then

$$P = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}.$$

A factorization theorem for 2×2 block matrices

- ▶ **Theorem 5.1:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- ▶ (i) If D is invertible, then

$$P = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}.$$

- ▶ (ii) If A is invertible, then

$$P = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}.$$

- ▶ **Remark** The terms $A - BD^{-1}C$ and $D - CA^{-1}B$ appearing above are known as **Schur Complements** and they appear in various block matrix computations.

A factorization theorem for 2×2 block matrices

- ▶ **Theorem 5.1:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- ▶ (i) If D is invertible, then

$$P = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}.$$

- ▶ (ii) If A is invertible, then

$$P = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}.$$

- ▶ **Remark** The terms $A - BD^{-1}C$ and $D - CA^{-1}B$ appearing above are known as **Schur Complements** and they appear in various block matrix computations.
- ▶ Here A and D need not be of same order.

Continuation

► **Proof.** By direct computation:

$$\begin{aligned} & \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\ = & \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\ = & \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \end{aligned}$$

Continuation

- **Proof.** By direct computation:

$$\begin{aligned} & \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\ = & \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\ = & \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \end{aligned}$$

- This proves (i). Similarly (ii) follows by multiplication. ■

Determinant of 2×2 block matrices

- Theorem 5.2: Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

Determinant of 2×2 block matrices

- Theorem 5.2: Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- (i) If D is invertible, then

$$\det(P) = \det(A - BD^{-1}C) \cdot \det(D).$$

Determinant of 2×2 block matrices

- ▶ **Theorem 5.2:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- ▶ (i) If D is invertible, then

$$\det(P) = \det(A - BD^{-1}C) \cdot \det(D).$$

- ▶ (ii) If A is invertible, then

$$\det(P) = \det(A) \cdot \det(D - CA^{-1}B).$$

Determinant of 2×2 block matrices

- ▶ **Theorem 5.2:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- ▶ (i) If D is invertible, then

$$\det(P) = \det(A - BD^{-1}C) \cdot \det(D).$$

- ▶ (ii) If A is invertible, then

$$\det(P) = \det(A) \cdot \det(D - CA^{-1}B).$$

- ▶ **Proof.** Clear from the factorization result and the fact that the determinant of a triangular block matrix is the product of determinants of diagonal blocks.

Inverses of 2×2 block matrices

- **Theorem 5.3:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

Inverses of 2×2 block matrices

- **Theorem 5.3:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- (i) Assume D is invertible and $S := (A - BD^{-1}C)$ is invertible. Then P is invertible and

$$P^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{bmatrix}$$

Inverses of 2×2 block matrices

- **Theorem 5.3:** Consider a block matrix

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices.

- (i) Assume D is invertible and $S := (A - BD^{-1}C)$ is invertible. Then P is invertible and

$$P^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{bmatrix}$$

- (ii) If A is invertible, and $T := D - CA^{-1}B$ is invertible, then P is invertible and

$$P^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BT^{-1}CA^{-1} & -A^{-1}BT^{-1} \\ -T^{-1}CA^{-1} & T^{-1} \end{bmatrix}.$$

Some special cases

► Theorem 5.4: Suppose

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, B, C, D are square matrices of same sizes and C, D commute ($CD = DC$). Then

$$\det(P) = \det(AD - BC).$$

Some special cases

- Theorem 5.4: Suppose

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, B, C, D are square matrices of same sizes and C, D commute ($CD = DC$). Then

$$\det(P) = \det(AD - BC).$$

- Theorem 5.5: Suppose

$$P = \begin{bmatrix} A & B \\ B & A \end{bmatrix}$$

where A, B , are square matrices. Then

$$\det(P) = \det(A - B) \cdot \det(A + B).$$

Some special cases

- Theorem 5.4: Suppose

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, B, C, D are square matrices of same sizes and C, D commute ($CD = DC$). Then

$$\det(P) = \det(AD - BC).$$

- Theorem 5.5: Suppose

$$P = \begin{bmatrix} A & B \\ B & A \end{bmatrix}$$

where A, B , are square matrices. Then

$$\det(P) = \det(A - B) \cdot \det(A + B).$$

- Exercise: Prove these theorems.

Matrix tricks

- ▶ How to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where B, C are $n \times n$ square matrices and either B or C is invertible.

Matrix tricks

- ▶ How to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where B, C are $n \times n$ square matrices and either B or C is invertible.

- ▶ Take

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

of same order.

Matrix tricks

- ▶ How to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where B, C are $n \times n$ square matrices and either B or C is invertible.

- ▶ Take

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

of same order.

- ▶ Then $\det(J) = (-1)^n$. (Prove this!.)

Matrix tricks

- ▶ How to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where B, C are $n \times n$ square matrices and either B or C is invertible.

- ▶ Take

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

of same order.

- ▶ Then $\det(J) = (-1)^n$. (Prove this!.)
- ▶ Now

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \cdot \begin{bmatrix} C & D \\ A & B \end{bmatrix}$$

- Therefore,

$$\det(P) = (-1)^n \cdot \det\left(\begin{bmatrix} C & D \\ A & B \end{bmatrix}\right)$$

which can be computed using the formulae derived earlier.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.
- ▶ Hence there exists $\epsilon > 0$ such that $f(t) \neq 0$ for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.
- ▶ Hence there exists $\epsilon > 0$ such that $f(t) \neq 0$ for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ Therefore $A + tI$ is invertible for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.
- ▶ Hence there exists $\epsilon > 0$ such that $f(t) \neq 0$ for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ Therefore $A + tI$ is invertible for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ So using results proved earlier we can compute the determinant of P_t for $t \in (-\epsilon, +\epsilon) \setminus \{0\}$.

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.
- ▶ Hence there exists $\epsilon > 0$ such that $f(t) \neq 0$ for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ Therefore $A + tI$ is invertible for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ So using results proved earlier we can compute the determinant of P_t for $t \in (-\epsilon, +\epsilon) \setminus \{0\}$.
- ▶ Taking the limit as t tends to 0, we get the determinant of P .

Another trick

- ▶ Suppose we want to compute the determinant of

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, D are square matrices, but not invertible.

- ▶ For $t \in \mathbb{R}$ consider

$$P_t = P + tI = \begin{bmatrix} A + tI & B \\ C & D + tI \end{bmatrix}$$

- ▶ Consider $f(t) = \det(A + tI)$ for $t \in \mathbb{R}$.
- ▶ Then f is a polynomial in t . So it has finite number of zeros.
- ▶ Hence there exists $\epsilon > 0$ such that $f(t) \neq 0$ for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ Therefore $A + tI$ is invertible for $t \in (-\epsilon, \epsilon) \setminus \{0\}$.
- ▶ So using results proved earlier we can compute the determinant of P_t for $t \in (-\epsilon, +\epsilon) \setminus \{0\}$.
- ▶ Taking the limit as t tends to 0, we get the determinant of P .
- ▶ **END OF LECTURE 5.**

Lecture 6: Cauchy Binet formula

- ▶ Suppose A, B are $m \times m$ matrices. Then we know that

$$\det(AB) = \det(A) \cdot \det(B).$$

Lecture 6: Cauchy Binet formula

- ▶ Suppose A, B are $m \times m$ matrices. Then we know that

$$\det(AB) = \det(A) \cdot \det(B).$$

- ▶ Now suppose A is an $m \times n$ matrix and B is a $n \times m$ matrix so that AB is again an $m \times m$ matrix.

Lecture 6: Cauchy Binet formula

- ▶ Suppose A, B are $m \times m$ matrices. Then we know that

$$\det(AB) = \det(A) \cdot \det(B).$$

- ▶ Now suppose A is an $m \times n$ matrix and B is a $n \times m$ matrix so that AB is again an $m \times m$ matrix.
- ▶ **Question:** How can we compute the determinant of AB using A and B ?

Lecture 6: Cauchy Binet formula

- ▶ Suppose A, B are $m \times m$ matrices. Then we know that

$$\det(AB) = \det(A) \cdot \det(B).$$

- ▶ Now suppose A is an $m \times n$ matrix and B is a $n \times m$ matrix so that AB is again an $m \times m$ matrix.
- ▶ **Question:** How can we compute the determinant of AB using A and B ?
- ▶ This question is answered by Cauchy-Binet formula.

Illustration



$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 5 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 \\ 7 & 8 \\ 0 & 1 \end{bmatrix}$$

Illustration



$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 5 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 \\ 7 & 8 \\ 0 & 1 \end{bmatrix}$$

▶ Then

$$\det(AB) = \det(A_1B_1) + \det(A_2B_2) + \det(A_3B_3),$$

Illustration



$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 5 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 \\ 7 & 8 \\ 0 & 1 \end{bmatrix}$$

► Then

$$\det(AB) = \det(A_1B_1) + \det(A_2B_2) + \det(A_3B_3),$$

► where A_j, B_j 's are 2×2 matrices formed by choosing columns of A and respective rows of B :

$$A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 4 & 6 \\ 7 & 8 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 4 & 6 \\ 0 & 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 1 & 0 \\ 5 & 7 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 7 & 8 \\ 0 & 1 \end{bmatrix}.$$

Notation

- ▶ To state the theorem we need some notation.

Notation

- ▶ To state the theorem we need some notation.
- ▶ Fix $m, n \in \mathbb{N}$ with $m \leq n$.

Notation

- ▶ To state the theorem we need some notation.
- ▶ Fix $m, n \in \mathbb{N}$ with $m \leq n$.
- ▶ Suppose $B = [b_{jk}]_{1 \leq j \leq n, 1 \leq k \leq m}$ so that B is an $n \times m$ matrix.

Notation

- ▶ To state the theorem we need some notation.
- ▶ Fix $m, n \in \mathbb{N}$ with $m \leq n$.
- ▶ Suppose $B = [b_{jk}]_{1 \leq j \leq n, 1 \leq k \leq m}$ so that B is an $n \times m$ matrix.
- ▶ We form $m \times m$ square matrices out of B by choosing m -rows of B .

Notation

- ▶ To state the theorem we need some notation.
- ▶ Fix $m, n \in \mathbb{N}$ with $m \leq n$.
- ▶ Suppose $B = [b_{jk}]_{1 \leq j \leq n, 1 \leq k \leq m}$ so that B is an $n \times m$ matrix.
- ▶ We form $m \times m$ square matrices out of B by choosing m -rows of B .
- ▶ So if $1 \leq j_1, j_2, \dots, j_m \leq n$, we take

$$B(j_1, j_2, \dots, j_m | 1, 2, \dots, m) = \begin{bmatrix} b_{j_1 1} & b_{j_1 2} & \dots & b_{j_1 m} \\ b_{j_2 1} & b_{j_2 2} & \dots & b_{j_2 m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{j_m 1} & b_{j_m 2} & \dots & b_{j_m m} \end{bmatrix}.$$

Notation

- ▶ To state the theorem we need some notation.
- ▶ Fix $m, n \in \mathbb{N}$ with $m \leq n$.
- ▶ Suppose $B = [b_{jk}]_{1 \leq j \leq n, 1 \leq k \leq m}$ so that B is an $n \times m$ matrix.
- ▶ We form $m \times m$ square matrices out of B by choosing m -rows of B .
- ▶ So if $1 \leq j_1, j_2, \dots, j_m \leq n$, we take

$$B(j_1, j_2, \dots, j_m | 1, 2, \dots, m) = \begin{bmatrix} b_{j_1 1} & b_{j_1 2} & \dots & b_{j_1 m} \\ b_{j_2 1} & b_{j_2 2} & \dots & b_{j_2 m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{j_m 1} & b_{j_m 2} & \dots & b_{j_m m} \end{bmatrix}.$$

- ▶ The notation indicates that the rows chosen are $j_1, j_2, \dots, j_m$, and columns chosen are $1, 2, \dots, m$.

Continuation

- ▶ Similarly if $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ is an $m \times n$ matrix with $m \leq n$, we form square matrices by choosing m -columns of A .

Continuation

- ▶ Similarly if $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ is an $m \times n$ matrix with $m \leq n$, we form square matrices by choosing m -columns of A .
- ▶ If $1 \leq j_1, j_2, \dots, j_m \leq n$, take

$$A(1, 2, \dots, m | j_1, j_2, \dots, j_m) := \begin{bmatrix} a_{1j_1} & a_{1j_2} & \dots & a_{1j_m} \\ a_{2j_1} & a_{2j_2} & \dots & a_{2j_m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{mj_1} & a_{mj_2} & \dots & a_{mj_m} \end{bmatrix}.$$

Continuation

- ▶ Similarly if $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ is an $m \times n$ matrix with $m \leq n$, we form square matrices by choosing m -columns of A .
- ▶ If $1 \leq j_1, j_2, \dots, j_m \leq n$, take

$$A(1, 2, \dots, m | j_1, j_2, \dots, j_m) := \begin{bmatrix} a_{1j_1} & a_{1j_2} & \dots & a_{1j_m} \\ a_{2j_1} & a_{2j_2} & \dots & a_{2j_m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{mj_1} & a_{mj_2} & \dots & a_{mj_m} \end{bmatrix}.$$

- ▶ Here all rows are chosen and columns $j_1, j_2, \dots, j_m$ are chosen to get a square matrix.

A Lemma

- **Lemma 6.1:** For any $n \times m$ matrix B ,

$$\det B(j_1, j_2, \dots, j_m | 1, 2, \dots, m) = 0$$

if $j_1, \dots, j_m$ are not distinct. If $j_1, j_2, \dots, j_m$ are distinct, then

$$\begin{aligned} & \det B(j_1, j_2, \dots, j_m | 1, \dots, m) \\ &= \epsilon(\tau) \det B(j_{\tau(1)}, j_{\tau(2)}, \dots, j_{\tau(m)} | 1, \dots, m) \end{aligned}$$

where $\tau \in S_m$ is the permutation such that $j_{\tau(1)} < j_{\tau(2)} < \dots < j_{\tau(m)}$.

A Lemma

- **Lemma 6.1:** For any $n \times m$ matrix B ,

$$\det B(j_1, j_2, \dots, j_m | 1, 2, \dots, m) = 0$$

if $j_1, \dots, j_m$ are not distinct. If $j_1, j_2, \dots, j_m$ are distinct, then

$$\begin{aligned} & \det B(j_1, j_2, \dots, j_m | 1, \dots, m) \\ &= \epsilon(\tau) \det B(j_{\tau(1)}, j_{\tau(2)}, \dots, j_{\tau(m)} | 1, \dots, m) \end{aligned}$$

where $\tau \in S_m$ is the permutation such that $j_{\tau(1)} < j_{\tau(2)} < \dots < j_{\tau(m)}$.

- **Proof.** Follows from the basic properties of the determinant.

Cauchy Binet formula

- **Theorem 6.2:** Suppose A, B are $m \times n$ and $n \times m$ matrices with $m \leq n$ and $C = AB$. Then $\det(C) =$

$$\sum_{1 \leq j_1 < \dots < j_m \leq n} \det(A(1, \dots, m | j_1, \dots, j_m)) \cdot \det(B(j_1, \dots, j_m | 1, \dots, m)).$$

Cauchy Binet formula

- **Theorem 6.2:** Suppose A, B are $m \times n$ and $n \times m$ matrices with $m \leq n$ and $C = AB$. Then $\det(C) =$

$$\sum_{1 \leq j_1 < \dots < j_m \leq n} \det(A(1, \dots, m | j_1, \dots, j_m)) \cdot \det(B(j_1, \dots, j_m | 1, \dots, m)).$$

- Note that there are $\binom{n}{m}$ terms in this summation.

Cauchy Binet formula

- **Theorem 6.2:** Suppose A, B are $m \times n$ and $n \times m$ matrices with $m \leq n$ and $C = AB$. Then $\det(C) =$

$$\sum_{1 \leq j_1 < \dots < j_m \leq n} \det(A(1, \dots, m | j_1, \dots, j_m)) \cdot \det(B(j_1, \dots, j_m | 1, \dots, m)).$$

- Note that there are $\binom{n}{m}$ terms in this summation.
- **Proof.** We have

$$\begin{aligned} & \det(AB) \\ &= \sum_{\sigma \in S_m} \epsilon(\sigma) (AB)_{1\sigma(1)} (AB)_{2\sigma(2)} \dots (AB)_{m\sigma(m)} \\ &= \sum_{\sigma \in S_m} \epsilon(\sigma) \left(\sum_{j=1}^n a_{1j} b_{j\sigma(1)} \right) \left(\sum_{j=1}^n a_{2j} b_{j\sigma(2)} \right) \dots \left(\sum_{j=1}^n a_{mj} b_{j\sigma(m)} \right) \end{aligned}$$

Continuation

$$\begin{aligned} &= \sum_{j_1, j_2, \dots, j_m=1}^n a_{1j_1} \dots a_{mj_m} \sum_{\sigma \in S_m} \epsilon(\sigma) b_{j_1\sigma(1)} b_{j_2\sigma(2)} \dots b_{j_m\sigma(m)} \\ &= \sum_{j_1, j_2, \dots, j_m=1}^n a_{1j_1} \dots a_{mj_m} \det B(j_1, \dots, j_m | 1, \dots, m) \\ &= \sum_{j_1, j_2, \dots, j_m \text{ distinct}} a_{1j_1} \dots a_{mj_m} \det B(j_1, \dots, j_m | 1, \dots, m) \\ &= \sum_{1 \leq j_1 < \dots < j_m \leq n} \sum_{\tau \in S_m} a_{1j_{\tau(1)}} \dots a_{mj_{\tau(m)}} \det B(j_{\tau(1)}, \dots, j_{\tau(m)} | 1, \dots, m) \\ &= \sum_{1 \leq j_1 < \dots < j_m \leq n, \tau \in S_m} a_{1j_{\tau(1)}} \dots a_{mj_{\tau(m)}} \epsilon(\tau) \det B(j_1, \dots, j_m | 1, \dots, m) \\ &= \sum_{1 \leq j_1 < \dots < j_m \leq n} \det A(1, \dots, m | j_1, \dots, j_m) \det B(j_1, \dots, j_m | 1, \dots, m) \end{aligned}$$

This is what we wanted to prove. ■

Continuation

- ▶ What happens if $m > n$.

Continuation

- ▶ What happens if $m > n$.
- ▶ We have $C = AB$, where A is an $m \times n$ matrix and B is an $n \times m$ matrix.

Continuation

- ▶ What happens if $m > n$.
- ▶ We have $C = AB$, where A is an $m \times n$ matrix and B is an $n \times m$ matrix.
- ▶ Now $m > n$, means that we can't form $m \times m$ matrices using distinct columns of A . This suggests that the determinant should be zero.

Continuation

- ▶ What happens if $m > n$.
- ▶ We have $C = AB$, where A is an $m \times n$ matrix and B is an $n \times m$ matrix.
- ▶ Now $m > n$, means that we can't form $m \times m$ matrices using distinct columns of A . This suggests that the determinant should be zero.
- ▶ **Exercise:** Show that $\det(AB) = 0$ when A, B are as above. (Hint: $\text{rank}(AB) \leq \min(\text{rank } A, \text{rank } B)$.)

Continuation

- ▶ What happens if $m > n$.
- ▶ We have $C = AB$, where A is an $m \times n$ matrix and B is an $n \times m$ matrix.
- ▶ Now $m > n$, means that we can't form $m \times m$ matrices using distinct columns of A . This suggests that the determinant should be zero.
- ▶ **Exercise:** Show that $\det(AB) = 0$ when A, B are as above. (Hint: $\text{rank}(AB) \leq \min(\text{rank } A, \text{rank } B)$.)
- ▶ Look up Wikipedia for a good account of Cauchy Binet formula.

Continuation

- ▶ What happens if $m > n$.
- ▶ We have $C = AB$, where A is an $m \times n$ matrix and B is an $n \times m$ matrix.
- ▶ Now $m > n$, means that we can't form $m \times m$ matrices using distinct columns of A . This suggests that the determinant should be zero.
- ▶ **Exercise:** Show that $\det(AB) = 0$ when A, B are as above. (Hint: $\text{rank}(AB) \leq \min(\text{rank } A, \text{rank } B)$.)
- ▶ Look up Wikipedia for a good account of Cauchy Binet formula.
- ▶ **END OF LECTURE 6.**

Lecture 7: Inner product spaces

- ▶ You have studied vector spaces. Now we introduce a new structure on them called 'inner product'.

Lecture 7: Inner product spaces

- ▶ You have studied vector spaces. Now we introduce a new structure on them called 'inner product'.
- ▶ This abstractly captures the notions of 'length' and 'angle'.

Lecture 7: Inner product spaces

- ▶ You have studied vector spaces. Now we introduce a new structure on them called 'inner product'.
- ▶ This abstractly captures the notions of 'length' and 'angle'.
- ▶ Once we have an inner product we can talk about the distance between elements of the vector space. This allows us to define convergence of a sequence vectors.

Lecture 7: Inner product spaces

- ▶ You have studied vector spaces. Now we introduce a new structure on them called 'inner product'.
- ▶ This abstractly captures the notions of 'length' and 'angle'.
- ▶ Once we have an inner product we can talk about the distance between elements of the vector space. This allows us to define convergence of a sequence vectors.
- ▶ The notion of inner product also allows us to define as to when one vector is 'orthogonal' to another.

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ We recall that for any complex number $z = a + ib$ with $a, b \in \mathbb{R}$, $\bar{z} := a - ib$.

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ We recall that for any complex number $z = a + ib$ with $a, b \in \mathbb{R}$, $\bar{z} := a - ib$.
- ▶ For any two complex numbers z, w we have $\overline{(zw)} = \bar{z} \cdot \bar{w}$.

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ We recall that for any complex number $z = a + ib$ with $a, b \in \mathbb{R}$, $\bar{z} := a - ib$.
- ▶ For any two complex numbers z, w we have $\overline{(zw)} = \bar{z} \cdot \bar{w}$.
- ▶ For $z \in \mathbb{C}$ as above,

$$|z| := (z\bar{z})^{\frac{1}{2}} = (a^2 + b^2)^{\frac{1}{2}}.$$

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ We recall that for any complex number $z = a + ib$ with $a, b \in \mathbb{R}$, $\bar{z} := a - ib$.
- ▶ For any two complex numbers z, w we have $\overline{(zw)} = \bar{z} \cdot \bar{w}$.
- ▶ For $z \in \mathbb{C}$ as above,

$$|z| := (z\bar{z})^{\frac{1}{2}} = (a^2 + b^2)^{\frac{1}{2}}.$$

- ▶ Recall that any complex number $z \neq 0$ has the unique polar decomposition as $z = re^{i\theta}$ where $r = |z|$ and $0 \leq \theta < 2\pi$.

The setting

- ▶ In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ We recall that for any complex number $z = a + ib$ with $a, b \in \mathbb{R}$, $\bar{z} := a - ib$.
- ▶ For any two complex numbers z, w we have $\overline{(zw)} = \bar{z} \cdot \bar{w}$.
- ▶ For $z \in \mathbb{C}$ as above,

$$|z| := (z\bar{z})^{\frac{1}{2}} = (a^2 + b^2)^{\frac{1}{2}}.$$

- ▶ Recall that any complex number $z \neq 0$ has the unique polar decomposition as $z = re^{i\theta}$ where $r = |z|$ and $0 \leq \theta < 2\pi$.
- ▶ We have $|z| = 0$ if and only if $z = 0$. Further, $|zw| = |z||w|$ and $|z + w| \leq |z| + |w|$ for all $z, w \in \mathbb{C}$.

Standard inner products

- ▶ Consider a vector space V over a field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$).

Standard inner products

- ▶ Consider a vector space V over a field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$).
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).

Standard inner products

- ▶ Consider a vector space V over a field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$).
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:

Standard inner products

- ▶ Consider a vector space V over a field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$).
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:
- ▶ **Definition 7.1:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{R}^n$. The **standard inner product** on $\mathbb{R}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n x_j y_j.$$

Standard inner products

- ▶ Consider a vector space V over a field $\mathbb{F}$ (which is either $\mathbb{R}$ or $\mathbb{C}$).
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:
- ▶ **Definition 7.1:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{R}^n$. The **standard inner product** on $\mathbb{R}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n x_j y_j.$$

- ▶ We note that $\langle x, x \rangle \geq 0$ for every x in $\mathbb{R}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

$\mathbb{C}^n$. as an inner product space

- **Definition 7.2:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{C}^n$. The **standard inner product** on $\mathbb{C}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n \overline{x_j} y_j.$$

$\mathbb{C}^n$. as an inner product space

- **Definition 7.2:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{C}^n$. The **standard inner product** on $\mathbb{C}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n \overline{x_j} y_j.$$

- Here also we note that $\langle x, x \rangle \geq 0$ for every x in $\mathbb{C}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Definitions

- **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)
- ▶ (iv) $\langle x, x \rangle = 0$ if and only if $x = 0$. (Definiteness.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)
- ▶ (iv) $\langle x, x \rangle = 0$ if and only if $x = 0$. (Definiteness.)
- ▶ Some authors take inner product as linear in the first variable. It is a matter of convention. A vector space with a specified inner product is called an inner product space.

Examples

- ▶ It is clear that standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner products.

Examples

- ▶ It is clear that standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner products.
- ▶ **Example 7.4:** For $n \in \mathbb{N}$, fix scalars $a_1, a_2, \dots, a_n$.

Examples

- ▶ It is clear that standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner products.
- ▶ **Example 7.4:** For $n \in \mathbb{N}$, fix scalars $a_1, a_2, \dots, a_n$.
- ▶ Then

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n a_j x_j y_j,$$

Examples

- ▶ It is clear that standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner products.
- ▶ **Example 7.4:** For $n \in \mathbb{N}$, fix scalars $a_1, a_2, \dots, a_n$.
- ▶ Then

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n a_j x_j y_j,$$

- ▶ is an inner-product on $\mathbb{R}^n$ if and only if $a_j > 0$ for every j .

Examples

- ▶ It is clear that standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$ are inner products.

- ▶ **Example 7.4:** For $n \in \mathbb{N}$, fix scalars $a_1, a_2, \dots, a_n$.

- ▶ Then

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n a_j x_j y_j,$$

- ▶ is an inner-product on $\mathbb{R}^n$ if and only if $a_j > 0$ for every j .
- ▶ Note that if $a_j \geq 0$, then conditions (i)-(iii) of the inner product are satisfied but the definiteness may not be satisfied. In such cases, $\langle \cdot, \cdot \rangle$ is known as semi-inner product.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and (conjugate) symmetry of the inner-product we get also linearity in the first variable.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and (conjugate) symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable in the real case.)

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and (conjugate) symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable in the real case.)
- ▶ On the other hand if the field $\mathbb{F} = \mathbb{C}$, from linearity and conjugate symmetry of the inner-product we get anti-linearity in the first variable.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and (conjugate) symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable in the real case.)
- ▶ On the other hand if the field $\mathbb{F} = \mathbb{C}$, from linearity and conjugate symmetry of the inner-product we get anti-linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = \bar{c}\langle x, z \rangle + \bar{d}\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$.

Another example

- ▶ Take $\mathbb{F} = \mathbb{C}$. Fix $m, n \in \mathbb{N}$. Then we know that $M_{m,n}(\mathbb{C})$ is a vector space of dimension mn over $\mathbb{C}$.

Another example

- ▶ Take $\mathbb{F} = \mathbb{C}$. Fix $m, n \in \mathbb{N}$. Then we know that $M_{m,n}(\mathbb{C})$ is a vector space of dimension mn over $\mathbb{C}$.
- ▶ For $X, Y \in M_{m,n}(\mathbb{C})$ take

$$\langle X, Y \rangle = \text{trace}(X^* Y)$$

where $(X^*)_{jk} = \overline{x_{kj}}$, $1 \leq k \leq m$; $1 \leq j \leq n$.

Another example

- ▶ Take $\mathbb{F} = \mathbb{C}$. Fix $m, n \in \mathbb{N}$. Then we know that $M_{m,n}(\mathbb{C})$ is a vector space of dimension mn over $\mathbb{C}$.
- ▶ For $X, Y \in M_{m,n}(\mathbb{C})$ take

$$\langle X, Y \rangle = \text{trace}(X^* Y)$$

where $(X^*)_{jk} = \overline{x_{kj}}$, $1 \leq k \leq m$; $1 \leq j \leq n$.

- ▶ Then $\langle \cdot, \cdot \rangle$ is an inner product on $M_{m,n}(\mathbb{C})$.

Another example

- ▶ Take $\mathbb{F} = \mathbb{C}$. Fix $m, n \in \mathbb{N}$. Then we know that $M_{m,n}(\mathbb{C})$ is a vector space of dimension mn over $\mathbb{C}$.
- ▶ For $X, Y \in M_{m,n}(\mathbb{C})$ take

$$\langle X, Y \rangle = \text{trace}(X^* Y)$$

where $(X^*)_{jk} = \overline{x_{kj}}$, $1 \leq k \leq m; 1 \leq j \leq n$.

- ▶ Then $\langle \cdot, \cdot \rangle$ is an inner product on $M_{m,n}(\mathbb{C})$.
- ▶ **Proof:** We have,

$$\begin{aligned}\langle X, Y \rangle &= \text{trace}(X^* Y) \\ &= \sum_{j=1}^n (X^* Y)_{jj} \\ &= \sum_{j=1}^n \sum_{k=1}^m (X^*)_{jk} (Y)_{kj} \\ &= \sum_{j=1}^n \sum_{k=1}^m \overline{x_{kj}} y_{kj}\end{aligned}$$

- ▶ Now it is clear that this is essentially the standard inner product on $\mathbb{C}^{mn}$.

Continuation

- ▶ Now it is clear that this is essentially the standard inner product on $\mathbb{C}^{mn}$.
- ▶ Similarly,

$$\langle X, Y \rangle = \text{trace}(X^t Y),$$

is an inner product on $M_{m,n}(\mathbb{R})$.

$\mathbb{R}^2$.

► On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\mathbb{R}^2$.

- On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

- More generally the distance between any two points $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{\frac{1}{2}} = (\langle x - y, x - y \rangle)^{\frac{1}{2}}.$$

$\mathbb{R}^2$.

- ▶ On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

- ▶ More generally the distance between any two points $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{\frac{1}{2}} = (\langle x - y, x - y \rangle)^{\frac{1}{2}}.$$

- ▶ This suggests the following definitions.

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

- If x, y are vectors in V , the distance of y from x is defined as

$$d(x, y) = \|y - x\|.$$

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

- If x, y are vectors in V , the distance of y from x is defined as

$$d(x, y) = \|y - x\|.$$

- The 'distance function' $d : V \times V \rightarrow \mathbb{R}$ is also known as **metric**.

Basic properties of the norm

- **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.
- ▶ (ii) $\|ax\| = |a|\|x\|$, $\forall a \in \mathbb{F}$ and $x \in V$.

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.
- ▶ (ii) $\|ax\| = |a|\|x\|$, $\forall a \in \mathbb{F}$ and $x \in V$.
- ▶ (iii) $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in V$.

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Continuation

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- ▶ (ii) We have,

$$\langle ax, ax \rangle = \bar{a}.a\langle x, x \rangle = |a|^2\langle x, x \rangle$$

from anti-linearity of the inner product in the first variable and linear in the second variable. Now (ii) is immediate.

Continuation

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- ▶ (ii) We have,

$$\langle ax, ax \rangle = \bar{a}.a\langle x, x \rangle = |a|^2\langle x, x \rangle$$

from anti-linearity of the inner product in the first variable and linear in the second variable. Now (ii) is immediate.

- ▶ (iii) This is a consequence of Cauchy-Schwarz inequality and we will prove it in the next class.

Continuation

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- ▶ (ii) We have,

$$\langle ax, ax \rangle = \bar{a}.a\langle x, x \rangle = |a|^2\langle x, x \rangle$$

from anti-linearity of the inner product in the first variable and linear in the second variable. Now (ii) is immediate.

- ▶ (iii) This is a consequence of Cauchy-Schwarz inequality and we will prove it in the next class.
- ▶ **END OF LECTURE 7.**

Lecture 8: Cauchy-Schwarz inequality

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.

Lecture 8: Cauchy-Schwarz inequality

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).

Lecture 8: Cauchy-Schwarz inequality

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:

Lecture 8: Cauchy-Schwarz inequality

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:
- ▶ **Definition 7.1:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{R}^n$. The **standard inner product** on $\mathbb{R}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n x_j y_j.$$

Lecture 8: Cauchy-Schwarz inequality

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product between two vectors x, y in V , usually denoted by $\langle x, y \rangle$ would be a scalar (an element of $\mathbb{F}$).
- ▶ The main examples we have in mind are the following:
- ▶ **Definition 7.1:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{R}^n$. The **standard inner product** on $\mathbb{R}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n x_j y_j.$$

- ▶ We note that $\langle x, x \rangle \geq 0$ for every x in $\mathbb{R}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

- **Definition 7.2:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{C}^n$. The **standard inner product** on $\mathbb{C}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n \overline{x_j} y_j.$$

- **Definition 7.2:** For $n \in \mathbb{N}$, consider the vector space $\mathbb{C}^n$. The **standard inner product** on $\mathbb{C}^n$ is defined by:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n \overline{x_j} y_j.$$

- Here also we note that $\langle x, x \rangle \geq 0$ for every x in $\mathbb{C}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)
- ▶ (iv) $\langle x, x \rangle = 0$ if and only if $x = 0$. (Definiteness.)

Definitions

- ▶ **Definition 7.3:** Let V be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ An inner product on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$$

such that

- ▶ (i) $\langle x, cy + dz \rangle = c\langle x, y \rangle + d\langle x, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in second variable.)
- ▶ (ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Conjugate Symmetry.)
- ▶ (iii) $\langle x, x \rangle \geq 0$ for all x in V . (Positivity.)
- ▶ (iv) $\langle x, x \rangle = 0$ if and only if $x = 0$. (Definiteness.)
- ▶ Some authors take inner product as linear in the first variable. It is a matter of convention. A vector space with a specified inner product is called an inner product space.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and symmetry of the inner-product we get also linearity in the first variable.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable.)

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable.)
- ▶ On the other hand if the field $\mathbb{F} = \mathbb{C}$, from linearity and symmetry of the inner-product we get anti-linearity in the first variable.

Linearity or anti-linearity in the first variable

- ▶ If the field $\mathbb{F} = \mathbb{R}$, from linearity and symmetry of the inner-product we get also linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = c\langle x, z \rangle + d\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in first variable.)
- ▶ On the other hand if the field $\mathbb{F} = \mathbb{C}$, from linearity and symmetry of the inner-product we get anti-linearity in the first variable.
- ▶ That is, $\langle cx + dy, z \rangle = \bar{c}\langle x, z \rangle + \bar{d}\langle y, z \rangle$, for all $x, y, z \in V, c, d \in \mathbb{F}$ (Linearity in the second variable.)

$\mathbb{R}^2$.

► On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\mathbb{R}^2$.

- On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

- More generally the distance between any two points $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$
and $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{\frac{1}{2}} = (\langle (x - y), (x - y) \rangle)^{\frac{1}{2}}.$$

$\mathbb{R}^2$.

- ▶ On $\mathbb{R}^2$, for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

$$(\langle x, x \rangle)^{\frac{1}{2}} = (x_1^2 + x_2^2)^{\frac{1}{2}}$$

is the distance of point x from the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

- ▶ More generally the distance between any two points $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{\frac{1}{2}} = (\langle (x - y), (x - y) \rangle)^{\frac{1}{2}}.$$

- ▶ This suggests the following definitions.

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

- If x, y are vectors in V , the distance of y from x is defined as

$$d(x, y) = \|y - x\|.$$

The norm on an inner product space

- **Definition 7.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then the **norm** of a vector $x \in V$ is defined as

$$\|x\| = (\langle x, x \rangle)^{\frac{1}{2}}.$$

- If x, y are vectors in V , the distance of y from x is defined as

$$d(x, y) = \|y - x\|.$$

- The 'distance function' $d : V \rightarrow V \rightarrow \mathbb{R}$ is also known as **metric**.

Basic properties of the norm

- **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.
- ▶ (ii) $\|ax\| = |a|\|x\|$, $\forall a \in \mathbb{F}$ and $x \in V$.

Basic properties of the norm

- ▶ **Theorem 7.6:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\| \cdot \|$ be the associated norm on V . Then
- ▶ (i) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.
- ▶ (ii) $\|ax\| = |a|\|x\|$, $\forall a \in \mathbb{F}$ and $x \in V$.
- ▶ (iii) $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in V$.

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Continuation

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- ▶ (ii) We have,

$$\langle ax, ax \rangle = \bar{a}.a\langle x, x \rangle = |a|^2\langle x, x \rangle$$

from anti-linearity of the inner product in the first variable and linear in the second variable. Now (ii) is immediate.

Continuation

- ▶ **Proof.** (i) This is clear, as $\langle x, x \rangle \geq 0$ for all x and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- ▶ (ii) We have,

$$\langle ax, ax \rangle = \bar{a}.a\langle x, x \rangle = |a|^2\langle x, x \rangle$$

from anti-linearity of the inner product in the first variable and linear in the second variable. Now (ii) is immediate.

- ▶ (iii) This is a consequence of Cauchy-Schwarz inequality and will be proved in the next class.

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

- ▶ **Theorem 8.1 (Cauchy-Schwarz inequality):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Then for x, y in V ,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

- ▶ **Theorem 8.1 (Cauchy-Schwarz inequality):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Then for x, y in V ,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

- ▶ **Theorem 8.1 (Cauchy-Schwarz inequality):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Then for x, y in V ,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

- ▶ The equality takes place if and only if x, y are linearly dependent.

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

- ▶ **Theorem 8.1 (Cauchy-Schwarz inequality):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Then for x, y in V ,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

- ▶ The equality takes place if and only if x, y are linearly dependent.
- ▶ Case (i): $\langle x, x \rangle = 0$. This means $x = 0$. Now $\langle 0, y \rangle = \langle 0 + 0, y \rangle = \langle 0, y \rangle + \langle 0, y \rangle$. This implies $\langle x, y \rangle = 0$.

Cauchy-Schwarz inequality

- ▶ You may have seen the inequality

$$\sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n b_j^2 \right)^{\frac{1}{2}}$$

for $n \in \mathbb{N}$ and real numbers $a_j, b_j, 1 \leq j \leq n$. Here we have a generalization of this result. Cauchy-Schwarz inequality is among the most important inequalities of mathematics.

- ▶ **Theorem 8.1 (Cauchy-Schwarz inequality):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$ (where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Then for x, y in V ,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

- ▶ The equality takes place if and only if x, y are linearly dependent.
- ▶ Case (i): $\langle x, x \rangle = 0$. This means $x = 0$. Now $\langle 0, y \rangle = \langle 0 + 0, y \rangle = \langle 0, y \rangle + \langle 0, y \rangle$. This implies $\langle x, y \rangle = 0$.
- ▶ Hence the required inequality is trivially true and is an equality

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.
- ▶ We know $\langle z, z \rangle \geq 0$.

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.
- ▶ We know $\langle z, z \rangle \geq 0$.
- ▶ Recalling that $\|x\|^2 = \langle x, x \rangle$ and using anti-linearity of the inner-product in the first variable and linearity in the second variable we get:

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.
- ▶ We know $\langle z, z \rangle \geq 0$.
- ▶ Recalling that $\|x\|^2 = \langle x, x \rangle$ and using anti-linearity of the inner-product in the first variable and linearity in the second variable we get:
- ▶

$$\begin{aligned} 0 &\leq \langle y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x, y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x \rangle \\ &= \|y\|^2 - 2 \frac{|\langle x, y \rangle|^2}{\|x\|^2} + \frac{|\langle x, y \rangle|^2 \cdot \|x\|^2}{\|x\|^4} \\ &= \|y\|^2 - \frac{|\langle x, y \rangle|^2}{\|x\|^2}, \end{aligned}$$

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.
- ▶ We know $\langle z, z \rangle \geq 0$.
- ▶ Recalling that $\|x\|^2 = \langle x, x \rangle$ and using anti-linearity of the inner-product in the first variable and linearity in the second variable we get:
- ▶

$$\begin{aligned} 0 &\leq \langle y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x, y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x \rangle \\ &= \|y\|^2 - 2 \frac{|\langle x, y \rangle|^2}{\|x\|^2} + \frac{|\langle x, y \rangle|^2 \cdot \|x\|^2}{\|x\|^4} \\ &= \|y\|^2 - \frac{|\langle x, y \rangle|^2}{\|x\|^2}, \end{aligned}$$

- ▶ which is same as $|\langle x, y \rangle|^2 \leq \|x\|^2 \|y\|^2$.

Continuation

- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ Consider the vector $z := y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x$.
- ▶ We know $\langle z, z \rangle \geq 0$.
- ▶ Recalling that $\|x\|^2 = \langle x, x \rangle$ and using anti-linearity of the inner-product in the first variable and linearity in the second variable we get:



$$\begin{aligned} 0 &\leq \langle y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x, y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x \rangle \\ &= \|y\|^2 - 2 \frac{|\langle x, y \rangle|^2}{\|x\|^2} + \frac{|\langle x, y \rangle|^2 \cdot \|x\|^2}{\|x\|^4} \\ &= \|y\|^2 - \frac{|\langle x, y \rangle|^2}{\|x\|^2}, \end{aligned}$$

- ▶ which is same as $|\langle x, y \rangle|^2 \leq \|x\|^2 \|y\|^2$.
- ▶ Taking positive square root, we have the required inequality.

The case of equality

- ▶ Case (i): $\langle x, x \rangle = 0$. Here $x = 0$ and hence x, y are linearly dependent.

The case of equality

- ▶ Case (i): $\langle x, x \rangle = 0$. Here $x = 0$ and hence x, y are linearly dependent.
- ▶ Case (ii): $\langle x, x \rangle \neq 0$.

The case of equality

- ▶ Case (i): $\langle x, x \rangle = 0$. Here $x = 0$ and hence x, y are linearly dependent.
- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ From the proof of Cauchy-Schwarz inequality, to have equality, we must have $z = 0$, where

$$z = y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x.$$

The case of equality

- ▶ Case (i): $\langle x, x \rangle = 0$. Here $x = 0$ and hence x, y are linearly dependent.
- ▶ Case (ii): $\langle x, x \rangle \neq 0$.
- ▶ From the proof of Cauchy-Schwarz inequality, to have equality, we must have $z = 0$, where

$$z = y - \frac{\langle x, y \rangle}{\langle x, x \rangle} x.$$

- ▶ Hence z is a scalar multiple of x . In particular, x, y are linearly dependent.
- ▶ **Exercise:** We proved the Cauchy-Schwarz inequality with the assumption of definiteness, that is $\langle x, x \rangle = 0$ implies $x = 0$. Prove the inequality without this assumption. (Hint: Consider the function $p(t) = \|y - t\langle x, y \rangle x\|^2$ for $t \in \mathbb{R}$.)

Triangle inequality

- ▶ Recall that we wanted to show

$$\|x + y\| \leq \|x\| + \|y\|$$

for $x, y \in V$.

Triangle inequality

- ▶ Recall that we wanted to show

$$\|x + y\| \leq \|x\| + \|y\|$$

for $x, y \in V$.

- ▶ Now using Cauchy-Schwarz inequality,

$$\begin{aligned}\langle x + y, x + y \rangle &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle \\&= \|x\|^2 + \|y\|^2 + 2 \operatorname{Re} \langle x, y \rangle \\&\leq \|x\|^2 + \|y\|^2 + 2|\langle x, y \rangle| \\&\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \\&= (\|x\| + \|y\|)^2.\end{aligned}$$

Triangle inequality

- ▶ Recall that we wanted to show

$$\|x + y\| \leq \|x\| + \|y\|$$

for $x, y \in V$.

- ▶ Now using Cauchy-Schwarz inequality,

$$\begin{aligned}\langle x + y, x + y \rangle &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle \\ &= \|x\|^2 + \|y\|^2 + 2 \operatorname{Re} \langle x, y \rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2|\langle x, y \rangle| \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \\ &= (\|x\| + \|y\|)^2.\end{aligned}$$

- ▶ Taking positive square root we have the required inequality. ■.

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

- ▶ We observe that d satisfies:

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

- ▶ We observe that d satisfies:
- ▶ (i) $d(x, y) \geq 0$ for every x, y in V and $d(x, y) = 0$ if and only if $x = y$ (Positivity and definiteness).

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

- ▶ We observe that d satisfies:
- ▶ (i) $d(x, y) \geq 0$ for every x, y in V and $d(x, y) = 0$ if and only if $x = y$ (Positivity and definiteness).
- ▶ (ii) $d(x, y) = d(y, x)$ for all x, y in V (symmetry).

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

- ▶ We observe that d satisfies:
- ▶ (i) $d(x, y) \geq 0$ for every x, y in V and $d(x, y) = 0$ if and only if $x = y$ (Positivity and definiteness).
- ▶ (ii) $d(x, y) = d(y, x)$ for all x, y in V (symmetry).
- ▶ (iii) $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y, z in V . (triangle inequality).

Triangle inequality for the metric

- ▶ In any inner product space $(V, \langle \cdot, \cdot \rangle)$ we have the **metric** (distance function) as

$$d(x, y) = \|y - x\|.$$

- ▶ We observe that d satisfies:
- ▶ (i) $d(x, y) \geq 0$ for every x, y in V and $d(x, y) = 0$ if and only if $x = y$ (Positivity and definiteness).
- ▶ (ii) $d(x, y) = d(y, x)$ for all x, y in V (symmetry).
- ▶ (iii) $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y, z in V . (triangle inequality).
- ▶ All these are immediate from respective properties of the norm.

Parallelogram law

- **Theorem 8.2:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for any x, y in V :

Parallelogram law

- ▶ **Theorem 8.2:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for any x, y in V :



$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Parallelogram law

- ▶ **Theorem 8.2:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for any x, y in V :



$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

- ▶ **Proof.** This is clear by direct computation:

$$\begin{aligned} & \|x + y\|^2 + \|x - y\|^2 \\ &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= 2\langle x, x \rangle + 2\langle y, y \rangle \end{aligned}$$

Parallelogram law

- ▶ **Theorem 8.2:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for any x, y in V :



$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

- ▶ **Proof.** This is clear by direct computation:

$$\begin{aligned} & \|x + y\|^2 + \|x - y\|^2 \\ &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= 2\langle x, x \rangle + 2\langle y, y \rangle \end{aligned}$$

- ▶ as all cross-terms cancel with each other.

Parallelogram law

- ▶ **Theorem 8.2:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for any x, y in V :



$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

- ▶ **Proof.** This is clear by direct computation:

$$\begin{aligned} & \|x + y\|^2 + \|x - y\|^2 \\ &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= 2\langle x, x \rangle + 2\langle y, y \rangle \end{aligned}$$

- ▶ as all cross-terms cancel with each other.
- ▶ **Exercise:** Find out as to why this is called parallelogram law.

Recovering the inner product from the norm

- ▶ It is possible to compute the inner product from the norm using the following formulae.

Recovering the inner product from the norm

- ▶ It is possible to compute the inner product from the norm using the following formulae.
- ▶ **Theorem 8.3:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{R}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

Recovering the inner product from the norm

- ▶ It is possible to compute the inner product from the norm using the following formulae.
- ▶ **Theorem 8.3:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{R}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

- ▶ **Proof.** We have,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle$$

By symmetry,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle.$$

Recovering the inner product from the norm

- ▶ It is possible to compute the inner product from the norm using the following formulae.
- ▶ **Theorem 8.3:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{R}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

- ▶ **Proof.** We have,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle$$

By symmetry,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle.$$

- ▶ Similarly,

$$\|(x - y)\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle.$$

Recovering the inner product from the norm

- ▶ It is possible to compute the inner product from the norm using the following formulae.
- ▶ **Theorem 8.3:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{R}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

- ▶ **Proof.** We have,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle$$

By symmetry,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle.$$

- ▶ Similarly,

$$\|(x - y)\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle.$$

- ▶ Now the result is clear by taking the difference of two equations and dividing by four.

Continuation

- **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

Continuation

- **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

Continuation

- **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.
- ▶ Recall: $i^2 = -1, i^3 = -i, i^4 = 1$ and $\bar{i} = -i$.

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.
- ▶ Recall: $i^2 = -1, i^3 = -i, i^4 = 1$ and $\bar{i} = -i$.
- ▶ Now

$$\begin{aligned} (-i)^0 \|x + i^0 y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle. \\ (-i)^1 \|x + i^1 y\|^2 &= (-i)\|x\|^2 + (-i)\|y\|^2 + \langle x, y \rangle + (-i)^2 \langle y, x \rangle \\ (-i)^2 \|x + i^2 y\|^2 &= -\|x\|^2 - \|y\|^2 + \langle x, y \rangle + (-1) \cdot (-1) \langle y, x \rangle \\ (-i)^3 \|x + i^3 y\|^2 &= i\|x\|^2 + i\|y\|^2 + \langle x, y \rangle + (i) \cdot (i) \langle y, x \rangle \end{aligned}$$

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.
- ▶ Recall: $i^2 = -1, i^3 = -i, i^4 = 1$ and $\bar{i} = -i$.
- ▶ Now

$$\begin{aligned} (-i)^0 \|x + i^0 y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle. \\ (-i)^1 \|x + i^1 y\|^2 &= (-i)\|x\|^2 + (-i)\|y\|^2 + \langle x, y \rangle + (-i)^2 \langle y, x \rangle \\ (-i)^2 \|x + i^2 y\|^2 &= -\|x\|^2 - \|y\|^2 + \langle x, y \rangle + (-1) \cdot (-1) \langle y, x \rangle \\ (-i)^3 \|x + i^3 y\|^2 &= i\|x\|^2 + i\|y\|^2 + \langle x, y \rangle + (i) \cdot (i) \langle y, x \rangle \end{aligned}$$

- ▶ Adding these equations and dividing by 4 we have the theorem. ■.

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.
- ▶ Recall: $i^2 = -1, i^3 = -i, i^4 = 1$ and $\bar{i} = -i$.
- ▶ Now

$$\begin{aligned} (-i)^0 \|x + i^0 y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle. \\ (-i)^1 \|x + i^1 y\|^2 &= (-i)\|x\|^2 + (-i)\|y\|^2 + \langle x, y \rangle + (-i)^2 \langle y, x \rangle \\ (-i)^2 \|x + i^2 y\|^2 &= -\|x\|^2 - \|y\|^2 + \langle x, y \rangle + (-1) \cdot (-1) \langle y, x \rangle \\ (-i)^3 \|x + i^3 y\|^2 &= i\|x\|^2 + i\|y\|^2 + \langle x, y \rangle + (i) \cdot (i) \langle y, x \rangle \end{aligned}$$

- ▶ Adding these equations and dividing by 4 we have the theorem. ■.
- ▶ The formula of this theorem is known as **polarization identity**.

Continuation

- ▶ **Theorem 8.4 (polarization identity):** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{C}$. Then for any $x, y \in V$

$$\langle x, y \rangle = \frac{1}{4} \sum_{n=0}^3 (-i)^n \|x + i^n y\|^2.$$

- ▶ **Proof.** The proof is by direct computation.
- ▶ Recall: $i^2 = -1, i^3 = -i, i^4 = 1$ and $\bar{i} = -i$.
- ▶ Now

$$\begin{aligned} (-i)^0 \|x + i^0 y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle. \\ (-i)^1 \|x + i^1 y\|^2 &= (-i)\|x\|^2 + (-i)\|y\|^2 + \langle x, y \rangle + (-i)^2 \langle y, x \rangle \\ (-i)^2 \|x + i^2 y\|^2 &= -\|x\|^2 - \|y\|^2 + \langle x, y \rangle + (-1) \cdot (-1) \langle y, x \rangle \\ (-i)^3 \|x + i^3 y\|^2 &= i\|x\|^2 + i\|y\|^2 + \langle x, y \rangle + (i) \cdot (i) \langle y, x \rangle \end{aligned}$$

- ▶ Adding these equations and dividing by 4 we have the theorem. ■.
- ▶ The formula of this theorem is known as **polarization identity**.
- ▶ **END OF LECTURE 8.**

Lecture 9: Gram-Schmidt orthogonalization

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.

Lecture 9: Gram-Schmidt orthogonalization

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ Let V be a vector space.

Lecture 9: Gram-Schmidt orthogonalization

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ Let V be a vector space.
- ▶ A collection of vectors $\{v_1, v_2, \dots, v_n\}$ is said to be linearly independent if

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0$$

implies $c_1 = c_2 = \dots = c_n = 0$.

Lecture 9: Gram-Schmidt orthogonalization

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ Let V be a vector space.
- ▶ A collection of vectors $\{v_1, v_2, \dots, v_n\}$ is said to be linearly independent if

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0$$

implies $c_1 = c_2 = \dots = c_n = 0$.

- ▶ A linearly independent collection of vectors $\{v_1, v_2, \dots, v_n\}$ is a basis for V if

$$V = \text{span}\{v_1, v_2, \dots, v_n\},$$

Lecture 9: Gram-Schmidt orthogonalization

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ Let V be a vector space.
- ▶ A collection of vectors $\{v_1, v_2, \dots, v_n\}$ is said to be linearly independent if

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0$$

implies $c_1 = c_2 = \dots = c_n = 0$.

- ▶ A linearly independent collection of vectors $\{v_1, v_2, \dots, v_n\}$ is a basis for V if

$$V = \text{span}\{v_1, v_2, \dots, v_n\},$$

- ▶ that is, given any vector $x \in V$, there exist, $c_1, c_2, \dots, c_n$ in $\mathbb{F}$ such that $x = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$. Note that given x , these coefficients are uniquely determined due to linear independence of v_j 's.

Orthogonality

- **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.

Orthogonality

- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.
- ▶ More generally, two subsets S, T of V are said to be **mutually orthogonal** if

$$\langle u, v \rangle = 0, \quad \forall u \in S, v \in T.$$

Orthogonality

- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.
- ▶ More generally, two subsets S, T of V are said to be **mutually orthogonal** if

$$\langle u, v \rangle = 0, \quad \forall u \in S, v \in T.$$

- ▶ **Theorem 9.2 (The Pythagorean theorem for inner product spaces:)** Suppose v, w are mutually orthogonal vectors in an inner product space V . Then

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Orthogonality

- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.
- ▶ More generally, two subsets S, T of V are said to be **mutually orthogonal** if

$$\langle u, v \rangle = 0, \quad \forall u \in S, v \in T.$$

- ▶ **Theorem 9.2 (The Pythagorean theorem for inner product spaces:)** Suppose v, w are mutually orthogonal vectors in an inner product space V . Then

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

- ▶ **Proof.** We have

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w \rangle \\ &= \langle v, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + \|w\|^2,\end{aligned}$$

Orthogonality

- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.
- ▶ More generally, two subsets S, T of V are said to be **mutually orthogonal** if

$$\langle u, v \rangle = 0, \quad \forall u \in S, v \in T.$$

- ▶ **Theorem 9.2 (The Pythagorean theorem for inner product spaces:)** Suppose v, w are mutually orthogonal vectors in an inner product space V . Then

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

- ▶ **Proof.** We have

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w \rangle \\ &= \langle v, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + \|w\|^2,\end{aligned}$$

- ▶ as cross-terms are equal to zero due to orthogonality.

Basic properties

- **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.
- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.
- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_m\}$ is linearly independent.

- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_m v_m \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■
- ▶ **Corollary 9.4:** Suppose $\{v_1, \dots, v_m\}$ is a set of mutually orthogonal non-zero vectors in an inner product space V , then

$$m \leq \dim V.$$

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.

- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■
- ▶ **Corollary 9.4:** Suppose $\{v_1, \dots, v_m\}$ is a set of mutually orthogonal non-zero vectors in an inner product space V , then

$$m \leq \dim V.$$

- ▶ **Proof.** This is clear, as the dimension of V is same as the maximum possible size of linearly independent sets. ■

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.
- ▶ **Example 9.6:** For $\mathbb{R}^n$ (or $\mathbb{C}^n$) the standard basis $\{e_1, e_2, \dots, e_n\}$, where e_j is the vector whose j -th coordinate is one and all other coordinates are equal to zero, is an orthonormal basis with respect to the standard inner product.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.
- ▶ **Theorem 9.7:** Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then for any vector $w \in V$,

$$w = \sum_{j=1}^n \langle v_j, w \rangle v_j.$$

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.
- ▶ **Theorem 9.7:** Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then for any vector $w \in V$,

$$w = \sum_{j=1}^n \langle v_j, w \rangle v_j.$$

- ▶ **Proof.** As $\{v_1, v_2, \dots, v_n\}$ is a basis for V , $w = \sum_{i=1}^n c_i v_i$ for some $c_1, c_2, \dots, c_n$ in $\mathbb{F}$.

Continuation

- ▶ Now for any j , using linearity of the inner product in second variable,

$$\begin{aligned}\langle v_j, w \rangle &= \langle v_j, \sum_{i=1}^n c_i v_i \rangle \\ &= \sum_{i=1}^n c_i \langle v_j, v_i \rangle \\ &= \sum_{i=1}^n c_i \delta_{ji} \\ &= c_j.\end{aligned}$$

Continuation

- ▶ Now for any j , using linearity of the inner product in second variable,

$$\begin{aligned}\langle v_j, w \rangle &= \langle v_j, \sum_{i=1}^n c_i v_i \rangle \\ &= \sum_{i=1}^n c_i \langle v_j, v_i \rangle \\ &= \sum_{i=1}^n c_i \delta_{ji} \\ &= c_j.\end{aligned}$$

- ▶ As this is true for every j , $w = \sum_{j=1}^n \langle v_j, w \rangle v_j$. ■

Gram-Schmidt orthogonalization

- ▶ **Gram-Schmidt Orthogonalization** : This is an iterative scheme to obtain an orthonormal basis out of a given basis.

Gram-Schmidt orthogonalization

- ▶ **Gram-Schmidt Orthogonalization** : This is an iterative scheme to obtain an orthonormal basis out of a given basis.
- ▶ Let $\{u_1, u_2, \dots, u_n\}$ be a basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$.

Gram-Schmidt orthogonalization

- ▶ **Gram-Schmidt Orthogonalization** : This is an iterative scheme to obtain an orthonormal basis out of a given basis.
- ▶ Let $\{u_1, u_2, \dots, u_n\}$ be a basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$.
- ▶ Since $\{u_1, \dots, u_n\}$ are linearly independent each of them is non-zero.

Gram-Schmidt orthogonalization

- ▶ **Gram-Schmidt Orthogonalization** : This is an iterative scheme to obtain an orthonormal basis out of a given basis.
- ▶ Let $\{u_1, u_2, \dots, u_n\}$ be a basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$.
- ▶ Since $\{u_1, \dots, u_n\}$ are linearly independent each of them is non-zero.
- ▶ In particular, $\|u_1\| \neq 0$.

Gram-Schmidt orthogonalization

- ▶ **Gram-Schmidt Orthogonalization** : This is an iterative scheme to obtain an orthonormal basis out of a given basis.
- ▶ Let $\{u_1, u_2, \dots, u_n\}$ be a basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$.
- ▶ Since $\{u_1, \dots, u_n\}$ are linearly independent each of them is non-zero.
- ▶ In particular, $\|u_1\| \neq 0$.
- ▶ Take $y_1 = u_1$ and $v_1 = \frac{u_1}{\|u_1\|}$.

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.
- ▶ We claim that y_2 is orthogonal to v_1 .

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.
- ▶ We claim that y_2 is orthogonal to v_1 .
- ▶ In fact,

$$\begin{aligned}\langle v_1, y_2 \rangle &= \langle v_1, u_2 - \langle v_1, u_2 \rangle v_1 \rangle \\ &= \langle v_1, u_2 \rangle - \langle v_1, u_2 \rangle \\ &= 0\end{aligned}$$

as $\langle v_1, v_1 \rangle = 1$.

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.
- ▶ We claim that y_2 is orthogonal to v_1 .
- ▶ In fact,

$$\begin{aligned}\langle v_1, y_2 \rangle &= \langle v_1, u_2 - \langle v_1, u_2 \rangle v_1 \rangle \\ &= \langle v_1, u_2 \rangle - \langle v_1, u_2 \rangle \\ &= 0\end{aligned}$$

as $\langle v_1, v_1 \rangle = 1$.

- ▶ Also as u_1, u_2 are linearly independent, $y_2 \neq 0$.

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.
- ▶ We claim that y_2 is orthogonal to v_1 .
- ▶ In fact,

$$\begin{aligned}\langle v_1, y_2 \rangle &= \langle v_1, u_2 - \langle v_1, u_2 \rangle v_1 \rangle \\ &= \langle v_1, u_2 \rangle - \langle v_1, u_2 \rangle \\ &= 0\end{aligned}$$

as $\langle v_1, v_1 \rangle = 1$.

- ▶ Also as u_1, u_2 are linearly independent, $y_2 \neq 0$.
- ▶ Now take, $v_2 = \frac{y_2}{\|y_2\|}$.

Continuation

- ▶ Take $y_2 = u_2 - \langle v_1, u_2 \rangle v_1$.
- ▶ We claim that y_2 is orthogonal to v_1 .
- ▶ In fact,

$$\begin{aligned}\langle v_1, y_2 \rangle &= \langle v_1, u_2 - \langle v_1, u_2 \rangle v_1 \rangle \\ &= \langle v_1, u_2 \rangle - \langle v_1, u_2 \rangle \\ &= 0\end{aligned}$$

as $\langle v_1, v_1 \rangle = 1$.

- ▶ Also as u_1, u_2 are linearly independent, $y_2 \neq 0$.
- ▶ Now take, $v_2 = \frac{y_2}{\|y_2\|}$.
- ▶ This way, $\{v_1, v_2\}$ are orthonormal (that is, they have norm one and are mutually orthogonal.)

Continuation

- ▶ Inductively, after we construct $\{v_1, v_2, \dots, v_k\}$ such that they are orthonormal and

$$\text{span} \{v_1, v_2, \dots, v_k\} = \text{span} \{u_1, u_2, \dots, u_k\},$$

Continuation

- ▶ Inductively, after we construct $\{v_1, v_2, \dots, v_k\}$ such that they are orthonormal and

$$\text{span } \{v_1, v_2, \dots, v_k\} = \text{span } \{u_1, u_2, \dots, u_k\},$$

- ▶ we take

$$y_{k+1} = u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j.$$

Continuation

- ▶ Inductively, after we construct $\{v_1, v_2, \dots, v_k\}$ such that they are orthonormal and

$$\text{span} \{v_1, v_2, \dots, v_k\} = \text{span} \{u_1, u_2, \dots, u_k\},$$

- ▶ we take

$$y_{k+1} = u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j.$$

- ▶ Now, for any $1 \leq i \leq k$,

$$\langle v_i, y_{k+1} \rangle = \langle v_i, u_{k+1} \rangle - \langle v_i, u_{k+1} \rangle = 0.$$

- ▶ $y_{k+1} \neq 0$ follows as

$$u_{k+1} \notin \text{span} \{v_1, \dots, v_k\} = \text{span} \{u_1, \dots, u_k\}.$$

Continuation

- ▶ Inductively, after we construct $\{v_1, v_2, \dots, v_k\}$ such that they are orthonormal and

$$\text{span} \{v_1, v_2, \dots, v_k\} = \text{span} \{u_1, u_2, \dots, u_k\},$$

- ▶ we take

$$y_{k+1} = u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j.$$

- ▶ Now, for any $1 \leq i \leq k$,

$$\langle v_i, y_{k+1} \rangle = \langle v_i, u_{k+1} \rangle - \langle v_i, u_{k+1} \rangle = 0.$$

- ▶ $y_{k+1} \neq 0$ follows as
 $u_{k+1} \notin \text{span} \{v_1, \dots, v_k\} = \text{span} \{u_1, \dots, u_k\}.$
- ▶ Then take $v_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|}.$

Continuation

- ▶ Inductively, after we construct $\{v_1, v_2, \dots, v_k\}$ such that they are orthonormal and

$$\text{span} \{v_1, v_2, \dots, v_k\} = \text{span} \{u_1, u_2, \dots, u_k\},$$

- ▶ we take

$$y_{k+1} = u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j.$$

- ▶ Now, for any $1 \leq i \leq k$,

$$\langle v_i, y_{k+1} \rangle = \langle v_i, u_{k+1} \rangle - \langle v_i, u_{k+1} \rangle = 0.$$

- ▶ $y_{k+1} \neq 0$ follows as

$$u_{k+1} \notin \text{span} \{v_1, \dots, v_k\} = \text{span} \{u_1, \dots, u_k\}.$$

- ▶ Then take $v_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|}$.

- ▶ We see that $\{v_1, \dots, v_{k+1}\}$ are orthonormal and $\text{span} \{v_1, \dots, v_{k+1}\} = \text{span} \{u_1, \dots, u_{k+1}\}$ so that the induction can be continued.

Continuation

- ▶ Continuing this way, up to $k = n$, we have an ortho-normal collection $\{v_1, v_2, \dots, v_n\}$. Since it spans whole of V it is an orthonormal basis. ■

Continuation

- ▶ Continuing this way, up to $k = n$, we have an ortho-normal collection $\{v_1, v_2, \dots, v_n\}$. Since it spans whole of V it is an orthonormal basis. ■
- ▶ **Remark:** The formula of Gram-Schmidt orthogonalization:

$$v_{k+1} = \frac{u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j}{\|u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j\|}$$

is worth remembering.

Continuation

- ▶ Continuing this way, up to $k = n$, we have an orthonormal collection $\{v_1, v_2, \dots, v_n\}$. Since it spans whole of V it is an orthonormal basis. ■
- ▶ **Remark:** The formula of Gram-Schmidt orthogonalization:

$$v_{k+1} = \frac{u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j}{\|u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j\|}$$

is worth remembering.

- ▶ **Exercise:** Obtain an orthonormal basis for $\mathbb{R}^3$ by Gram-Schmidt orthogonalization applied to the basis:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Continuation

- ▶ Continuing this way, up to $k = n$, we have an ortho-normal collection $\{v_1, v_2, \dots, v_n\}$. Since it spans whole of V it is an orthonormal basis. ■
- ▶ **Remark:** The formula of Gram-Schmidt orthogonalization:

$$v_{k+1} = \frac{u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j}{\|u_{k+1} - \sum_{j=1}^k \langle v_j, u_{k+1} \rangle v_j\|}$$

is worth remembering.

- ▶ **Exercise:** Obtain an orthonormal basis for $\mathbb{R}^3$ by Gram-Schmidt orthogonalization applied to the basis:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

- ▶ **END OF LECTURE 9.**

Lecture 10: Adjoint of a linear map

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.

Lecture 10: Adjoint of a linear map

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.

Lecture 10: Adjoint of a linear map

- ▶ Recall: In the following the field $\mathbb{F}$ would be either $\mathbb{R}$ or $\mathbb{C}$.
- ▶ **Definition 9.1:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Two vectors u, v in V are said to be **mutually orthogonal** if $\langle u, v \rangle = 0$.
- ▶ More generally, two subsets S, T of V are said to be **mutually orthogonal** if

$$\langle u, v \rangle = 0, \quad \forall u \in S, v \in T.$$

Basic properties

- **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.
- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.
- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.
- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_m\}$ is linearly independent.

- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_m v_m \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■
- ▶ **Corollary 9.4:** Suppose $\{v_1, \dots, v_m\}$ is a set of mutually orthogonal non-zero vectors in an inner product space V , then

$$m \leq \dim V.$$

Basic properties

- ▶ **Proposition 9.3:** Suppose $\{v_1, v_2, \dots, v_m\}$ is an orthogonal collection of non-zero vectors in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then the collection $\{v_1, v_2, \dots, v_n\}$ is linearly independent.

- ▶ **Proof.** Suppose $c_1 v_1 + \dots + c_m v_m = 0$.

- ▶ For any j , $1 \leq j \leq m$, taking inner product with v_j , as $\langle v_j, v_i \rangle = \delta_{ij} \langle v_j, v_j \rangle$, we get

$$0 = \langle v_j, c_1 v_1 + \dots + c_n v_n \rangle = c_j \langle v_j, v_j \rangle.$$

- ▶ Therefore $c_j = 0$, $\forall j$, as $\langle v_j, v_j \rangle \neq 0$. ■
- ▶ **Corollary 9.4:** Suppose $\{v_1, \dots, v_m\}$ is a set of mutually orthogonal non-zero vectors in an inner product space V , then

$$m \leq \dim V.$$

- ▶ **Proof.** This is clear, as the dimension of V is same as the maximum possible size of linearly independent sets. ■

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.

Orthonormal basis

- ▶ A vector v in V is said to be a unit vector if $\|v\| = 1$. Note that if $y \in V$ is non-zero then $v := \frac{y}{\|y\|}$ is a unit vector.
- ▶ If $\{v_1, \dots, v_n\}$ is a collection of mutually orthogonal non-zero vectors, then $\{\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|}\}$ is a collection of mutually orthogonal unit vectors.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.
- ▶ **Example 9.6:** For $\mathbb{R}^n$ (or $\mathbb{C}^n$) the standard basis $\{e_1, e_2, \dots, e_n\}$, where e_j is the vector whose j -th coordinate is one and all other coordinates are equal to zero, is an orthonormal basis with respect to the standard inner product.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.
- ▶ **Theorem 9.7:** Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then for any vector $w \in V$,

$$w = \sum_{j=1}^n \langle v_j, w \rangle v_j.$$

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$
- ▶ Fix any $j, 1 \leq j \leq n$ and consider the basis vector v_j .

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$
- ▶ Fix any $j, 1 \leq j \leq n$ and consider the basis vector v_j .
- ▶ Now Tv_j is a vector in W and $\mathcal{C}$ is a basis for W .

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

- This defines the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ and is denoted as $c[T]_B$. Observe that if $x = \sum_{j=1}^n x_j v_j$ then by linearity

$$\begin{aligned} Tx &= \sum_{j=1}^n x_j (Tv_j) \\ &= \sum_{j=1}^n \sum_{i=1}^m x_j (a_{ij} w_i) \\ &= \sum_{i=1}^m \left[\sum_{j=1}^n a_{ij} x_j \right] w_i. \end{aligned}$$

- **Conclusion:** For a linear map $T : V \rightarrow W$, the matrix of T in bases $\mathcal{B}, \mathcal{C}$ is the unique matrix A which satisfies

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n a_{ij} x_j \right] w_i.$$

for $x = \sum_{j=1}^n x_j v_j$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.
- ▶ For general $x \in V$, we get

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n \langle w_i, Tv_j \rangle \langle v_j, x \rangle \right] w_i$$

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.
- ▶ For general $x \in V$, we get

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n \langle w_i, Tv_j \rangle \langle v_j, x \rangle \right] w_i$$

- ▶ We summarize this as a theorem.

The matrix of a linear transformation under orthonormal bases

- **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

The matrix of a linear transformation under orthonormal bases

- **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

- Conversely, given any $m \times n$ matrix $A = [a_{ij}]$, there exists unique linear map $T : V \rightarrow W$ satisfying

$$a_{ij} = \langle w_i, Tv_j \rangle, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

The matrix of a linear transformation under orthonormal bases

- ▶ **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

- ▶ Conversely, given any $m \times n$ matrix $A = [a_{ij}]$, there exists unique linear map $T : V \rightarrow W$ satisfying

$$a_{ij} = \langle w_i, Tv_j \rangle, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ Note that here:

$$Tv_j = \sum_{i=1}^m \langle w_i, Tv_j \rangle w_i = \sum_{i=1}^m a_{ij} w_i.$$

(Hermitian) adjoint

- **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

(Hermitian) adjoint

- **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.
- ▶ Consider the $n \times m$ matrix A^* defined by

$$(A^*)_{ji} = \overline{a_{ij}}, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.
- ▶ Consider the $n \times m$ matrix A^* defined by

$$(A^*)_{ji} = \overline{a_{ij}}, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ We know that A^* determines a linear map $S : W \rightarrow V$ satisfying

$$\langle v_j, Sw_i \rangle = (A^*)_{ji} = \overline{a_{ij}}.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or
$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.
- ▶ The uniqueness is clear, as we can see that any linear map S with required property has the matrix A^* as the matrix in the given bases. ■

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.
- ▶ The uniqueness is clear, as we can see that any linear map S with required property has the matrix A^* as the matrix in the given bases. ■
- ▶ **Definition 10.3:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then the unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad x \in V, y \in W,$$

is known as the (Hermitian) adjoint of T and is denoted by T^* .

Basic properties of the adjoint

- **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
- ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$.
(Anti-linearity).

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
 - ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$. (Anti-linearity).
 - ▶ (ii) $((T_1)^*)^* = T_1$. (Involution property).

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
 - ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$. (Anti-linearity).
 - ▶ (ii) $((T_1)^*)^* = T_1$. (Involution property).
- ▶ **Proof.** Exercise.

Composition

- **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

Composition

- **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

- **Proof.** For $x \in U$ and $z \in W$,

$$\langle S^*T^*z, x \rangle = \langle T^*z, Sx \rangle = \langle z, TSx \rangle.$$

Composition

- ▶ **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

- ▶ **Proof.** For $x \in U$ and $z \in W$,

$$\langle S^*T^*z, x \rangle = \langle T^*z, Sx \rangle = \langle z, TSx \rangle.$$

- ▶ Now from the uniqueness of the adjoint, we get $(TS)^* = S^*T^*$.

Composition

- ▶ **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

- ▶ **Proof.** For $x \in U$ and $z \in W$,

$$\langle S^*T^*z, x \rangle = \langle T^*z, Sx \rangle = \langle z, TSx \rangle.$$

- ▶ Now from the uniqueness of the adjoint, we get $(TS)^* = S^*T^*$.
- ▶ **END OF LECTURE 10.**

Lecture 11: Isometries and unitaries

- ▶ We recall a few things from the last lecture.

Lecture 11: Isometries and unitaries

- ▶ We recall a few things from the last lecture.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

Lecture 11: Isometries and unitaries

- ▶ We recall a few things from the last lecture.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.

Lecture 11: Isometries and unitaries

- ▶ We recall a few things from the last lecture.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.
- ▶ **Example 9.6:** For $\mathbb{R}^n$ (or $\mathbb{C}^n$) the standard basis $\{e_1, e_2, \dots, e_n\}$, where e_j is the vector whose j -th coordinate is one and all other coordinates are equal to zero, is an orthonormal basis with respect to the standard inner product.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.
- ▶ **Theorem 9.7:** Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then for any vector $w \in V$,

$$w = \sum_{j=1}^n \langle v_j, w \rangle v_j.$$

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$
- ▶ Fix any $j, 1 \leq j \leq n$ and consider the basis vector v_j .

Matrix of a linear map

- ▶ We recall how we define a matrix for a linear map from one finite dimensional space to another on fixing bases for these spaces.
- ▶ Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V and $\mathcal{C} = \{w_1, \dots, w_m\}$ is a basis for W . In particular, the dimension of V is n and the dimension of W is m .
- ▶ Let $T : V \rightarrow W$ be a linear map. We associate an $m \times n$ matrix A to T as described below and call it the matrix of T in bases $\mathcal{B}, \mathcal{C}$
- ▶ Fix any $j, 1 \leq j \leq n$ and consider the basis vector v_j .
- ▶ Now Tv_j is a vector in W and $\mathcal{C}$ is a basis for W .

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

Continuation

- Therefore, Tv_j is a linear combination of w_i 's. Denote the corresponding coefficients as a_{ij} 's. That is, a_{ij} is determined by requiring:

$$Tv_j = \sum_{i=1}^m a_{ij} w_i, \quad 1 \leq j \leq n.$$

- This defines the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ and is denoted as ${}_C[T]_B$. Observe that if $x = \sum_{j=1}^n x_j v_j$ then by linearity

$$\begin{aligned} Tx &= \sum_{j=1}^n x_j (Tv_j) \\ &= \sum_{j=1}^n \sum_{i=1}^m x_j (a_{ij} w_i) \\ &= \sum_{i=1}^m \left[\sum_{j=1}^n a_{ij} x_j \right] w_i. \end{aligned}$$

- **Conclusion:** For a linear map $T : V \rightarrow W$, the matrix of T in bases $\mathcal{B}, \mathcal{C}$ is the unique matrix A which satisfies

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n a_{ij} x_j \right] w_i.$$

for $x = \sum_{j=1}^n x_j v_j$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.
- ▶ For general $x \in V$, we get

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n \langle w_i, Tv_j \rangle \langle v_j, x \rangle \right] w_i$$

Maps on inner product spaces

- ▶ Consider the set up as above, with additional assumptions that V, W are inner product spaces and $\mathcal{B}, \mathcal{C}$ are orthonormal bases.
- ▶ Recall that for any vector $x \in V$, if $x = \sum_{j=1}^n x_j v_j$ then $x_j = \langle v_j, x \rangle$ so that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ Similarly, considering the orthonormal basis $\mathcal{C}$ in W , for fixed j , $Tv_j = \sum_{i=1}^m a_{ij} w_i$ implies that $a_{ij} = \langle w_i, Tv_j \rangle$.
- ▶ For general $x \in V$, we get

$$Tx = \sum_{i=1}^m \left[\sum_{j=1}^n \langle w_i, Tv_j \rangle \langle v_j, x \rangle \right] w_i$$

- ▶ We summarize this as a theorem.

The matrix of a linear transformation under orthonormal bases

- **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

The matrix of a linear transformation under orthonormal bases

- ▶ **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

- ▶ Conversely, given any $m \times n$ matrix $A = [a_{ij}]$, there exists unique linear map $T : V \rightarrow W$ satisfying

$$a_{ij} = \langle w_i, Tv_j \rangle, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

The matrix of a linear transformation under orthonormal bases

- ▶ **Theorem 10.1:** Let V, W be inner product spaces with orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ for some $m, n \in \mathbb{N}$. Let $T : V \rightarrow W$ be a linear map. Then the matrix of T in these bases is given by the $m \times n$ matrix $A = [a_{ij}]_{1 \leq i \leq m; 1 \leq j \leq n}$ where

$$a_{ij} = \langle w_i, Tv_j \rangle.$$

- ▶ Conversely, given any $m \times n$ matrix $A = [a_{ij}]$, there exists unique linear map $T : V \rightarrow W$ satisfying

$$a_{ij} = \langle w_i, Tv_j \rangle, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ Note that here:

$$Tv_j = \sum_{i=1}^m \langle w_i, Tv_j \rangle w_i = \sum_{i=1}^m a_{ij} w_i.$$

(Hermitian) adjoint

- **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

(Hermitian) adjoint

- **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.
- ▶ Consider the $n \times m$ matrix A^* defined by

$$(A^*)_{ji} = \overline{a_{ij}}, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

(Hermitian) adjoint

- ▶ **Theorem 10.2:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then there exists a unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ **Proof.** Choose an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V and an orthonormal basis $\mathcal{C} = \{w_1, \dots, w_m\}$ for W . (Note that such orthonormal bases exist as we can apply Gram-Schmidt orthogonalization on some bases).
- ▶ Let $A = [a_{ij}]$ be the matrix of T in this bases.
- ▶ Consider the $n \times m$ matrix A^* defined by

$$(A^*)_{ji} = \overline{a_{ij}}, \quad 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ We know that A^* determines a linear map $S : W \rightarrow V$ satisfying

$$\langle v_j, Sw_i \rangle = (A^*)_{ji} = \overline{a_{ij}}.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or
$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.
- ▶ The uniqueness is clear, as we can see that any linear map S with required property has the matrix A^* as the matrix in the given bases. ■

Continuation

- ▶ Taking complex conjugation, we have, $\langle Sw_i, v_j \rangle = a_{ij}$ or

$$\langle Sw_i, v_j \rangle = \langle w_i, Tv_j \rangle, \quad \forall 1 \leq i \leq m; 1 \leq j \leq n.$$

- ▶ By linearity of S, T we have

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad \forall x \in V, y \in W.$$

- ▶ This proves the existence.
- ▶ The uniqueness is clear, as we can see that any linear map S with required property has the matrix A^* as the matrix in the given bases. ■
- ▶ **Definition 10.3:** Let V, W be finite dimensional inner product spaces and let $T : V \rightarrow W$ be a linear map. Then the unique linear map $S : W \rightarrow V$ satisfying

$$\langle Sy, x \rangle = \langle y, Tx \rangle, \quad x \in V, y \in W,$$

is known as the (Hermitian) adjoint of T and is denoted by T^* .

Basic properties of the adjoint

- **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
- ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$.
(Anti-linearity).

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
 - ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$. (Anti-linearity).
 - ▶ (ii) $((T_1)^*)^* = T_1$. (Involution property).

Basic properties of the adjoint

- ▶ **Theorem 10.4:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $T_1 : V \rightarrow W$ and $T_2 : V \rightarrow W$ be linear maps. Then
 - ▶ (i) For $c_1, c_2 \in \mathbb{F}$, $(c_1 T_1 + c_2 T_2)^* = \overline{c_1} T_1^* + \overline{c_2} T_2^*$. (Anti-linearity).
 - ▶ (ii) $((T_1)^*)^* = T_1$. (Involution property).
- ▶ **Proof.** Exercise.

Composition

- **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

Composition

- **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^*T^*.$$

- **Proof.** For $x \in U$ and $z \in W$,

$$\langle S^*T^*z, x \rangle = \langle T^*z, Sx \rangle = \langle z, TSx \rangle.$$

Composition

- ▶ **Theorem 10.5:** Let U, V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : U \rightarrow V$ and $T : V \rightarrow W$ be linear maps. Then

$$(TS)^* = S^* T^*.$$

- ▶ **Proof.** For $x \in U$ and $z \in W$,

$$\langle S^* T^* z, x \rangle = \langle T^* z, Sx \rangle = \langle z, TSx \rangle.$$

- ▶ Now from the uniqueness of the adjoint, we get $(TS)^* = S^* T^*$.

Isometries and unitaries

- **Definition 11.1:** Let V, W be inner product spaces over a field $\mathbb{F}$. Then a linear map $S : V \rightarrow W$ is said to be an **isometry** if

$$\|Sx\| = \|x\|, \quad \forall x \in V.$$

A bijective isometry is said to be an **unitary**.

Isometries and unitaries

- **Definition 11.1:** Let V, W be inner product spaces over a field $\mathbb{F}$. Then a linear map $S : V \rightarrow W$ is said to be an **isometry** if

$$\|Sx\| = \|x\|, \quad \forall x \in V.$$

A bijective isometry is said to be an **unitary**.

- **Example 11.2:** Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Then

Isometries and unitaries

- **Definition 11.1:** Let V, W be inner product spaces over a field $\mathbb{F}$. Then a linear map $S : V \rightarrow W$ is said to be an **isometry** if

$$\|Sx\| = \|x\|, \quad \forall x \in V.$$

A bijective isometry is said to be an **unitary**.

- **Example 11.2:** Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Then
- (i) $S_1 : V \rightarrow V$ defined by

$$S_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$

is a unitary.

Isometries and unitaries

- **Definition 11.1:** Let V, W be inner product spaces over a field $\mathbb{F}$. Then a linear map $S : V \rightarrow W$ is said to be an **isometry** if

$$\|Sx\| = \|x\|, \quad \forall x \in V.$$

A bijective isometry is said to be an **unitary**.

- **Example 11.2:** Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Then
- (i) $S_1 : V \rightarrow V$ defined by

$$S_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$

is a unitary.

- (ii) $S_2 : V \rightarrow W$ defined by

$$S_2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$$

is an isometry.

- ▶ (iii) $S_3 : V \rightarrow V$ defined by

$$S_3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2}{\sqrt{2}} \\ \frac{x_1-x_2}{\sqrt{2}} \end{pmatrix}$$

is a unitary.

Characterizations of isometries

- **Theorem 11.2:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:

Characterizations of isometries

- ▶ **Theorem 11.2:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is an isometry, that is, $\|Sx\| = \|x\|$, $\forall x \in V$. (S preserves the norm.)

Characterizations of isometries

- ▶ **Theorem 11.2:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is an isometry, that is, $\|Sx\| = \|x\|$, $\forall x \in V$. (S preserves the norm.)
 - ▶ (ii) S preserves the metric:

$$d(Sx, Sy) = d(x, y), \quad \forall x, y \in V$$

where $d(x, y) = \|y - x\|$, $x, y \in V$.

Characterizations of isometries

- ▶ **Theorem 11.2:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is an isometry, that is, $\|Sx\| = \|x\|$, $\forall x \in V$. (S preserves the norm.)
 - ▶ (ii) S preserves the metric:

$$d(Sx, Sy) = d(x, y), \quad \forall x, y \in V$$

where $d(x, y) = \|y - x\|$, $x, y \in V$.

- ▶ (iii) S preserves the inner product:

$$\langle Sx, Sy \rangle = \langle x, y \rangle, \quad \forall x, y \in V.$$

Characterizations of isometries

- ▶ **Theorem 11.2:** Let V, W be finite dimensional inner product spaces over a field $\mathbb{F}$. Let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is an isometry, that is, $\|Sx\| = \|x\|$, $\forall x \in V$. (S preserves the norm.)
 - ▶ (ii) S preserves the metric:

$$d(Sx, Sy) = d(x, y), \quad \forall x, y \in V$$

where $d(x, y) = \|y - x\|$, $x, y \in V$.

- ▶ (iii) S preserves the inner product:

$$\langle Sx, Sy \rangle = \langle x, y \rangle, \quad \forall x, y \in V.$$

- ▶ (iv) $S^*S = I_V$, where I_V denotes the identity of V .

Continuation

► **Proof.** (i) $\Rightarrow$ (ii). This is clear, as

$$d(Sx, Sy) = \|Sy - Sx\| = \|S(y - x)\| = \|y - x\| = d(x, y), \quad \forall x, y \in V$$

Continuation

- ▶ **Proof.** $(i) \Rightarrow (ii)$. This is clear, as

$$d(Sx, Sy) = \|Sy - Sx\| = \|S(y - x)\| = \|y - x\| = d(x, y), \quad \forall x, y \in V$$

- ▶ $(ii) \Rightarrow (i)$ is clear by taking $y = 0$.

Continuation

- ▶ **Proof.** (i) $\Rightarrow$ (ii). This is clear, as

$$d(Sx, Sy) = \|Sy - Sx\| = \|S(y - x)\| = \|y - x\| = d(x, y), \quad \forall x, y \in V$$

- ▶ (ii) $\Rightarrow$ (i) is clear by taking $y = 0$.
- ▶ (i) $\Rightarrow$ (iii). If $\mathbb{F} = \mathbb{R}$, we have

$$\begin{aligned}\langle Sx, Sy \rangle &= \frac{1}{4}(\langle S(x + y), S(x + y) \rangle - \langle S(x - y), S(x - y) \rangle) \\ &= \frac{1}{4}(\langle (x + y), (x + y) \rangle - \langle (x - y), (x - y) \rangle) \\ &= \langle x, y \rangle.\end{aligned}$$

Continuation

- ▶ **Proof.** (i) $\Rightarrow$ (ii). This is clear, as

$$d(Sx, Sy) = \|Sy - Sx\| = \|S(y - x)\| = \|y - x\| = d(x, y), \quad \forall x, y \in V$$

- ▶ (ii) $\Rightarrow$ (i) is clear by taking $y = 0$.
- ▶ (i) $\Rightarrow$ (iii). If $\mathbb{F} = \mathbb{R}$, we have

$$\begin{aligned}\langle Sx, Sy \rangle &= \frac{1}{4}(\langle S(x + y), S(x + y) \rangle - \langle S(x - y), S(x - y) \rangle) \\ &= \frac{1}{4}(\langle (x + y), (x + y) \rangle - \langle (x - y), (x - y) \rangle) \\ &= \langle x, y \rangle.\end{aligned}$$

- ▶ Similarly, using polarization identity we have the result for $\mathbb{F} = \mathbb{C}$.

Continuation

- ▶ (iii) $\Rightarrow$ (iv) From the defining property of the adjoint and (iii): For x, y in V ,

$$\langle x, S^* Sy \rangle = \langle Sx, Sy \rangle = \langle x, y \rangle.$$

Continuation

- ▶ (iii) $\Rightarrow$ (iv) From the defining property of the adjoint and (iii): For x, y in V ,

$$\langle x, S^*Sy \rangle = \langle Sx, Sy \rangle = \langle x, y \rangle.$$

- ▶ Hence

$$\langle x, (S^*S - I_V)y \rangle = 0, \quad \forall x, y \in V.$$

Continuation

- ▶ (iii) $\Rightarrow$ (iv) From the defining property of the adjoint and (iii): For x, y in V ,

$$\langle x, S^*Sy \rangle = \langle Sx, Sy \rangle = \langle x, y \rangle.$$

- ▶ Hence

$$\langle x, (S^*S - I_V)y \rangle = 0, \quad \forall x, y \in V.$$

- ▶ For $y \in V$, taking $x = (S^*S - I_V)y$, we get

$$\langle (S^*S - I_V)y, (S^*S - I_V)y \rangle = 0,$$

Continuation

- ▶ (iii) $\Rightarrow$ (iv) From the defining property of the adjoint and (iii): For x, y in V ,

$$\langle x, S^*Sy \rangle = \langle Sx, Sy \rangle = \langle x, y \rangle.$$

- ▶ Hence

$$\langle x, (S^*S - I_V)y \rangle = 0, \quad \forall x, y \in V.$$

- ▶ For $y \in V$, taking $x = (S^*S - I_V)y$, we get

$$\langle (S^*S - I_V)y, (S^*S - I_V)y \rangle = 0,$$

- ▶ or, $(S^*S - I_V)y = 0$ for all $y \in V$.

Continuation

- ▶ (iii) $\Rightarrow$ (iv) From the defining property of the adjoint and (iii): For x, y in V ,

$$\langle x, S^*Sy \rangle = \langle Sx, Sy \rangle = \langle x, y \rangle.$$

- ▶ Hence

$$\langle x, (S^*S - I_V)y \rangle = 0, \quad \forall x, y \in V.$$

- ▶ For $y \in V$, taking $x = (S^*S - I_V)y$, we get

$$\langle (S^*S - I_V)y, (S^*S - I_V)y \rangle = 0,$$

- ▶ or, $(S^*S - I_V)y = 0$ for all $y \in V$.
- ▶ Equivalently $S^*S - I_V = 0$ or $S^*S = I_V$.

Continuation

- (iv) $\Rightarrow$ (i) : For $x \in V$, we have

$$\begin{aligned}\langle Sx, Sx \rangle &= \langle x, S^* Sx \rangle \\ &= \langle x, x \rangle\end{aligned}$$

- ▶ $(iv) \Rightarrow (i)$: For $x \in V$, we have

$$\begin{aligned}\langle Sx, Sx \rangle &= \langle x, S^* Sx \rangle \\ &= \langle x, x \rangle\end{aligned}$$

- ▶ This completes the proof of Theorem 11.2.

Continuation

- ▶ $(iv) \Rightarrow (i)$: For $x \in V$, we have

$$\begin{aligned}\langle Sx, Sx \rangle &= \langle x, S^* Sx \rangle \\ &= \langle x, x \rangle\end{aligned}$$

- ▶ This completes the proof of Theorem 11.2.
- ▶ **Corollary 11.3:** Suppose V, W are finite dimensional inner product spaces and $S : V \rightarrow W$ is an isometry. Then for any orthonormal collection $\{v_1, \dots, v_k\}$ in V , $\{Sv_1, \dots, Sv_k\}$ is orthonormal in W .

Continuation

- ▶ $(iv) \Rightarrow (i)$: For $x \in V$, we have

$$\begin{aligned}\langle Sx, Sx \rangle &= \langle x, S^* Sx \rangle \\ &= \langle x, x \rangle\end{aligned}$$

- ▶ This completes the proof of Theorem 11.2.
- ▶ **Corollary 11.3:** Suppose V, W are finite dimensional inner product spaces and $S : V \rightarrow W$ is an isometry. Then for any orthonormal collection $\{v_1, \dots, v_k\}$ in V , $\{Sv_1, \dots, Sv_k\}$ is orthonormal in W .
- ▶ **Proof:** Clear from (iii) of previous theorem. ■

Existence of isometries

- **Corollary 11.4:** Suppose V, W are finite dimensional inner product spaces on a field $\mathbb{F}$. Then there exists an isometry $S : V \rightarrow W$ if and only if $\dim(V) \leq \dim(W)$.

Existence of isometries

- ▶ **Corollary 11.4:** Suppose V, W are finite dimensional inner product spaces on a field $\mathbb{F}$. Then there exists an isometry $S : V \rightarrow W$ if and only if $\dim(V) \leq \dim(W)$.
- ▶ **Proof.** Suppose there exists an isometry $S : V \rightarrow W$.

Existence of isometries

- ▶ **Corollary 11.4:** Suppose V, W are finite dimensional inner product spaces on a field $\mathbb{F}$. Then there exists an isometry $S : V \rightarrow W$ if and only if $\dim(V) \leq \dim(W)$.
- ▶ **Proof.** Suppose there exists an isometry $S : V \rightarrow W$.
- ▶ If $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , then $\{Sv_1, \dots, Sv_n\}$ is an orthonormal collection in W .

Existence of isometries

- ▶ **Corollary 11.4:** Suppose V, W are finite dimensional inner product spaces on a field $\mathbb{F}$. Then there exists an isometry $S : V \rightarrow W$ if and only if $\dim(V) \leq \dim(W)$.
- ▶ **Proof.** Suppose there exists an isometry $S : V \rightarrow W$.
- ▶ If $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , then $\{Sv_1, \dots, Sv_n\}$ is an orthonormal collection in W .
- ▶ In particular, $\{Sv_1, \dots, Sv_n\}$ are linearly independent in W . Hence $\dim(W) \geq n = \dim(V)$.

Existence of isometries

- ▶ **Corollary 11.4:** Suppose V, W are finite dimensional inner product spaces on a field $\mathbb{F}$. Then there exists an isometry $S : V \rightarrow W$ if and only if $\dim(V) \leq \dim(W)$.
- ▶ **Proof.** Suppose there exists an isometry $S : V \rightarrow W$.
- ▶ If $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , then $\{Sv_1, \dots, Sv_n\}$ is an orthonormal collection in W .
- ▶ In particular, $\{Sv_1, \dots, Sv_n\}$ are linearly independent in W . Hence $\dim(W) \geq n = \dim(V)$.
- ▶ The converse is an exercise.

- ▶ Recall that bijective isometries are called unitaries.

Unitaries

- ▶ Recall that bijective isometries are called unitaries.
- ▶ **Theorem 11.5:** Let V, W be finite dimensional inner product spaces and let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:

Unitaries

- ▶ Recall that bijective isometries are called unitaries.
- ▶ **Theorem 11.5:** Let V, W be finite dimensional inner product spaces and let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is a unitary;

Unitaries

- ▶ Recall that bijective isometries are called unitaries.
- ▶ **Theorem 11.5:** Let V, W be finite dimensional inner product spaces and let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is a unitary;
 - ▶ (ii) $\langle Sx, Sy \rangle = \langle x, y \rangle$ for all x, y in X and S is onto.

Unitaries

- ▶ Recall that bijective isometries are called unitaries.
- ▶ **Theorem 11.5:** Let V, W be finite dimensional inner product spaces and let $S : V \rightarrow W$ be a linear map. Then the following are equivalent:
 - ▶ (i) S is a unitary;
 - ▶ (ii) $\langle Sx, Sy \rangle = \langle x, y \rangle$ for all x, y in X and S is onto.
 - ▶ (iii) $S^*S = I_V$ and $SS^* = I_W$.

Continuation

- **Proof:** $(i) \Rightarrow (ii)$ is clear as S is an isometry and is given to be bijective. In particular, it is onto.

Continuation

- ▶ **Proof:** $(i) \Rightarrow (ii)$ is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ $(ii) \Rightarrow (iii)$. Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.

Continuation

- ▶ **Proof:** $(i) \Rightarrow (ii)$ is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ $(ii) \Rightarrow (iii)$. Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.
- ▶ S is also given to be surjective. Suppose there exists x, y in V such that $Sx = Sy$.

Continuation

- ▶ **Proof:** $(i) \Rightarrow (ii)$ is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ $(ii) \Rightarrow (iii)$. Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.
- ▶ S is also given to be surjective. Suppose there exists x, y in V such that $Sx = Sy$.
- ▶ This means that $S(x - y) = 0$. Hence

$$0 = \langle S(x - y), S(x - y) \rangle = \langle (x - y), (x - y) \rangle.$$

Continuation

- ▶ **Proof:** $(i) \Rightarrow (ii)$ is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ $(ii) \Rightarrow (iii)$. Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.
- ▶ S is also given to be surjective. Suppose there exists x, y in V such that $Sx = Sy$.
- ▶ This means that $S(x - y) = 0$. Hence

$$0 = \langle S(x - y), S(x - y) \rangle = \langle (x - y), (x - y) \rangle.$$

- ▶ Consequently $x - y = 0$ or $x = y$. This shows that S is injective. Therefore S is invertible as a function.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii) is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ (ii) $\Rightarrow$ (iii). Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.
- ▶ S is also given to be surjective. Suppose there exists x, y in V such that $Sx = Sy$.
- ▶ This means that $S(x - y) = 0$. Hence

$$0 = \langle S(x - y), S(x - y) \rangle = \langle (x - y), (x - y) \rangle.$$

- ▶ Consequently $x - y = 0$ or $x = y$. This shows that S is injective. Therefore S is invertible as a function.
- ▶ As $S^*S = I_V$, we get $S^* = S^{-1}$ and hence $SS^* = I_W$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii) is clear as S is an isometry and is given to be bijective. In particular, it is onto.
- ▶ (ii) $\Rightarrow$ (iii). Assuming (ii) we already know that S is an isometry and consequently, $S^*S = I_V$.
- ▶ S is also given to be surjective. Suppose there exists x, y in V such that $Sx = Sy$.
- ▶ This means that $S(x - y) = 0$. Hence

$$0 = \langle S(x - y), S(x - y) \rangle = \langle (x - y), (x - y) \rangle.$$

- ▶ Consequently $x - y = 0$ or $x = y$. This shows that S is injective. Therefore S is invertible as a function.
- ▶ As $S^*S = I_V$, we get $S^* = S^{-1}$ and hence $SS^* = I_W$.
- ▶ (iii) $\Rightarrow$ (i). Now as $S^*S = I_V$, S is isometric. As S is invertible, it is a bijection.

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.
- ▶ (i) A linear map $S : V \rightarrow W$ is a unitary if and only if $S^* : W \rightarrow V$ is a unitary;

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.
- ▶ (i) A linear map $S : V \rightarrow W$ is a unitary if and only if $S^* : W \rightarrow V$ is a unitary;
- ▶ (ii) There exists a unitary $S : V \rightarrow W$ if and only if $\dim(V) = \dim(W)$;

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.
- ▶ (i) A linear map $S : V \rightarrow W$ is a unitary if and only if $S^* : W \rightarrow V$ is a unitary;
- ▶ (ii) There exists a unitary $S : V \rightarrow W$ if and only if $\dim(V) = \dim(W)$;
- ▶ (iii) Suppose $S : V \rightarrow W$ is a unitary. Then whenever $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , $\{Sv_1, \dots, Sv_n\}$ is an orthonormal basis of W .

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.
- ▶ (i) A linear map $S : V \rightarrow W$ is a unitary if and only if $S^* : W \rightarrow V$ is a unitary;
- ▶ (ii) There exists a unitary $S : V \rightarrow W$ if and only if $\dim(V) = \dim(W)$;
- ▶ (iii) Suppose $S : V \rightarrow W$ is a unitary. Then whenever $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , $\{Sv_1, \dots, Sv_n\}$ is an orthonormal basis of W .
- ▶ (iv) If $S : V \rightarrow W$ is a linear map such that it maps some orthonormal basis $\{v_1, \dots, v_n\}$ of V to some orthonormal basis $\{Sv_1, \dots, Sv_n\}$ of W , then S is a unitary.

More on unitaries

- ▶ **Corollary 11.6:** Let V, W be finite dimensional inner product spaces.
- ▶ (i) A linear map $S : V \rightarrow W$ is a unitary if and only if $S^* : W \rightarrow V$ is a unitary;
- ▶ (ii) There exists a unitary $S : V \rightarrow W$ if and only if $\dim(V) = \dim(W)$;
- ▶ (iii) Suppose $S : V \rightarrow W$ is a unitary. Then whenever $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , $\{Sv_1, \dots, Sv_n\}$ is an orthonormal basis of W .
- ▶ (iv) If $S : V \rightarrow W$ is a linear map such that it maps some orthonormal basis $\{v_1, \dots, v_n\}$ of V to some orthonormal basis $\{Sv_1, \dots, Sv_n\}$ of W , then S is a unitary.
- ▶ (v) Suppose $S : V \rightarrow V$ is an isometry then it is a unitary.
- ▶ **Proof:** Exercise.

Matrices of unitaries

- **Definition 11.7:** An $n \times n$ complex matrix A is said to be a **unitary matrix** if and only if

$$A^*A = I = AA^*.$$

Matrices of unitaries

- **Definition 11.7:** An $n \times n$ complex matrix A is said to be a **unitary matrix** if and only if

$$A^*A = I = AA^*.$$

- Note that here $A^*A = I$ implies $I = AA^*$ and vice versa. Any unitary matrix is matrix of a unitary map in some orthonormal bases.

Matrices of unitaries

- **Definition 11.7:** An $n \times n$ complex matrix A is said to be a **unitary matrix** if and only if

$$A^*A = I = AA^*.$$

- Note that here $A^*A = I$ implies $I = AA^*$ and vice versa. Any unitary matrix is matrix of a unitary map in some orthonormal bases.
- Observe that an $n \times n$ complex matrix A is a unitary if and only if columns (equivalently rows) of A form an orthonormal basis of $\mathbb{C}^n$.

Matrices of unitaries

- **Definition 11.7:** An $n \times n$ complex matrix A is said to be a **unitary matrix** if and only if

$$A^*A = I = AA^*.$$

- Note that here $A^*A = I$ implies $I = AA^*$ and vice versa. Any unitary matrix is matrix of a unitary map in some orthonormal bases.
- Observe that an $n \times n$ complex matrix A is a unitary if and only if columns (equivalently rows) of A form an orthonormal basis of $\mathbb{C}^n$.
- **Definition 11.8:** An $n \times n$ real matrix A is said to be an **orthogonal matrix** if

$$A^tA = I = AA^t.$$

Matrices of unitaries

- ▶ **Definition 11.7:** An $n \times n$ complex matrix A is said to be a **unitary matrix** if and only if

$$A^*A = I = AA^*.$$

- ▶ Note that here $A^*A = I$ implies $I = AA^*$ and vice versa. Any unitary matrix is matrix of a unitary map in some orthonormal bases.
- ▶ Observe that an $n \times n$ complex matrix A is a unitary if and only if columns (equivalently rows) of A form an orthonormal basis of $\mathbb{C}^n$.
- ▶ **Definition 11.8:** An $n \times n$ real matrix A is said to be an **orthogonal matrix** if

$$A^tA = I = AA^t.$$

- ▶ Here too, $A^tA = I$ implies $AA^t = I$ and vice versa and these are matrices of unitary maps on finite dimensional real inner product spaces, on orthonormal bases.

Examples

- **Examples 11.9:** The following are orthogonal matrices: $\theta \in \mathbb{R}$.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$

Examples

- ▶ **Examples 11.9:** The following are orthogonal matrices: $\theta \in \mathbb{R}$.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$

- ▶ All permutation matrices are orthogonal matrices.

Examples

- **Examples 11.9:** The following are orthogonal matrices: $\theta \in \mathbb{R}$.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$

- All permutation matrices are orthogonal matrices.
- **Examples 11.10:** The following are unitary matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & z_1 & 0 \\ 0 & 0 & z_2 \\ z_3 & 0 & 0 \end{bmatrix},$$

where $|z_1| = |z_2| = |z_3| = 1$.

Examples

- **Examples 11.9:** The following are orthogonal matrices: $\theta \in \mathbb{R}$.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$

- All permutation matrices are orthogonal matrices.
- **Examples 11.10:** The following are unitary matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & z_1 & 0 \\ 0 & 0 & z_2 \\ z_3 & 0 & 0 \end{bmatrix},$$

where $|z_1| = |z_2| = |z_3| = 1$.

- **END OF LECTURE 11.**

Lecture 12: Orthogonal decomposition

- ▶ We recall a few things from previous lectures.

Lecture 12: Orthogonal decomposition

- ▶ We recall a few things from previous lectures.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

Lecture 12: Orthogonal decomposition

- ▶ We recall a few things from previous lectures.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.

Lecture 12: Orthogonal decomposition

- ▶ We recall a few things from previous lectures.
- ▶ **Definition 9.5:** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then a basis $\{v_1, v_2, \dots, v_n\}$ is said to be an **orthonormal basis** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

- ▶ In other words, an orthonormal basis is a basis consisting of mutually orthogonal unit vectors.
- ▶ **Example 9.6:** For $\mathbb{R}^n$ (or $\mathbb{C}^n$) the standard basis $\{e_1, e_2, \dots, e_n\}$, where e_j is the vector whose j -th coordinate is one and all other coordinates are equal to zero, is an orthonormal basis with respect to the standard inner product.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.

A formula for coefficients

- ▶ What is the advantage of having an orthonormal basis instead of ordinary basis? This is answered by the following theorem.
- ▶ It gives a formula for the coefficients in the expansion of any vector in terms of the basis.
- ▶ **Theorem 9.7:** Let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then for any vector $w \in V$,

$$w = \sum_{j=1}^n \langle v_j, w \rangle v_j.$$

Orthogonal complement

- **Definition 12.1** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Orthogonal complement

- **Definition 12.1** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

Orthogonal complement

- **Definition 12.1** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

- Then

$$S^\perp = \left\{ \begin{pmatrix} c \\ c \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Continuation

- **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.
- ▶ We have already seen that orthogonal complement of any non-empty subset is a subspace. In particular, $(S^\perp)^\perp$ is a subspace.

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.
- ▶ Clearly,

$$V_2 = \left\{ \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} : x_3 \in \mathbb{R} \right\}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

- ▶ We want to show that this is a general phenomenon.

Extending bases and orthonormal bases

- **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .
- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

- ▶ If $V = M_{k+1}$ then $\{v_1, \dots, v_{k+1}\}$ is a basis for V and we are done. If not, take $v_{k+2} \in V \setminus M_{k+1}$ and continue the induction process.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.
- ▶ Therefore $V = M_n$ for some n and $\{v_1, \dots, v_n\}$ is a basis for V .

Extending orthonormal bases

- **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .
- ▶ It is an elementary exercise to see that $e_j = v_j$ for $1 \leq j \leq k$ as $v_1, \dots, v_k$ are already orthonormal. ■

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.
- ▶ This shows $\langle x, y \rangle = 0$ for all $x \in V_0$ and $y \in V_1$. Hence $V_1 \subseteq (V_0)^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get
$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j.$$

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.
- ▶ This proves $(V_0)^\perp \subseteq V_1$ and completes the proof of our claim.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .
- ▶ Now we know that any $x \in V$ decomposes as

$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

Uniqueness

► Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.
- ▶ Hence $\langle y - y', y - y' \rangle = 0$. Consequently $y = y'$ and $z' = z$.
This proves the uniqueness.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .

Continuation

- ▶ We have $\|z\|^2 \geq 0$.

Continuation

- ▶ We have $\|z\|^2 \geq 0$.
- ▶ This means that:

$$\|x - \langle v, x \rangle v\|^2 \geq 0.$$

Continuation

- ▶ We have $\|z\|^2 \geq 0$.
- ▶ This means that:

$$\|x - \langle v, x \rangle v\|^2 \geq 0.$$

- ▶ Here

$$\begin{aligned}\|x - \langle v, x \rangle v\|^2 &= \langle x, x \rangle - 2|\langle x, v \rangle|^2 + |\langle x, v \rangle|^2 \\ &= \|x\|^2 - |\langle x, v \rangle|^2 \\ &= \|x\|^2 - \left| \left\langle x, \frac{y}{\|y\|} \right\rangle \right|^2\end{aligned}$$

Continuation

- ▶ We have $\|z\|^2 \geq 0$.
- ▶ This means that:

$$\|x - \langle v, x \rangle v\|^2 \geq 0.$$

- ▶ Here

$$\begin{aligned}\|x - \langle v, x \rangle v\|^2 &= \langle x, x \rangle - 2|\langle x, v \rangle|^2 + |\langle x, v \rangle|^2 \\ &= \|x\|^2 - |\langle x, v \rangle|^2 \\ &= \|x\|^2 - |\langle x, \frac{y}{\|y\|} \rangle|^2\end{aligned}$$

- ▶ and the positivity of this is same as the Cauchy-Schwarz inequality:

$$\|x\|^2 \|y\|^2 \geq |\langle x, y \rangle|^2.$$

Continuation

- ▶ The equality holds, only when $z = 0$, that is when $x \in \text{span}\{y\}$. (We have assumed $y \neq 0$.) This explains our proof of Cauchy-Schwarz inequality.

Continuation

- ▶ The equality holds, only when $z = 0$, that is when $x \in \text{span}\{y\}$. (We have assumed $y \neq 0$.) This explains our proof of Cauchy-Schwarz inequality.
- ▶ **Exercise 12.7 :** Consider examples of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ and try to understand the projection theorem in concrete cases.

Continuation

- ▶ The equality holds, only when $z = 0$, that is when $x \in \text{span}\{y\}$. (We have assumed $y \neq 0$.) This explains our proof of Cauchy-Schwarz inequality.
- ▶ **Exercise 12.7 :** Consider examples of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ and try to understand the projection theorem in concrete cases.
- ▶ **END OF LECTURE 12.**

Projection as linear map

- ▶ We recall: [Definition 12.1](#). Let S be a non-empty subset of an inner product space V . Then the [orthogonal complement](#) of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Projection as linear map

- We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

Projection as linear map

- ▶ We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

- ▶ Then

$$S^\perp = \left\{ \begin{pmatrix} c \\ c \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Continuation

- **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.
- ▶ We have already seen that orthogonal complement of any non-empty subset is a subspace. In particular, $(S^\perp)^\perp$ is a subspace.

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.
- ▶ Clearly,

$$V_2 = \left\{ \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} : x_3 \in \mathbb{R} \right\}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

- ▶ We want to show that this is a general phenomenon.

Extending bases and orthonormal bases

- **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .
- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

- ▶ If $V = M_{k+1}$ then $\{v_1, \dots, v_{k+1}\}$ is a basis for V and we are done. If not, take $v_{k+2} \in V \setminus M_{k+1}$ and continue the induction process.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.
- ▶ Therefore $V = M_n$ for some n and $\{v_1, \dots, v_n\}$ is a basis for V .

Extending orthonormal bases

- **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .
- ▶ It is an elementary exercise to see that $e_j = v_j$ for $1 \leq j \leq k$ as $v_1, \dots, v_k$ are already orthonormal. ■

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.
- ▶ This shows $\langle x, y \rangle = 0$ for all $x \in V_0$ and $y \in V_1$. Hence $V_1 \subseteq (V_0)^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get
$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j.$$

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.
- ▶ This proves $(V_0)^\perp \subseteq V_1$ and completes the proof of our claim.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .
- ▶ Now we know that any $x \in V$ decomposes as

$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

Uniqueness

► Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.
- ▶ Hence $\langle y - y', y - y' \rangle = 0$. Consequently $y = y'$ and $z' = z$.
This proves the uniqueness.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .
- ▶ As shown in the previous lecture this is related to Cauchy-Schwarz inequality.

Example

- **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product.
Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

- ▶ Let us apply Gram-Schmidt on this to get an orthonormal basis for V_0 .

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

- ▶ Now take

$$\begin{aligned} w_2 &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \left\langle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}. \end{aligned}$$

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

Continuation

- ▶ Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- ▶ $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .
- Given $x \in \mathbb{R}^3$, it decomposes as $y + z$, where $y \in V_0$, $z \in V_0^\perp$.

$$\begin{aligned} y &= \langle v_1, x \rangle v_1 + \langle v_2, x \rangle v_2 \\ &= \frac{x_1 - x_2}{\sqrt{2}} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} + \frac{(x_1 + x_2 - 2x_3)}{\sqrt{6}} \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 2x_1 - x_2 - x_3 \\ -x_1 + 2x_2 - x_3 \\ -x_1 - x_2 + 2x_3 \end{pmatrix} \end{aligned}$$

Continuation

$$\blacktriangleright z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

► It is easy to see that $y \in V_0$, $z \in (V_0)^\perp$ and $x = y + z$.

Lecture 13. Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

Lecture 13. Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .

Lecture 13. Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,

Lecture 13. Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.

Lecture 13. Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- (iii) $P(V) = V_0$.

Lecture 13. Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.

Lecture 13. Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.
- ▶ (v) $P_{V_1} = I - P$ where $V_1 = (V_0)^\perp$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

- ▶ Since the inner product is linear in the second variable, P is a linear map. This proves (i).

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.
- ▶ This proves (ii).

- ▶ (iii). We want to show $P(V) = V_0$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.
- ▶ Hence $P^2(x) = P(x)$ for every x , or $P^2 = P$.

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .
- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

- ▶ This shows that $P^* = P$ from the defining property of the adjoint of P .

Continuation

- (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

- ▶ We have just revisited our formula for the expansion of x in terms of an orthonormal basis.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

- ▶ We have just revisited our formula for the expansion of x in terms of an orthonormal basis.
- ▶ **END OF LECTURE 13.**

Lecture 14: Best approximation property of projections

- ▶ We recall: [Definition 12.1](#). Let S be a non-empty subset of an inner product space V . Then the [orthogonal complement](#) of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Lecture 14: Best approximation property of projections

- ▶ We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

Lecture 14: Best approximation property of projections

- ▶ We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

- ▶ Then

$$S^\perp = \left\{ \begin{pmatrix} c \\ c \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Continuation

- **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.
- ▶ We have already seen that orthogonal complement of any non-empty subset is a subspace. In particular, $(S^\perp)^\perp$ is a subspace.

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.
- ▶ Clearly,

$$V_2 = \left\{ \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} : x_3 \in \mathbb{R} \right\}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

- ▶ We want to show that this is a general phenomenon.

Extending bases and orthonormal bases

- **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .
- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

- ▶ If $V = M_{k+1}$ then $\{v_1, \dots, v_{k+1}\}$ is a basis for V and we are done. If not, take $v_{k+2} \in V \setminus M_{k+1}$ and continue the induction process.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.
- ▶ Therefore $V = M_n$ for some n and $\{v_1, \dots, v_n\}$ is a basis for V .

Extending orthonormal bases

- **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .
- ▶ It is an elementary exercise to see that $e_j = v_j$ for $1 \leq j \leq k$ as $v_1, \dots, v_k$ are already orthonormal. ■

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.
- ▶ This shows $\langle x, y \rangle = 0$ for all $x \in V_0$ and $y \in V_1$. Hence $V_1 \subseteq (V_0)^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get
$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j.$$

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.
- ▶ This proves $(V_0)^\perp \subseteq V_1$ and completes the proof of our claim.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .
- ▶ Now we know that any $x \in V$ decomposes as

$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

Uniqueness

► Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

Uniqueness

- Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.
- ▶ Hence $\langle y - y', y - y' \rangle = 0$. Consequently $y = y'$ and $z' = z$.
This proves the uniqueness.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .
- ▶ As shown in the previous lecture this is related to Cauchy-Schwarz inequality.

Example

- **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product.
Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

- ▶ Let us apply Gram-Schmidt on this to get an orthonormal basis for V_0 .

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

- ▶ Now take

$$\begin{aligned} w_2 &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \left\langle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}. \end{aligned}$$

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

Continuation

- ▶ Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- ▶ $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .
- Given $x \in \mathbb{R}^3$, it decomposes as $y + z$, where $y \in V_0$, $z \in V_0^\perp$.

$$\begin{aligned} y &= \langle v_1, x \rangle v_1 + \langle v_2, x \rangle v_2 \\ &= \frac{x_1 - x_2}{\sqrt{2}} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} + \frac{(x_1 + x_2 - 2x_3)}{\sqrt{6}} \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 2x_1 - x_2 - x_3 \\ -x_1 + 2x_2 - x_3 \\ -x_1 - x_2 + 2x_3 \end{pmatrix} \end{aligned}$$

Continuation

$$\blacktriangleright z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

► It is easy to see that $y \in V_0$, $z \in (V_0)^\perp$ and $x = y + z$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- (iii) $P(V) = V_0$.

Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.

Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.
- ▶ (v) $P_{V_1} = I - P$ where $V_1 = (V_0)^\perp$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

- ▶ Since the inner product is linear in the second variable, P is a linear map. This proves (i).

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.
- ▶ This proves (ii).

- ▶ (iii). We want to show $P(V) = V_0$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.
- ▶ Hence $P^2(x) = P(x)$ for every x , or $P^2 = P$.

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .
- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

- ▶ This shows that $P^* = P$ from the defining property of the adjoint of P .

Continuation

- (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

- ▶ We have just revisited our formula for the expansion of x in terms of an orthonormal basis.

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.

Distance between sets

- ▶ **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- ▶ We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- ▶ In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .
- **Example 14.1:** Take $V = \mathbb{R}^2$. Take $a = (1, 0)$. Consider $B_1 = \{(x_1, x_2) : x_1 < 0\}$ and $B_2 = \{(x_1, x_2) : |x_1 - 1| \geq 1\}$

Distance between sets

- ▶ **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- ▶ We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- ▶ In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .
- ▶ **Example 14.1:** Take $V = \mathbb{R}^2$. Take $a = (1, 0)$. Consider $B_1 = \{(x_1, x_2) : x_1 < 0\}$ and $B_2 = \{(x_1, x_2) : |x_1 - 1| \geq 1\}$
- ▶ Then $d(a, B_1) = 1$ is not attained at any point. $d(a, B_2) = 1$ gets attained at two points.

Best approximation property

- **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

Best approximation property

- ▶ **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

- ▶ This theorem tells us that Px is the unique 'best approximation' for x in V_0 .

Best approximation property

- ▶ **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

- ▶ This theorem tells us that Px is the unique 'best approximation' for x in V_0 .
- ▶ **Proof:** Suppose $x = y + z$, is the unique decomposition of x , with $y \in V_0, z \in V_0^\perp$.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|y - v\|^2 + \|z\|^2.\end{aligned}$$

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

- ▶ This proves the theorem.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

- ▶ This proves the theorem.
- ▶ Note that we are using the 'Pythagoras theorem' of inner product spaces.

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.
- ▶

$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.



$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

- ▶ Let P_1 be the projection onto V_1 .

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.



$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

- ▶ Let P_1 be the projection onto V_1 .
- ▶ Then $P_1 x = \frac{1}{n}(x_1 + \cdots + x_n) =: \bar{x}$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

- ▶ This value is n times the variance of the tuple $\{x_1, \dots, x_n\}$. In other words,

$$\inf_{c \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (x_j - c)^2 = \text{Var} \{x_1, \dots, x_n\},$$

and the infimum gets attained only at $\bar{x}$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

- ▶ This value is n times the variance of the tuple $\{x_1, \dots, x_n\}$. In other words,

$$\inf_{c \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (x_j - c)^2 = \text{Var} \{x_1, \dots, x_n\},$$

and the infimum gets attained only at $\bar{x}$.

- ▶ **Exercise:** Work out more examples.

Example

► **Example 14.3:** Consider $V = \mathbb{R}^2$. Let

$$V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\} \text{ where } \theta \text{ is a fixed real number.}$$

Write down the matrix of the projection onto V_0 .

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let
- $$V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\} \text{ where } \theta \text{ is a fixed real number.}$$
- Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}.$$

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let
 $V_0 = \{c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R}\}$ where θ is a fixed real number.
Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}.$$

- Then for any vector $x \in \mathbb{R}^2$,

$$Px = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = v \langle v, x \rangle = \langle v, x \rangle \cdot v,$$

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let
 $V_0 = \{c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R}\}$ where θ is a fixed real number.
Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}.$$

- Then for any vector $x \in \mathbb{R}^2$,

$$Px = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = v \langle v, x \rangle = \langle v, x \rangle \cdot v,$$

- where

$$v = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}.$$

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

- ▶ You may verify $P = P^2 = P^*$ and $P(\mathbb{R}^2) = V_0$.

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

- ▶ You may verify $P = P^2 = P^*$ and $P(\mathbb{R}^2) = V_0$.

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

- ▶ You may verify $P = P^2 = P^*$ and $P(\mathbb{R}^2) = V_0$.
- ▶ **END OF LECTURE 14.**

Lecture 15: Algebraic description of projections

- ▶ We recall: [Definition 12.1](#). Let S be a non-empty subset of an inner product space V . Then the [orthogonal complement](#) of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Lecture 15: Algebraic description of projections

- ▶ We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

Lecture 15: Algebraic description of projections

- ▶ We recall: **Definition 12.1.** Let S be a non-empty subset of an inner product space V . Then the **orthogonal complement** of S is defined as:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ **Example 12.2:** Consider $S \subset \mathbb{R}^3$ where

$$S = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$$

- ▶ Then

$$S^\perp = \left\{ \begin{pmatrix} c \\ c \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Continuation

- **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.

Continuation

- ▶ **Proposition 12.2:** Let S be a non-empty subset of an inner product space V . Then $S^\perp$ is a subspace of V . Further, $(S^\perp)^\perp$ is a subspace containing S .
- ▶ **Proof:** We recall the definition of $S^\perp$:

$$S^\perp = \{v \in V : \langle x, v \rangle = 0, \quad \forall x \in S\}.$$

- ▶ Now if $v, w \in S^\perp$ and $c, d \in \mathbb{F}$: For $x \in S$,

$$\langle x, cv + dw \rangle = c\langle x, v \rangle + d\langle x, w \rangle = c \cdot 0 + d \cdot 0 = 0.$$

- ▶ Hence $cv + dw \in S^\perp$. This proves that $S^\perp$ is a subspace of V .
- ▶ It is easy to see that if $x \in S$ then $x \in (S^\perp)^\perp$. Therefore $S \subseteq (S^\perp)^\perp$.
- ▶ We have already seen that orthogonal complement of any non-empty subset is a subspace. In particular, $(S^\perp)^\perp$ is a subspace.

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.

$\mathbb{R}^2$ in $\mathbb{R}^3$

- ▶ Consider $V = \mathbb{R}^3$ with standard inner product.
- ▶ Consider the subspace

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

- ▶ Take $V_1 = (V_0)^\perp$.
- ▶ Clearly,

$$V_1 = \left\{ \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} : x_3 \in \mathbb{R} \right\}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

Continuation

- ▶ We see that any vector $x \in V$ decomposes uniquely as $x = y + z$ with $y \in V_0$ and $z \in V_1$.
- ▶ Indeed for

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

the only choice is:

$$y = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}; z = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}.$$

- ▶ We want to show that this is a general phenomenon.

Extending bases and orthonormal bases

- **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .
- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

Extending bases and orthonormal bases

- ▶ **Theorem 12.4:** Let V_0 be a non-trivial subspace of a finite dimensional vector space V . Then any basis of V_0 extends to a basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is a basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is a basis of V .

- ▶ **Proof:** Take

$$M_k := \text{span} \{v_1, v_2, \dots, v_k\}$$

- ▶ If $M_k = V$ then $V_0 = V$, $\{v_1, \dots, v_k\}$ is a basis for V and so no extension is required.
- ▶ If not, choose any $v_{k+1} \in V \setminus M_k$. Then $\{v_1, \dots, v_{k+1}\}$ is a linearly independent set (Why?). Take

$$M_{k+1} := \text{span}\{v_1, \dots, v_{k+1}\}.$$

- ▶ If $V = M_{k+1}$ then $\{v_1, \dots, v_{k+1}\}$ is a basis for V and we are done. If not, take $v_{k+2} \in V \setminus M_{k+1}$ and continue the induction process.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.

- ▶ The process terminates after a finite number of steps as V is finite dimensional and so it can have at most $\dim(V)$ linearly independent elements.
- ▶ Therefore $V = M_n$ for some n and $\{v_1, \dots, v_n\}$ is a basis for V .

Extending orthonormal bases

- **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .

Extending orthonormal bases

- ▶ **Theorem 12.5:** Let V_0 be a non-trivial subspace of a finite dimensional inner product space V . Then any orthonormal basis of V_0 extends to an orthonormal basis of V , that is, if $\{v_1, v_2, \dots, v_k\}$ is an orthonormal basis of V_0 then there exists $\{v_{k+1}, \dots, v_n\}$ such that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ **Proof:** By the previous theorem we may extend $\{v_1, \dots, v_k\}$ to a basis $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ of V .
- ▶ Now apply the Gram-Schmidt procedure on $\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ to get an orthonormal basis $\{e_1, \dots, e_n\}$ of V .
- ▶ It is an elementary exercise to see that $e_j = v_j$ for $1 \leq j \leq k$ as $v_1, \dots, v_k$ are already orthonormal. ■

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .
- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.

Orthogonal complement of a subspace

- ▶ Consider the set up as above, that is, V_0 is a non-trivial subspace of a finite dimensional inner product space V . Suppose $\{v_1, \dots, v_k\}$ is an orthonormal basis of V_0 and $\{v_1, \dots, v_n\}$ is an orthonormal basis of V .

- ▶ Take

$$V_1 = \text{span} \{v_{k+1}, \dots, v_n\}.$$

- ▶ We claim that $V_1 = (V_0)^\perp$ and $\{v_{k+1}, \dots, v_n\}$ is an ortho-normal basis of V_1 .
- ▶ The second part is obvious. We only need to prove $V_1 = (V_0)^\perp$.
- ▶ Note that $\langle v_i, v_j \rangle = 0$ for all $1 \leq i \leq k$ and $(k+1) \leq j \leq n$
- ▶ Therefore $\langle \sum_{i=1}^k c_i v_i, \sum_{j=(k+1)}^n c_j v_j \rangle$ for any scalars $c_1, \dots, c_n$.
- ▶ This shows $\langle x, y \rangle = 0$ for all $x \in V_0$ and $y \in V_1$. Hence $V_1 \subseteq (V_0)^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get
$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j.$$

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.

Continuation

- ▶ Suppose $x \in V_0^\perp$.
- ▶ As $\{v_1, \dots, v_n\}$ is an orthonormal basis of V , we get $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$.
- ▶ As x is orthogonal to V_0 , we get $\langle v_j, x \rangle = 0$ for $1 \leq j \leq k$.
- ▶ Hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$ and therefore $x \in V_1$.
- ▶ This proves $(V_0)^\perp \subseteq V_1$ and completes the proof of our claim.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

Projection theorem

- **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .

Projection theorem

- ▶ **Theorem 12.6:** Let V_0 be a subspace of a finite dimensional inner product space V . Then every $x \in V$ decomposes uniquely as

$$x = y + z$$

where $y \in V_0$ and $z \in V_0^\perp$.

- ▶ **Proof:** Suppose $V_0 = \{0\}$. Then $V_0^\perp = V$ and we can decompose x as $x = 0 + x$, with $0 \in V_0$ and $x \in V_0^\perp$.
- ▶ If $V_0 \neq \{0\}$, choose an orthonormal basis $\{v_1, \dots, v_k\}$ for V_0 . Extend it to an orthonormal basis $\{v_1, \dots, v_n\}$ of V .
- ▶ Now we know that any $x \in V$ decomposes as

$$x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

Uniqueness

► Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.

Uniqueness

- ▶ Take

$$y = \sum_{j=1}^k \langle v_j, x \rangle v_j$$

and

$$z = \sum_{j=(k+1)}^n \langle v_j, x \rangle v_j.$$

- ▶ Clearly $y \in V_0$ and $z \in V_0^\perp$. This proves the existence.
- ▶ Suppose $x = y + z$ and $x = y' + z'$ are two decompositions of x with $y, y' \in V_0$ and $z, z' \in V_0^\perp$.
- ▶ We have,

$$y + z = y' + z'.$$

- ▶ Therefore $y - y' = z' - z$. As $y, y' \in V_0$, $y - y' \in V_0$.
- ▶ Also as $z, z' \in V_0^\perp$, $y - y' = z' - z \in V_0^\perp$.
- ▶ Hence $\langle y - y', y - y' \rangle = 0$. Consequently $y = y'$ and $z' = z$.
This proves the uniqueness.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .

A special case

- ▶ Suppose V is a finite dimensional inner product space and let y be a non-zero vector in V .
- ▶ Consider the one dimensional space $V_0 = \{cy : c \in \mathbb{F}\}$.
- ▶ Now $\{v\}$ is an ortho-normal basis for V_0 where

$$v = \frac{y}{\|y\|}.$$

- ▶ Therefore any $x \in V$ decomposes as $x = \langle v, x \rangle v + z$ where z is orthogonal to v .
- ▶ As shown in the previous lecture this is related to Cauchy-Schwarz inequality.

Example

- **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product.
Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

Example

- ▶ **Example 13.1:** Let $V = \mathbb{R}^n$ with the standard inner product. Let $V_0 = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.
- ▶ We first analyze the case when $n = 3$. Now $V = \mathbb{R}^3$ and

$$V_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}.$$

- ▶ One can see that

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

is a basis for V_0 .

- ▶ Let us apply Gram-Schmidt on this to get an orthonormal basis for V_0 .

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

Continuation

- ▶ We get the first vector as

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

- ▶ Now take

$$\begin{aligned} w_2 &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \left\langle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}. \end{aligned}$$

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

Continuation

- ▶ Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- ▶ $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .

Continuation

► Now

$$v_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

- $\{v_1, v_2\}$ is an ortho-normal basis for V_0 .
- Given $x \in \mathbb{R}^3$, it decomposes as $y + z$, where $y \in V_0$, $z \in V_0^\perp$.

$$\begin{aligned} y &= \langle v_1, x \rangle v_1 + \langle v_2, x \rangle v_2 \\ &= \frac{x_1 - x_2}{\sqrt{2}} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} + \frac{(x_1 + x_2 - 2x_3)}{\sqrt{6}} \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 2x_1 - x_2 - x_3 \\ -x_1 + 2x_2 - x_3 \\ -x_1 - x_2 + 2x_3 \end{pmatrix} \end{aligned}$$

Continuation

$$\blacktriangleright z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

Continuation

► $z = \frac{1}{3} \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$

► For general n , with $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$,

$$y = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad z = \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

► It is easy to see that $y \in V_0$, $z \in (V_0)^\perp$ and $x = y + z$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.

Projection as a linear map

- **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- **Theorem 13.3:** Under the set up as above,
- (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- (iii) $P(V) = V_0$.

Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.

Projection as a linear map

- ▶ **Definition 13.2:** Let V_0 be a subspace of a finite dimensional inner product space V . Then **the projection on to V_0** , is the map

$$P : V \rightarrow V$$

defined by

$$P(x) = y$$

where $x = y + z$, with $y \in V_0, z \in (V_0)^\perp$.

- ▶ Note that since every $x \in V$ decomposes uniquely as above, P is well-defined. If we want to emphasize the dependence of P on V_0 , we may denote it by P_{V_0} .
- ▶ **Theorem 13.3:** Under the set up as above,
- ▶ (i) P is a linear map. (ii) $Px = x$ if and only if $x \in V_0$ and $Px = 0$ if and only if $x \in (V_0)^\perp$.
- ▶ (iii) $P(V) = V_0$.
- ▶ (iv) $P = P^2 = P^*$.
- ▶ (v) $P_{V_1} = I - P$ where $V_1 = (V_0)^\perp$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

Continuation

- ▶ **Proof.** If $V_0 = \{0\}$ then $P = 0$ and all the properties mentioned above are easy to see.
- ▶ So assume $V_0 \neq \{0\}$.
- ▶ (i). Let $\{v_1, \dots, v_k\}$ be an orthonormal basis of V_0 . Extend it to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of V .
- ▶ Then we know that

$$P(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

(Note that P does not depend upon the choice of this basis!)

- ▶ Since the inner product is linear in the second variable, P is a linear map. This proves (i).

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.

Continuation

- ▶ (ii). We know that $x = \sum_{j=1}^n \langle v_j, x \rangle v_j$. Therefore $Px = x$ implies

$$\sum_{j=k+1}^n \langle v_j, x \rangle v_j = 0.$$

Therefore $x = \sum_{j=1}^k \langle v_j, x \rangle v_j$ and hence $x \in V_0$.

- ▶ The converse is easy to see from the definition of P .
- ▶ Now if $Px = 0$, then $\sum_{j=1}^k \langle v_j, x \rangle v_j = 0$ and hence $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, that is, $x \in (V_0)^\perp$.
- ▶ Conversely if $x \in (V_0)^\perp$, then $x = \sum_{j=k+1}^n \langle v_j, x \rangle v_j$, and consequently $Px = 0$.
- ▶ This proves (ii).

- ▶ (iii). We want to show $P(V) = V_0$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.

Continuation

- ▶ (iii). We want to show $P(V) = V_0$.
- ▶ From the formula given for P , $Px \in V_0$ for every $x \in V$ and hence $P(V) \subseteq V_0$. Since $Px = x$ for every $x \in V_0$, the range of P includes whole of V_0 . This proves (iii).
- ▶ (iv). If $x = \sum_{j=1}^n c_j v_j$, then $Px = \sum_{j=1}^k c_j v_j$.
- ▶ Now $P(P(x)) = P(\sum_{j=1}^k c_j v_j) = \sum_{j=1}^k c_j v_j = Px$.
- ▶ Hence $P^2(x) = P(x)$ for every x , or $P^2 = P$.

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .
- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

Continuation

- ▶ Suppose x_1, x_2 are in V . Let $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ be the unique decompositions of x_1, x_2 so that

$$y_1, y_2 \in V_0; \quad z_1, z_2 \in V_0^\perp.$$

- ▶ Note that $\langle y_i, z_j \rangle = 0$ for all i, j .

- ▶ Now

$$\langle Px_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Similarly,

$$\langle x_1, Px_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle y_1, y_2 \rangle.$$

- ▶ Consequently,

$$\langle Px_1, x_2 \rangle = \langle x_1, Px_2 \rangle$$

for all x_1, x_2 in V .

- ▶ This shows that $P^* = P$ from the defining property of the adjoint of P .

Continuation

- (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

Continuation

- ▶ (v). If $x = \sum_{j=1}^n c_j v_j$,

$$P_{V_0}(x) = \sum_{j=1}^k c_j v_j, \quad P_{V_1}(x) = \sum_{j=k+1}^n c_j v_j.$$

- ▶ From these formulae, it is easy to see that $P_{V_1} = 1 - P_{V_0}$.
- ▶ This completes the proof Theorem 13.2.
- ▶ **Remark 13.4:** Observe that $P_{\{0\}} = 0$ and $P_V = I$. In particular,

$$P_V(x) = x = \sum_{j=1}^n \langle v_j, x \rangle v_j$$

independent of the choice of the basis.

- ▶ We have just revisited our formula for the expansion of x in terms of an orthonormal basis.

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.

Distance between sets

- ▶ **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- ▶ We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- ▶ In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .

Distance between sets

- **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .
- **Example 14.1:** Take $V = \mathbb{R}^2$. Take $a = (1, 0)$. Consider $B_1 = \{(x_1, x_2) : x_1 < 0\}$ and $B_2 = \{(x_1, x_2) : |x_1 - 1| \geq 1\}$

Distance between sets

- ▶ **Notation:** Let A, B be non-empty subsets of an inner product space V and let $a \in V$. Then

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\}$$

and

$$d(a, B) := \inf\{d(a, b) : b \in B\}.$$

- ▶ We may informally call $d(A, B)$ as the distance between A, B and $d(a, B)$ as the distance between a and B . But note that now we may have $d(A, B) = 0$ without having $A = B$.
- ▶ In general, $d(a, B)$ may not be attained at any point in B . Also when it is attained it may not be at some unique point in B .
- ▶ **Example 14.1:** Take $V = \mathbb{R}^2$. Take $a = (1, 0)$. Consider $B_1 = \{(x_1, x_2) : x_1 < 0\}$ and $B_2 = \{(x_1, x_2) : |x_1 - 1| \geq 1\}$
- ▶ Then $d(a, B_1) = 1$ is not attained at any point. $d(a, B_2) = 1$ gets attained at two points.

Best approximation property

- **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

Best approximation property

- ▶ **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

- ▶ This theorem tells us that Px is the unique 'best approximation' for x in V_0 .

Best approximation property

- ▶ **Theorem 14.2:** Let V_0 be a subspace of an inner product space V . Let P be the projection onto V_0 . Then for $x \in V$,

$$d(x, V_0) = d(x, Px).$$

Moreover, Px is the unique point v such that $d(x, v) = d(x, V_0)$.

- ▶ This theorem tells us that Px is the unique 'best approximation' for x in V_0 .
- ▶ **Proof:** Suppose $x = y + z$, is the unique decomposition of x , with $y \in V_0, z \in V_0^\perp$.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|y - v\|^2 + \|z\|^2.\end{aligned}$$

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

- ▶ This proves the theorem.

Continuation

- ▶ We have $Px = y$. Now consider any $v \in V_0$. Due to orthogonality of $y - v$ and z , we get



$$\begin{aligned}\|x - v\|^2 &= \|(y + z) - v\|^2 \\ &= \langle (y - v) + z, (y - v) + z \rangle \\ &= \|(y - v)\|^2 + \|z\|^2.\end{aligned}$$

- ▶ Hence

$$\inf_{v \in V_0} \|x - v\|^2 = \|z\|^2$$

and the infimum is attained only at $v = y$.

- ▶ This proves the theorem.
- ▶ Note that we are using the 'Pythagoras theorem' of inner product spaces.

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.
- ▶

$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.



$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

- ▶ Let P_1 be the projection onto V_1 .

Example

- ▶ Consider the Example 13.1, where $V = \mathbb{R}^n$ and
- ▶ $V_0 = \{x \in V : \sum_{j=1}^n x_j = 0\}$.



$$V_1 = (V_0)^\perp = \left\{ \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} : c \in \mathbb{R} \right\}.$$

- ▶ Let P_1 be the projection onto V_1 .
- ▶ Then $P_1 x = \frac{1}{n}(x_1 + \cdots + x_n) =: \bar{x}$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

- ▶ This value is n times the variance of the tuple $\{x_1, \dots, x_n\}$. In other words,

$$\inf_{c \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (x_j - c)^2 = \text{Var} \{x_1, \dots, x_n\},$$

and the infimum gets attained only at $\bar{x}$.

Continuation

- ▶ So the best approximation for $x = (x_1, \dots, x_n)$ among constant sequences is $(\bar{x}, \dots, \bar{x})$.
- ▶ In other words, we have proved the theorem

$$\inf_{c \in \mathbb{R}} \sum_{j=1}^n (x_j - c)^2 = \sum_{j=1}^n (x_j - \bar{x})^2.$$

- ▶ This value is n times the variance of the tuple $\{x_1, \dots, x_n\}$. In other words,

$$\inf_{c \in \mathbb{R}} \frac{1}{n} \sum_{j=1}^n (x_j - c)^2 = \text{Var} \{x_1, \dots, x_n\},$$

and the infimum gets attained only at $\bar{x}$.

- ▶ **Exercise:** Work out more examples.

Example

► **Example 14.3:** Consider $V = \mathbb{R}^2$. Let

$$V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\} \text{ where } \theta \text{ is a fixed real number.}$$

Write down the matrix of the projection onto V_0 .

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let
- $$V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\} \text{ where } \theta \text{ is a fixed real number.}$$
- Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot (\cos \theta \quad \sin \theta).$$

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let

$$V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\} \text{ where } \theta \text{ is a fixed real number.}$$

Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}.$$

- Then for any vector $x \in \mathbb{R}^2$,

$$Px = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = v \langle v, x \rangle = \langle v, x \rangle \cdot v,$$

Example

- **Example 14.3:** Consider $V = \mathbb{R}^2$. Let
 $V_0 = \left\{ c \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} : c \in \mathbb{R} \right\}$ where θ is a fixed real number.
Write down the matrix of the projection onto V_0 .

- **Solution.** We take

$$P = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}.$$

- Then for any vector $x \in \mathbb{R}^2$,

$$Px = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = v \langle v, x \rangle = \langle v, x \rangle \cdot v,$$

- where

$$v = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}.$$

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

Continuation

- ▶ Observe that $\{v\}$ is an orthonormal basis for V_0 .
- ▶ Therefore P given as above, that is,

$$P = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix},$$

is the projection on to V_0 .

- ▶ You may verify $P = P^2 = P^*$ and $P(\mathbb{R}^2) = V_0$.

Algebraic description of projections

- **Theorem 15.1:** Let V be a finite dimensional inner product space. Then a linear map $P : V \rightarrow V$ is a projection if and only if

$$P = P^2 = P^*.$$

Algebraic description of projections

- ▶ **Theorem 15.1:** Let V be a finite dimensional inner product space. Then a linear map $P : V \rightarrow V$ is a projection if and only if

$$P = P^2 = P^*.$$

- ▶ **Proof:** We have already seen that if P is a projection then $P = P^2 = P^*$.

Algebraic description of projections

- ▶ **Theorem 15.1:** Let V be a finite dimensional inner product space. Then a linear map $P : V \rightarrow V$ is a projection if and only if

$$P = P^2 = P^*.$$

- ▶ **Proof:** We have already seen that if P is a projection then $P = P^2 = P^*$.
- ▶ Now suppose $P : V \rightarrow V$ is a linear map such that

$$P = P^2 = P^*.$$

Algebraic description of projections

- ▶ **Theorem 15.1:** Let V be a finite dimensional inner product space. Then a linear map $P : V \rightarrow V$ is a projection if and only if

$$P = P^2 = P^*.$$

- ▶ **Proof:** We have already seen that if P is a projection then $P = P^2 = P^*$.
- ▶ Now suppose $P : V \rightarrow V$ is a linear map such that

$$P = P^2 = P^*.$$

- ▶ Take $V_0 = P(V) = \{Px : x \in V\}$.

Algebraic description of projections

- ▶ **Theorem 15.1:** Let V be a finite dimensional inner product space. Then a linear map $P : V \rightarrow V$ is a projection if and only if

$$P = P^2 = P^*.$$

- ▶ **Proof:** We have already seen that if P is a projection then $P = P^2 = P^*$.
- ▶ Now suppose $P : V \rightarrow V$ is a linear map such that

$$P = P^2 = P^*.$$

- ▶ Take $V_0 = P(V) = \{Px : x \in V\}$.
- ▶ Clearly V_0 is a subspace of V .

- ▶ We wish to show that P is the projection onto V_0 .

Continuation

- ▶ We wish to show that P is the projection onto V_0 .
- ▶ For $x \in V$, take $y = Px$ and $z = x - Px$.

Continuation

- ▶ We wish to show that P is the projection onto V_0 .
- ▶ For $x \in V$, take $y = Px$ and $z = x - Px$.
- ▶ Clearly $x = y + z$ and $y \in V_0$.

Continuation

- ▶ We wish to show that P is the projection onto V_0 .
- ▶ For $x \in V$, take $y = Px$ and $z = x - Px$.
- ▶ Clearly $x = y + z$ and $y \in V_0$.
- ▶ It suffices to show that $z \in V_0^\perp$.

Continuation

- ▶ Consider any $w \in V_0$. So $w = Pv$ for some $v \in V$.

Continuation

- ▶ Consider any $w \in V_0$. So $w = Pv$ for some $v \in V$.
- ▶ Now

$$\begin{aligned}\langle w, z \rangle &= \langle Pv, x - Px \rangle \\ &= \langle v, P^*(x - Px) \rangle \\ &= \langle v, P(x - Px) \rangle \\ &= \langle v, Px - P^2x \rangle \\ &= \langle v, Px - Px \rangle \\ &= 0.\end{aligned}$$

Continuation

- ▶ Consider any $w \in V_0$. So $w = Pv$ for some $v \in V$.
- ▶ Now

$$\begin{aligned}\langle w, z \rangle &= \langle Pv, x - Px \rangle \\ &= \langle v, P^*(x - Px) \rangle \\ &= \langle v, P(x - Px) \rangle \\ &= \langle v, Px - P^2x \rangle \\ &= \langle v, Px - Px \rangle \\ &= 0.\end{aligned}$$

- ▶ This shows that $z \in V_0^\perp$. ■

Diagonalization of a projection

- **Theorem 15.2:** Let V_0 be a non-zero finite dimensional subspace of a finite dimensional inner product space V and let P the projection onto V_0 . Then there exists an orthonormal basis $\mathcal{B}$ such that on $\mathcal{B}$, the matrix of P is given by

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, with $k = \dim(V_0)$.

Diagonalization of a projection

- ▶ **Theorem 15.2:** Let V_0 be a non-zero finite dimensional subspace of a finite dimensional inner product space V and let P the projection onto V_0 . Then there exists an orthonormal basis $\mathcal{B}$ such that on $\mathcal{B}$, the matrix of P is given by

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, with $k = \dim(V_0)$.

- ▶ **Proof:** Let $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ be an orthonormal basis for V , where $\{v_1, \dots, v_k\}$ is an orthonormal basis for V_0 .

Diagonalization of a projection

- ▶ **Theorem 15.2:** Let V_0 be a non-zero finite dimensional subspace of a finite dimensional inner product space V and let P the projection onto V_0 . Then there exists an orthonormal basis $\mathcal{B}$ such that on $\mathcal{B}$, the matrix of P is given by

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, with $k = \dim(V_0)$.

- ▶ **Proof:** Let $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ be an orthonormal basis for V , where $\{v_1, \dots, v_k\}$ is an orthonormal basis for V_0 .
- ▶ Take $\mathcal{B} = \{v_1, \dots, v_n\}$.

Diagonalization of a projection

- ▶ **Theorem 15.2:** Let V_0 be a non-zero finite dimensional subspace of a finite dimensional inner product space V and let P the projection onto V_0 . Then there exists an orthonormal basis $\mathcal{B}$ such that on $\mathcal{B}$, the matrix of P is given by

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, with $k = \dim(V_0)$.

- ▶ **Proof:** Let $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ be an orthonormal basis for V , where $\{v_1, \dots, v_k\}$ is an orthonormal basis for V_0 .
- ▶ Take $\mathcal{B} = \{v_1, \dots, v_n\}$.
- ▶ We have $Pv_j = v_j$ for $1 \leq j \leq k$ and $Pv_j = 0$ for $(k+1) \leq j \leq n$.

Eigenvalues of projections

- **Theorem 15.3:** Let V be a finite dimensional inner product space and let $P : V \rightarrow V$ be a projection. Suppose λ is an eigenvalue of P then $\lambda \in \{0, 1\}$.

Eigenvalues of projections

- ▶ **Theorem 15.3:** Let V be a finite dimensional inner product space and let $P : V \rightarrow V$ be a projection. Suppose λ is an eigenvalue of P then $\lambda \in \{0, 1\}$.
- ▶ **Proof:** This is clear as the characteristic polynomial of the matrix

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, is given by $p(x) = (x - 1)^k x^{n-k}$. ■

Eigenvalues of projections

- ▶ **Theorem 15.3:** Let V be a finite dimensional inner product space and let $P : V \rightarrow V$ be a projection. Suppose λ is an eigenvalue of P then $\lambda \in \{0, 1\}$.
- ▶ **Proof:** This is clear as the characteristic polynomial of the matrix

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is of order $k \times k$, is given by $p(x) = (x - 1)^k x^{n-k}$. ■

- ▶ Note that 0 is the only eigenvalue of a projection P iff $P = 0$. Similarly 1 is the only eigenvalue of P if and only if $P = I$.

► Consequently

$$\langle v_i, P v_j \rangle = \begin{cases} \delta_{ij} & \text{if } 1 \leq i, j \leq k \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ Consequently

$$\langle v_i, P v_j \rangle = \begin{cases} \delta_{ij} & \text{if } 1 \leq i, j \leq k \\ 0 & \text{otherwise.} \end{cases}$$

- ▶ Hence the matrix of P on the orthonormal basis $\mathcal{B}$ is A as above.

Rank and trace

- **Corollary 15.4:** Let R be the matrix of a projection map on some orthonormal basis. Then $\text{trace}(R) = \text{rank}(R)$.

Rank and trace

- ▶ **Corollary 15.4:** Let R be the matrix of a projection map on some orthonormal basis. Then $\text{trace}(R) = \text{rank}(R)$.
- ▶ **Proof:** Clear from the description of the matrix of P .

Rank and trace

- ▶ **Corollary 15.4:** Let R be the matrix of a projection map on some orthonormal basis. Then $\text{trace}(R) = \text{rank}(R)$.
- ▶ **Proof:** Clear from the description of the matrix of P .
- ▶ **Example 15.5:** Show that $R : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by

$$R \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \end{pmatrix}$$

is a projection. Write down the matrix of R on standard basis. Find a suitable orthonormal basis such that the entries of the matrix of R on that basis are 0's and 1's.

Rank and trace

- ▶ **Corollary 15.4:** Let R be the matrix of a projection map on some orthonormal basis. Then $\text{trace}(R) = \text{rank}(R)$.
- ▶ **Proof:** Clear from the description of the matrix of P .
- ▶ **Example 15.5:** Show that $R : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by

$$R \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \end{pmatrix}$$

is a projection. Write down the matrix of R on standard basis. Find a suitable orthonormal basis such that the entries of the matrix of R on that basis are 0's and 1's.

Rank and trace

- ▶ **Corollary 15.4:** Let R be the matrix of a projection map on some orthonormal basis. Then $\text{trace}(R) = \text{rank}(R)$.
- ▶ **Proof:** Clear from the description of the matrix of P .
- ▶ **Example 15.5:** Show that $R : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by

$$R \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \end{pmatrix}$$

is a projection. Write down the matrix of R on standard basis. Find a suitable orthonormal basis such that the entries of the matrix of R on that basis are 0's and 1's.

- ▶ **END OF LECTURE 15.**

Lecture 16: Eigenvalues and eigenvectors

- **Definition 16.1:** Let A be a real/complex square matrix. Then the characteristic polynomial of A is defined as the polynomial

$$p(x) = \det(xI - A).$$

Lecture 16: Eigenvalues and eigenvectors

- **Definition 16.1:** Let A be a real/complex square matrix. Then the characteristic polynomial of A is defined as the polynomial

$$p(x) = \det(xI - A).$$

- If $A = [a_{ij}]_{1 \leq i, j \leq n}$

$$p(x) = \det \begin{bmatrix} x - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & x - a_{nn} \end{bmatrix}.$$

Lecture 16: Eigenvalues and eigenvectors

- **Definition 16.1:** Let A be a real/complex square matrix. Then the characteristic polynomial of A is defined as the polynomial

$$p(x) = \det(xI - A).$$

- If $A = [a_{ij}]_{1 \leq i, j \leq n}$

$$p(x) = \det \begin{bmatrix} x - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & x - a_{nn} \end{bmatrix}.$$

- Note that the characteristic polynomial of an $n \times n$ matrix is polynomial of degree n . Also its leading coefficient (the coefficient of x^n) is equal to 1. Such polynomials are known as **monic** polynomials.

Example

► Example 16.2: $A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -3 & -8 \\ 0 & 0 & 2i \end{bmatrix}.$

Example

► Example 16.2: $A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -3 & -8 \\ 0 & 0 & 2i \end{bmatrix}$.

► Then the characteristic polynomial of A is given by,

$$\begin{aligned} p(x) &= \det(xI - A) \\ &= \det\left(\begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix} - \begin{bmatrix} 2 & -1 & 0 \\ 0 & -3 & -8 \\ 0 & 0 & 2i \end{bmatrix}\right) \\ &= \det\begin{bmatrix} x-2 & +1 & 0 \\ 0 & x+3 & +8 \\ 0 & 0 & x-2i \end{bmatrix} \\ &= (x-2)(x+3)(x-2i) \\ &= (x^2 + x - 6)(x - 2i) \\ &= x^3 + x^2 - 6x - 2ix^2 - 2ix + 12i \\ &= x^3 + (1 - 2i)x^2 - (6 + 2i)x + 12i \end{aligned}$$

Fundamental theorem of algebra

- **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

Fundamental theorem of algebra

- ▶ **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

- ▶ Note that $\lambda_1, \dots, \lambda_n$ are roots of the polynomial, that is, these are the complex numbers λ such that $p(\lambda) = 0$.

Fundamental theorem of algebra

- ▶ **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

- ▶ Note that $\lambda_1, \dots, \lambda_n$ are roots of the polynomial, that is, these are the complex numbers λ such that $p(\lambda) = 0$.
- ▶ Some of the roots could be repeated. The number of times a particular root λ appears in the factorization is known as the multiplicity of λ . A complex polynomial of degree n has exactly n roots (counting multiplicity).

Fundamental theorem of algebra

- ▶ **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

- ▶ Note that $\lambda_1, \dots, \lambda_n$ are roots of the polynomial, that is, these are the complex numbers λ such that $p(\lambda) = 0$.
- ▶ Some of the roots could be repeated. The number of times a particular root λ appears in the factorization is known as the multiplicity of λ . A complex polynomial of degree n has exactly n roots (counting multiplicity).
- ▶ We will be assuming this theorem without giving proof.

Fundamental theorem of algebra

- ▶ **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

- ▶ Note that $\lambda_1, \dots, \lambda_n$ are roots of the polynomial, that is, these are the complex numbers λ such that $p(\lambda) = 0$.
- ▶ Some of the roots could be repeated. The number of times a particular root λ appears in the factorization is known as the multiplicity of λ . A complex polynomial of degree n has exactly n roots (counting multiplicity).
- ▶ We will be assuming this theorem without giving proof.
- ▶ **Example 16.4:** Consider the polynomial $p(x) = x^2 + 1$.

Fundamental theorem of algebra

- ▶ **Theorem 16.3(Fundamental theorem of algebra):** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial, with $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then p factorizes uniquely (up to permutation) as

$$p(x) = a_n(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$.

- ▶ Note that $\lambda_1, \dots, \lambda_n$ are roots of the polynomial, that is, these are the complex numbers λ such that $p(\lambda) = 0$.
- ▶ Some of the roots could be repeated. The number of times a particular root λ appears in the factorization is known as the multiplicity of λ . A complex polynomial of degree n has exactly n roots (counting multiplicity).
- ▶ We will be assuming this theorem without giving proof.
- ▶ **Example 16.4:** Consider the polynomial $p(x) = x^2 + 1$.
- ▶ We have $p(x) = (x + i)(x - i)$. So the roots of p can be complex even if the coefficients are real.

Eigenvalues and eigenvectors of complex matrices

- **Definition 16.5:** Let A be a complex square matrix. Then the **eigenvalues** of A are defined as roots of the characteristic polynomial of A . The **algebraic multiplicity** of an eigenvalue λ is defined as the multiplicity of λ in the factorization of the characteristic polynomial of A .

Eigenvalues and eigenvectors of complex matrices

- ▶ **Definition 16.5:** Let A be a complex square matrix. Then the **eigenvalues** of A are defined as roots of the characteristic polynomial of A . The **algebraic multiplicity** of an eigenvalue λ is defined as the multiplicity of λ in the factorization of the characteristic polynomial of A .
- ▶ An $n \times n$ matrix A is also viewed as the linear map $x \mapsto Ax$ on $\mathbb{C}^n$.

Eigenvalues and eigenvectors of complex matrices

- ▶ **Definition 16.5:** Let A be a complex square matrix. Then the **eigenvalues** of A are defined as roots of the characteristic polynomial of A . The **algebraic multiplicity** of an eigenvalue λ is defined as the multiplicity of λ in the factorization of the characteristic polynomial of A .
- ▶ An $n \times n$ matrix A is also viewed as the linear map $x \mapsto Ax$ on $\mathbb{C}^n$.
- ▶ **Definition 16.6:** Suppose A is an $n \times n$ complex matrix, and $\lambda \in \mathbb{C}$. If $x \in \mathbb{C}^n$ is a non-zero vector such that $Ax = \lambda x$, then x is said to be an eigenvector with eigenvalue λ .

Eigenvalues and eigenvectors of complex matrices

- ▶ **Definition 16.5:** Let A be a complex square matrix. Then the **eigenvalues** of A are defined as roots of the characteristic polynomial of A . The **algebraic multiplicity** of an eigenvalue λ is defined as the multiplicity of λ in the factorization of the characteristic polynomial of A .
- ▶ An $n \times n$ matrix A is also viewed as the linear map $x \mapsto Ax$ on $\mathbb{C}^n$.
- ▶ **Definition 16.6:** Suppose A is an $n \times n$ complex matrix, and $\lambda \in \mathbb{C}$. If $x \in \mathbb{C}^n$ is a non-zero vector such that $Ax = \lambda x$, then x is said to be an eigenvector with eigenvalue λ .
- ▶ It is to be noted that if $\lambda \in \mathbb{C}$ has an eigenvector x :

$$Ax = \lambda x.$$

This means that $(\lambda I - A)x = 0$. In particular $(\lambda I - A)$ is not injective, therefore $\det(\lambda I - A) = 0$ or $p(\lambda) = 0$ where p is the characteristic polynomial of A . So λ is an eigenvalue.

Geometric multiplicity

- **Definition 16.7:** Let A be an $n \times n$ complex matrix. Then the **geometric multiplicity** of an eigenvalue λ is defined as the dimension of the kernel of $(\lambda I - A)$, that is, the dimension of the **eigen space**:

$$\{x \in \mathbb{C}^n : Ax = \lambda x\}.$$

Geometric multiplicity

- **Definition 16.7:** Let A be an $n \times n$ complex matrix. Then the **geometric multiplicity** of an eigenvalue λ is defined as the dimension of the kernel of $(\lambda I - A)$, that is, the dimension of the **eigen space**:

$$\{x \in \mathbb{C}^n : Ax = \lambda x\}.$$

- **Example 16.8:** Consider $B = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$.

Geometric multiplicity

- **Definition 16.7:** Let A be an $n \times n$ complex matrix. Then the **geometric multiplicity** of an eigenvalue λ is defined as the dimension of the kernel of $(\lambda I - A)$, that is, the dimension of the **eigen space**:

$$\{x \in \mathbb{C}^n : Ax = \lambda x\}.$$

- **Example 16.8:** Consider $B = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$.
- Then the characteristic polynomial of B is given by $p(x) = (x - 5)^3$. Therefore 5 appears with algebraic multiplicity 3.

Geometric multiplicity

- **Definition 16.7:** Let A be an $n \times n$ complex matrix. Then the **geometric multiplicity** of an eigenvalue λ is defined as the dimension of the kernel of $(\lambda I - A)$, that is, the dimension of the **eigen space**:

$$\{x \in \mathbb{C}^n : Ax = \lambda x\}.$$

- **Example 16.8:** Consider $B = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$.
- Then the characteristic polynomial of B is given by $p(x) = (x - 5)^3$. Therefore 5 appears with algebraic multiplicity 3.
- However, if x is an eigenvector with eigenvalue 5, we see

$$\begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 5 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Continuation

► That is,

$$5x_1 + x_2 = 5x_1$$

$$5x_2 + x_3 = 5x_2$$

$$5x_3 = 5x_3$$

Continuation

- ▶ That is,

$$5x_1 + x_2 = 5x_1$$

$$5x_2 + x_3 = 5x_2$$

$$5x_3 = 5x_3$$

- ▶ Solving this, we see

$$\{x : Bx = 5x\} = \left\{ \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} \right\}$$

Therefore the geometric multiplicity of the eigenvalue 5 is 1.

Continuation

- More generally, for $n \geq 2$, and $c \in \mathbb{C}$, the $n \times n$ matrix

$$C = \begin{bmatrix} c & 1 & 0 & 0 & \dots \\ 0 & c & 1 & 0 & \dots \\ 0 & 0 & c & 1 & \dots \\ 0 & 0 & 0 & c & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

Continuation

- ▶ More generally, for $n \geq 2$, and $c \in \mathbb{C}$, the $n \times n$ matrix

$$C = \begin{bmatrix} c & 1 & 0 & 0 & \dots \\ 0 & c & 1 & 0 & \dots \\ 0 & 0 & c & 1 & \dots \\ 0 & 0 & 0 & c & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

- ▶ has characteristic polynomial $(x - c)^n$. However, the geometric multiplicity of the eigenvalue c is just 1.

Comparing two multiplicities

- **Theorem 16.9:** Let A be an $n \times n$ complex matrix and let λ be an eigenvalue of A . Then
$$1 \leq \text{geometric multiplicity of } \lambda \leq \text{algebraic multiplicity of } \lambda.$$

Comparing two multiplicities

- ▶ **Theorem 16.9:** Let A be an $n \times n$ complex matrix and let λ be an eigenvalue of A . Then

$$1 \leq \text{geometric multiplicity of } \lambda \leq \text{algebraic multiplicity of } \lambda.$$

- ▶ **Proof:** We have already seen that whenever λ is an eigenvalue there exists non-zero x such that $Ax = \lambda x$. Hence the geometric multiplicity of λ is at least 1.

Comparing two multiplicities

- ▶ **Theorem 16.9:** Let A be an $n \times n$ complex matrix and let λ be an eigenvalue of A . Then
$$1 \leq \text{geometric multiplicity of } \lambda \leq \text{algebraic multiplicity of } \lambda.$$
- ▶ **Proof:** We have already seen that whenever λ is an eigenvalue there exists non-zero x such that $Ax = \lambda x$. Hence the geometric multiplicity of λ is at least 1.
- ▶ Now suppose the geometric multiplicity of λ is k . Then there exist k linearly independent vectors $\{w_1, w_2, \dots, w_k\}$ such that $Aw_j = \lambda w_j$ for $1 \leq j \leq k$.

Continuation

- ▶ Extend $\{w_1, w_2, \dots, w_k\}$ to a basis $\{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ of $\mathbb{C}^n$.

Continuation

- ▶ Extend $\{w_1, w_2, \dots, w_k\}$ to a basis $\{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ of $\mathbb{C}^n$.
- ▶ Then the matrix of the linear map $x \mapsto Ax$ on this basis has the form:

$$B = \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix}$$

for some $C_{k \times n}, D_{(n-k) \times (n-k)}$, as $Aw_j = \lambda w_j$, for $1 \leq j \leq k$.

Continuation

- ▶ Extend $\{w_1, w_2, \dots, w_k\}$ to a basis $\{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ of $\mathbb{C}^n$.
- ▶ Then the matrix of the linear map $x \mapsto Ax$ on this basis has the form:

$$B = \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix}$$

for some $C_{k \times n}, D_{(n-k) \times (n-k)}$, as $Aw_j = \lambda w_j$, for $1 \leq j \leq k$.

- ▶ Equivalently there exists an invertible matrix S such that $B = S^{-1}AS$.

Continuation

- ▶ Extend $\{w_1, w_2, \dots, w_k\}$ to a basis $\{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ of $\mathbb{C}^n$.
- ▶ Then the matrix of the linear map $x \mapsto Ax$ on this basis has the form:

$$B = \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix}$$

for some $C_{k \times n}, D_{(n-k) \times (n-k)}$, as $Aw_j = \lambda w_j$, for $1 \leq j \leq k$.

- ▶ Equivalently there exists an invertible matrix S such that $B = S^{-1}AS$.
- ▶ Observe that,

$$\begin{aligned} \det(xI - B) &= \det(xI - S^{-1}AS) \\ &= \det(xS^{-1}S - S^{-1}AS) \\ &= \det S^{-1}(xI - A)S \\ &= \det(S^{-1}) \det(xI - A) \det(S) \\ &= \det(xI - A). \end{aligned}$$

- ▶ Hence, the characteristic polynomial of A , has the form

$$\begin{aligned} p(x) &= \det(xI - \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix}) \\ &= \det \begin{bmatrix} x - \lambda I_k & -C \\ 0 & xI - D \end{bmatrix} \\ &= (x - \lambda)^k \cdot \det(xI - D). \end{aligned}$$

- ▶ Hence, the characteristic polynomial of A , has the form

$$\begin{aligned} p(x) &= \det(xI - \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix}) \\ &= \det \begin{bmatrix} x - \lambda I_k & -C \\ 0 & xI - D \end{bmatrix} \\ &= (x - \lambda)^k \cdot \det(xI - D). \end{aligned}$$

- ▶ In particular, the algebraic multiplicity of λ is at least k . ■.

Real case

- ▶ Here also we will consider the eigenvalues of a real matrix as the (possibly complex) roots of the characteristic polynomial. The main issue here is that a polynomial with real coefficients can have complex roots.

Real case

- ▶ Here also we will consider the eigenvalues of a real matrix as the (possibly complex) roots of the characteristic polynomial. The main issue here is that a polynomial with real coefficients can have complex roots.
- ▶ For instance the characteristic polynomial of

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

is $p(x) = x^2 + 1 = (x - i)(x + i)$.

Real case

- ▶ Here also we will consider the eigenvalues of a real matrix as the (possibly complex) roots of the characteristic polynomial. The main issue here is that a polynomial with real coefficients can have complex roots.
- ▶ For instance the characteristic polynomial of

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

is $p(x) = x^2 + 1 = (x - i)(x + i)$.

- ▶ Therefore the eigenvalues of a real matrix can be complex. The algebraic multiplicity would be the multiplicity in the associated characteristic polynomial.

Real case

- ▶ Here also we will consider the eigenvalues of a real matrix as the (possibly complex) roots of the characteristic polynomial. The main issue here is that a polynomial with real coefficients can have complex roots.
- ▶ For instance the characteristic polynomial of

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

is $p(x) = x^2 + 1 = (x - i)(x + i)$.

- ▶ Therefore the eigenvalues of a real matrix can be complex. The algebraic multiplicity would be the multiplicity in the associated characteristic polynomial.
- ▶ However, we consider geometric multiplicity of an eigenvalue λ of a real matrix, considered as a linear map on $\mathbb{R}^n$, as the dimension of

$$\{x \in \mathbb{R}^n : Ax = \lambda x\}$$

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .
- ▶ (i) If $\lambda \notin \mathbb{R}$ then the geometric multiplicity of λ is zero.

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .
- ▶ (i) If $\lambda \notin \mathbb{R}$ then the geometric multiplicity of λ is zero.
- ▶ (ii) If $\lambda \in \mathbb{R}$ then the geometric multiplicity of λ is at least one.

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .
- ▶ (i) If $\lambda \notin \mathbb{R}$ then the geometric multiplicity of λ is zero.
- ▶ (ii) If $\lambda \in \mathbb{R}$ then the geometric multiplicity of λ is at least one.
- ▶ **Proof:** Now for $x \in \mathbb{R}^n$ clearly $Ax \in \mathbb{R}^n$. However as $\lambda \notin \mathbb{R}$ and $0 \neq x \in \mathbb{R}^n$, $\lambda x \notin \mathbb{R}^n$.

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .
- ▶ (i) If $\lambda \notin \mathbb{R}$ then the geometric multiplicity of λ is zero.
- ▶ (ii) If $\lambda \in \mathbb{R}$ then the geometric multiplicity of λ is at least one.
- ▶ **Proof:** Now for $x \in \mathbb{R}^n$ clearly $Ax \in \mathbb{R}^n$. However as $\lambda \notin \mathbb{R}$ and $0 \neq x \in \mathbb{R}^n$, $\lambda x \notin \mathbb{R}^n$.
- ▶ Therefore, $Ax = \lambda x$ is not possible.

Geometric multiplicity of real maps

- ▶ **Theorem 16.10:** Let A be a real matrix considered as a linear map on $\mathbb{R}^n$. Suppose λ is an eigenvalue of A .
- ▶ (i) If $\lambda \notin \mathbb{R}$ then the geometric multiplicity of λ is zero.
- ▶ (ii) If $\lambda \in \mathbb{R}$ then the geometric multiplicity of λ is at least one.
- ▶ **Proof:** Now for $x \in \mathbb{R}^n$ clearly $Ax \in \mathbb{R}^n$. However as $\lambda \notin \mathbb{R}$ and $0 \neq x \in \mathbb{R}^n$, $\lambda x \notin \mathbb{R}^n$.
- ▶ Therefore, $Ax = \lambda x$ is not possible.
- ▶ Hence the geometric multiplicity of non-real eigenvalues of real matrices (considered as real maps) is zero.

- ▶ (ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue.

- ▶ (ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue.
- ▶ Then $\det(\lambda I - A) = 0$. Hence $x \mapsto (\lambda I - A)x$ on $\mathbb{R}^n$ is not injective.

- ▶ (ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue.
- ▶ Then $\det(\lambda I - A) = 0$. Hence $x \mapsto (\lambda I - A)x$ on $\mathbb{R}^n$ is not injective.
- ▶ In particular, there exists non-zero $x \in \mathbb{R}^n$ such that $Ax = \lambda x$.

- ▶ (ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue.
- ▶ Then $\det(\lambda I - A) = 0$. Hence $x \mapsto (\lambda I - A)x$ on $\mathbb{R}^n$ is not injective.
- ▶ In particular, there exists non-zero $x \in \mathbb{R}^n$ such that $Ax = \lambda x$.
- ▶ Therefore, geometric multiplicity of any real eigenvalue of any real matrix is at least one.

Algebraic multiplicities of real matrices

- **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)
- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)
- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

- ▶ Now if λ is an eigenvalue of A , then $p(\lambda) = 0$.

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)
- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

- ▶ Now if λ is an eigenvalue of A , then $p(\lambda) = 0$.
- ▶ Now taking complex conjugate and observing that $\overline{a_j} = a_j$ for every j , we get $p(\bar{\lambda}) = 0$.

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)
- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

- ▶ Now if λ is an eigenvalue of A , then $p(\lambda) = 0$.
- ▶ Now taking complex conjugate and observing that $\overline{a_j} = a_j$ for every j , we get $p(\bar{\lambda}) = 0$.
- ▶ This shows that $(x - \lambda)(x - \bar{\lambda})$ is a factor of p .

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)
- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

- ▶ Now if λ is an eigenvalue of A , then $p(\lambda) = 0$.
- ▶ Now taking complex conjugate and observing that $\overline{a_j} = a_j$ for every j , we get $p(\bar{\lambda}) = 0$.
- ▶ This shows that $(x - \lambda)(x - \bar{\lambda})$ is a factor of p .
- ▶ We write $p(x) = (x - \lambda)(x - \bar{\lambda})q(x)$. Now $(x - (\bar{\lambda} + \lambda)x + |\lambda|^2)$ has only real coefficients. Then by the division algorithm, q also has only real coefficients.

Algebraic multiplicities of real matrices

- ▶ **Theorem 16.11:** Let A be a real square matrix. Then for any complex number λ the algebraic multiplicity of $\bar{\lambda}$ as an eigenvalue of A is same as the algebraic multiplicity of λ . (Eigenvalues appear in conjugate pairs.)

- ▶ **Proof:** Suppose

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

is the characteristic polynomial of A . Then $a_0, \dots, a_{n-1}$ are real numbers.

- ▶ Now if λ is an eigenvalue of A , then $p(\lambda) = 0$.
- ▶ Now taking complex conjugate and observing that $\overline{a_j} = a_j$ for every j , we get $p(\bar{\lambda}) = 0$.
- ▶ This shows that $(x - \lambda)(x - \bar{\lambda})$ is a factor of p .
- ▶ We write $p(x) = (x - \lambda)(x - \bar{\lambda})q(x)$. Now $(x - (\bar{\lambda} + \lambda)x + |\lambda|^2)$ has only real coefficients. Then by the division algorithm, q also has only real coefficients.
- ▶ Now the result follows by simple induction. ■

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.
- ▶ **Proof:** Since the eigenvalues of A appear in conjugate pairs and n is odd it follows that A has a real eigenvalue.

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.
- ▶ **Proof:** Since the eigenvalues of A appear in conjugate pairs and n is odd it follows that A has a real eigenvalue.
- ▶ If λ is a real eigenvalue of a real matrix then we know that its geometric multiplicity is at least one. ■

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.
- ▶ **Proof:** Since the eigenvalues of A appear in conjugate pairs and n is odd it follows that A has a real eigenvalue.
- ▶ If λ is a real eigenvalue of a real matrix then we know that its geometric multiplicity is at least one. ■
- ▶ It maybe noted that existence of a real root can also be got from real analysis:

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.
- ▶ **Proof:** Since the eigenvalues of A appear in conjugate pairs and n is odd it follows that A has a real eigenvalue.
- ▶ If λ is a real eigenvalue of a real matrix then we know that its geometric multiplicity is at least one. ■
- ▶ It maybe noted that existence of a real root can also be got from real analysis:
- ▶ As the characteristic polynomial p is a monic real polynomial of odd degree

$$\lim_{x \rightarrow \infty} p(x) = +\infty; \quad \lim_{x \rightarrow -\infty} p(x) = -\infty$$

and so by intermediate value theorem p has at least one real root.

Continuation

- ▶ **Corollary 16.12:** Suppose A is a real $n \times n$ matrix. If n is odd then A has at least one real eigenvalue with real eigenvector.
- ▶ **Proof:** Since the eigenvalues of A appear in conjugate pairs and n is odd it follows that A has a real eigenvalue.
- ▶ If λ is a real eigenvalue of a real matrix then we know that its geometric multiplicity is at least one. ■
- ▶ It maybe noted that existence of a real root can also be got from real analysis:
- ▶ As the characteristic polynomial p is a monic real polynomial of odd degree

$$\lim_{x \rightarrow \infty} p(x) = +\infty; \quad \lim_{x \rightarrow -\infty} p(x) = -\infty$$

and so by intermediate value theorem p has at least one real root.

- ▶ **END OF LECTURE 16.**

Lecture 17: Diagonalization

- ▶ Consider the sequence:

$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$

Lecture 17: Diagonalization

- ▶ Consider the sequence:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, . . .

- ▶ You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

Lecture 17: Diagonalization

- ▶ Consider the sequence:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, . . .

- ▶ You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

- ▶ How to compute F_{1000} ?

Lecture 17: Diagonalization

- ▶ Consider the sequence:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

- ▶ You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

- ▶ How to compute F_{1000} ?
- ▶ Consider the matrix

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

Continuation

► We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

- ▶ Therefore,

$$A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

- ▶ Therefore,

$$A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}.$$

- ▶ Hence we know F_{1000} if we know A^{999} .

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .
- ▶ Then

$$A = SDS^{-1}.$$

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .

- ▶ Then

$$A = SDS^{-1}.$$

- ▶ This implies $A^2 = SD^2S^{-1}$ and more generally,

$$A^m = SD^mS^{-1}, \quad \forall m \geq 1.$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

► then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

► then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

► and

$$A^m = S \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix} S^{-1}.$$

Continuation

- ▶ Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

- ▶ then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

- ▶ and

$$A^m = S \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix} S^{-1}.$$

- ▶ Hence computing A^m becomes easy.

Continuation

- **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?
- ▶ Let p be the characteristic polynomial of A . From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D . Hence

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?
- ▶ Let p be the characteristic polynomial of A . From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D . Hence

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ In particular the diagonal entries of D must be the eigenvalues of A .

Exmple

► Example 17.2.

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

is not diagonalizable.

Example

- ▶ Example 17.2.

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

is not diagonalizable.

- ▶ The eigenvalues of B are just 0 and 0.

Exmple

► Example 17.2.

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

is not diagonalizable.

- The eigenvalues of B are just 0 and 0.
- So if $B = SDS^{-1}$, the diagonal D must be the zero matrix.

Exmple

► Example 17.2.

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

is not diagonalizable.

- The eigenvalues of B are just 0 and 0.
- So if $B = SDS^{-1}$, the diagonal D must be the zero matrix.
- That would mean that $B = 0$, which is clearly not true. This is a contradiction. Hence B is not diagonalizable.

Linear independence

- ▶ We need to know how eigenvectors of distinct eigenvalues behave.

Linear independence

- ▶ We need to know how eigenvectors of distinct eigenvalues behave.
- ▶ **Theorem 17.3:** Let A be an $n \times n$ complex matrix. Suppose $a_1, \dots, a_k$ are some distinct eigenvalues of A and $w_1, \dots, w_k$ are eigenvectors with

$$Aw_j = a_j w_j, \quad 1 \leq j \leq k.$$

Then $\{w_1, \dots, w_k\}$ is a linearly independent set.

Linear independence

- ▶ We need to know how eigenvectors of distinct eigenvalues behave.
- ▶ **Theorem 17.3:** Let A be an $n \times n$ complex matrix. Suppose $a_1, \dots, a_k$ are some distinct eigenvalues of A and $w_1, \dots, w_k$ are eigenvectors with

$$Aw_j = a_j w_j, \quad 1 \leq j \leq k.$$

Then $\{w_1, \dots, w_k\}$ is a linearly independent set.

- ▶ **Proof:** Suppose $\sum_{j=1}^k c_j w_j = 0$.

Linear independence

- ▶ We need to know how eigenvectors of distinct eigenvalues behave.
- ▶ **Theorem 17.3:** Let A be an $n \times n$ complex matrix. Suppose $a_1, \dots, a_k$ are some distinct eigenvalues of A and $w_1, \dots, w_k$ are eigenvectors with

$$Aw_j = a_j w_j, \quad 1 \leq j \leq k.$$

Then $\{w_1, \dots, w_k\}$ is a linearly independent set.

- ▶ **Proof:** Suppose $\sum_{j=1}^k c_j w_j = 0$.
- ▶ On applying A , we get $\sum_{j=1}^k a_j c_j w_j = 0$.

Linear independence

- ▶ We need to know how eigenvectors of distinct eigenvalues behave.
- ▶ **Theorem 17.3:** Let A be an $n \times n$ complex matrix. Suppose $a_1, \dots, a_k$ are some distinct eigenvalues of A and $w_1, \dots, w_k$ are eigenvectors with

$$Aw_j = a_j w_j, \quad 1 \leq j \leq k.$$

Then $\{w_1, \dots, w_k\}$ is a linearly independent set.

- ▶ **Proof:** Suppose $\sum_{j=1}^k c_j w_j = 0$.
- ▶ On applying A , we get $\sum_{j=1}^k a_j c_j w_j = 0$.
- ▶ By repeated application of A we get

$$\sum_{j=1}^k a_j^s c_j w_j = 0, \quad \forall 1 \leq s \leq (k-1).$$

Continuation

- ▶ Let N be the $n \times k$ matrix formed by taking the vectors $c_1 w_1, \dots, c_k w_k$ as its columns:

$$N = [c_1 w_1, c_2 w_2, \dots, c_k w_k].$$

Continuation

- ▶ Let N be the $n \times k$ matrix formed by taking the vectors $c_1 w_1, \dots, c_k w_k$ as its columns:

$$N = [c_1 w_1, c_2 w_2, \dots, c_k w_k].$$

- ▶ Now we may write the linear equations above as:

$$\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = N \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_k & a_k^2 & \dots & a_k^{k-1} \end{bmatrix}$$

Continuation

- ▶ Let N be the $n \times k$ matrix formed by taking the vectors $c_1 w_1, \dots, c_k w_k$ as its columns:

$$N = [c_1 w_1, c_2 w_2, \dots, c_k w_k].$$

- ▶ Now we may write the linear equations above as:

$$\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = N \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_k & a_k^2 & \dots & a_k^{k-1} \end{bmatrix}$$

- ▶ Equivalently, $0 = NV$, where V is the $k \times k$ Vandermonde matrix formed out of $a_1, \dots, a_k$. Since a_j 's are distinct, V is invertible.

Continuation

- ▶ Let N be the $n \times k$ matrix formed by taking the vectors $c_1 w_1, \dots, c_k w_k$ as its columns:

$$N = [c_1 w_1, c_2 w_2, \dots, c_k w_k].$$

- ▶ Now we may write the linear equations above as:

$$\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = N \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_k & a_k^2 & \dots & a_k^{k-1} \end{bmatrix}$$

- ▶ Equivalently, $0 = NV$, where V is the $k \times k$ Vandermonde matrix formed out of $a_1, \dots, a_k$. Since a_j 's are distinct, V is invertible.
- ▶ Multiplying by V^{-1} from the right, we get $N = 0$ or $c_j w_j = 0$ for every j . Since w_j 's are assumed to be non-zero (they are eigenvectors), we get $c_j = 0$ for every j . ■

Continuation

- ▶ Let N be the $n \times k$ matrix formed by taking the vectors $c_1 w_1, \dots, c_k w_k$ as its columns:

$$N = [c_1 w_1, c_2 w_2, \dots, c_k w_k].$$

- ▶ Now we may write the linear equations above as:

$$\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = N \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_k & a_k^2 & \dots & a_k^{k-1} \end{bmatrix}$$

- ▶ Equivalently, $0 = NV$, where V is the $k \times k$ Vandermonde matrix formed out of $a_1, \dots, a_k$. Since a_j 's are distinct, V is invertible.
- ▶ Multiplying by V^{-1} from the right, we get $N = 0$ or $c_j w_j = 0$ for every j . Since w_j 's are assumed to be non-zero (they are eigenvectors), we get $c_j = 0$ for every j . ■
- ▶ **Challenge:** Find a different proof of this result.

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .

Continuation

- **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

- ▶ Note that the invertibility of S is equivalent to requiring that its columns $\{v_1, v_2, \dots, v_n\}$ forms a basis of $\mathbb{C}^n$.

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

- ▶ Note that the invertibility of S is equivalent to requiring that its columns $\{v_1, v_2, \dots, v_n\}$ forms a basis of $\mathbb{C}^n$.
- ▶ This proves $(i) \Leftrightarrow (ii)$.

Continuation

- ▶ (i) and (ii) $\Rightarrow$ (iii). From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D and hence the eigenvalues of A are $d_1, \dots, d_n$ (including multiplicity).

Continuation

- ▶ (i) and (ii) $\Rightarrow$ (iii). From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D and hence the eigenvalues of A are $d_1, \dots, d_n$ (including multiplicity).
- ▶ Suppose $a_1, \dots, a_k$ are distinct eigenvalues of A and r_j is the algebraic multiplicity of a_j , $1 \leq j \leq k$.

Continuation

- ▶ (i) and (ii) $\Rightarrow$ (iii). From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D and hence the eigenvalues of A are $d_1, \dots, d_n$ (including multiplicity).
- ▶ Suppose $a_1, \dots, a_k$ are distinct eigenvalues of A and r_j is the algebraic multiplicity of a_j , $1 \leq j \leq k$.
- ▶ Then taking a suitable permutation if necessary, we may assume that

$$(d_1, d_2, \dots, d_n) = (a_1, a_1, \dots, a_1, a_2, \dots, a_2, a_3, \dots, a_k)$$

where a_j appears r_j times, $1 \leq j \leq k$ and
 $r_1 + r_2 + \dots + r_k = n$.

- ▶ From $Av_j = d_j v_j$ and the fact that $\{v_1, \dots, v_n\}$ are linearly independent, we see that the geometric multiplicity of a_j is at least r_j .

Continuation

- ▶ From $Av_j = d_j v_j$ and the fact that $\{v_1, \dots, v_n\}$ are linearly independent, we see that the geometric multiplicity of a_j is at least r_j .
- ▶ It can't be more than r_j as we have proved that the geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity.

Continuation

- ▶ From $Av_j = d_j v_j$ and the fact that $\{v_1, \dots, v_n\}$ are linearly independent, we see that the geometric multiplicity of a_j is at least r_j .
- ▶ It can't be more than r_j as we have proved that the geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity.
- ▶ Therefore for every j , the geometric and algebraic multiplicity of a_j is r_j . This proves (iii).

- ▶ (iii) $\Rightarrow$ (ii). Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A and let the geometric/algebraic multiplicity of a_j be r_j .

Continuation

- ▶ (iii) $\Rightarrow$ (ii). Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A and let the geometric/algebraic multiplicity of a_j be r_j .
- ▶ By the fundamental theorem of algebra we know that $r_1 + \dots + r_k = n$.

Continuation

- ▶ (iii) $\Rightarrow$ (ii). Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A and let the geometric/algebraic multiplicity of a_j be r_j .
- ▶ By the fundamental theorem of algebra we know that $r_1 + \dots + r_k = n$.
- ▶ Let $\{v_{j1}, v_{j2}, \dots, v_{jr_j}\}$ be a basis for the eigenspace of A with eigenvalue a_j . In particular, for every j , $\{v_{j1}, v_{j2}, \dots, v_{jr_j}\}$ are linearly independent.

Continuation

- ▶ (iii) $\Rightarrow$ (ii). Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A and let the geometric/algebraic multiplicity of a_j be r_j .
- ▶ By the fundamental theorem of algebra we know that $r_1 + \dots + r_k = n$.
- ▶ Let $\{v_{j1}, v_{j2}, \dots, v_{jr_j}\}$ be a basis for the eigenspace of A with eigenvalue a_j . In particular, for every j , $\{v_{j1}, v_{j2}, \dots, v_{jr_j}\}$ are linearly independent.
- ▶ We also have $Av_{ji} = a_j v_{ji}$ for $1 \leq i \leq r_j$.

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.
- ▶ Note that $\sum_{j=1}^k w_j = 0$.

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.
- ▶ Note that $\sum_{j=1}^k w_j = 0$.
- ▶ Also $a_1, a_2, \dots, a_k$ are distinct and $Aw_j = a_j w_j, 1 \leq j \leq k$.

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.
- ▶ Note that $\sum_{j=1}^k w_j = 0$.
- ▶ Also $a_1, a_2, \dots, a_k$ are distinct and $Aw_j = a_j w_j, 1 \leq j \leq k$.
- ▶ Then by the previous theorem, $w_j = 0$ for every j .

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.
- ▶ Note that $\sum_{j=1}^k w_j = 0$.
- ▶ Also $a_1, a_2, \dots, a_k$ are distinct and $Aw_j = a_j w_j, 1 \leq j \leq k$.
- ▶ Then by the previous theorem, $w_j = 0$ for every j .
- ▶ For fixed j , by the linear independence of v_{ji} 's, we get $c_{ji} = 0$ for all i .

Continuation

- ▶ We obtain (ii) if we show that the whole collection

$$\{v_{ji} : 1 \leq i \leq r_j; 1 \leq j \leq k\}$$

is linearly independent.

- ▶ Suppose

$$\sum_{j=1}^k \sum_{i=1}^{r_j} c_{ji} v_{ji} = 0$$

- ▶ Take $w_j = \sum_{i=1}^{r_j} c_{ji} v_{ji}$.
- ▶ Note that $\sum_{j=1}^k w_j = 0$.
- ▶ Also $a_1, a_2, \dots, a_k$ are distinct and $Aw_j = a_j w_j, 1 \leq j \leq k$.
- ▶ Then by the previous theorem, $w_j = 0$ for every j .
- ▶ For fixed j , by the linear independence of v_{ji} 's, we get $c_{ji} = 0$ for all i .
- ▶ This proves the required linear independence. ■

Application

- Obtain a formula for Fibonacci number F_n by diagonalizing

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

and getting a formula for A^n .

Application

- ▶ Obtain a formula for Fibonacci number F_n by diagonalizing

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

and getting a formula for A^n .

- ▶ **END OF LECTURE 17.**

Lecture 18: Linear Recurrence relations

- ▶ Recall: Consider the sequence:

$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$

Lecture 18: Linear Recurrence relations

- Recall: Consider the sequence:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

- You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

Lecture 18: Linear Recurrence relations

- ▶ Recall: Consider the sequence:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, . . .

- ▶ You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

- ▶ How to compute F_{1000} ?

Lecture 18: Linear Recurrence relations

- ▶ Recall: Consider the sequence:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

- ▶ You may observe that this is the well-known **Fibonacci sequence** defined by: $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2.$$

- ▶ How to compute F_{1000} ?
- ▶ Consider the matrix

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

Continuation

► We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

- ▶ Therefore,

$$A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}.$$

Continuation

- ▶ We have:

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- ▶ Further,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- ▶ More generally,

$$A \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}.$$

- ▶ Therefore,

$$A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}.$$

- ▶ Hence we know F_{1000} if we know A^{999} .

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .
- ▶ Then

$$A = SDS^{-1}.$$

How to compute the power?

- ▶ This raises the question as to how to efficiently compute powers of a matrix?
- ▶ **Observation:** Suppose there exists an invertible matrix S such that $S^{-1}AS = D$ for some diagonal matrix D .

- ▶ Then

$$A = SDS^{-1}.$$

- ▶ This implies $A^2 = SD^2S^{-1}$ and more generally,

$$A^m = SD^mS^{-1}, \quad \forall m \geq 1.$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

► then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

Continuation

► Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

► then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

► and

$$A^m = S \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix} S^{-1}.$$

Continuation

- ▶ Now if

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

- ▶ then

$$D^m = \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix},$$

- ▶ and

$$A^m = S \begin{bmatrix} d_1^m & 0 & \dots & 0 \\ 0 & d_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^m \end{bmatrix} S^{-1}.$$

- ▶ Hence computing A^m becomes easy.

Continuation

- **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?
- ▶ Let p be the characteristic polynomial of A . From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D . Hence

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Continuation

- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ Let us try to understand diagonalizability.
- ▶ Suppose $A = SDS^{-1}$ with D diagonal. What can be the diagonal entries?
- ▶ Let p be the characteristic polynomial of A . From $A = SDS^{-1}$, we know that the characteristic polynomial of A is same as that of D . Hence

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ In particular the diagonal entries of D must be the eigenvalues of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .

Continuation

- **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

- ▶ Note that the invertibility of S is equivalent to requiring that its columns $\{v_1, v_2, \dots, v_n\}$ forms a basis of $\mathbb{C}^n$.

Continuation

- ▶ **Proof:** $(i) \Leftrightarrow (ii)$. The equation $A = SDS^{-1}$ is same as $AS = SD$. Let $v_1, v_2, \dots, v_n$ be the columns of S .
- ▶ Now $AS = SD$ means:

$$A[v_1, v_2, \dots, v_n] = [v_1, \dots, v_n]D.$$

- ▶ Equivalently,

$$[Av_1, Av_2, \dots, Av_n] = [d_1 v_1, d_2 v_2, \dots, d_n v_n].$$

- ▶ This is same as:

$$Av_j = d_j v_j, 1 \leq j \leq n.$$

In particular columns of S are eigenvectors of A .

- ▶ Note that the invertibility of S is equivalent to requiring that its columns $\{v_1, v_2, \dots, v_n\}$ forms a basis of $\mathbb{C}^n$.
- ▶ This proves $(i) \Leftrightarrow (ii)$.

Summary

- ▶ We understand that to diagonalize a matrix we need to find eigenvalues and eigenvectors of the matrix and we need to get a basis consisting of eigenvectors.

Summary

- ▶ We understand that to diagonalize a matrix we need to find eigenvalues and eigenvectors of the matrix and we need to get a basis consisting of eigenvectors.
- ▶ We now apply this idea to solve some linear recurrence relations.

Linear recurrence relations

- ▶ Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

Linear recurrence relations

- ▶ Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- ▶ We want to get a formula for a_n .

Linear recurrence relations

- ▶ Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- ▶ We want to get a formula for a_n .
- ▶ Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

Linear recurrence relations

- ▶ Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- ▶ We want to get a formula for a_n .
- ▶ Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

- ▶ We have

$$A \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}.$$

Linear recurrence relations

- Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- We want to get a formula for a_n .
- Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

- We have

$$A \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}.$$

- Therefore,

$$\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = A^{n-1} \begin{pmatrix} v_1 \\ v_0 \end{pmatrix}.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

► The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

► The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

► Solving $p(x) = 0$, we get the eigenvalues of A as

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

- The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

- Solving $p(x) = 0$, we get the eigenvalues of A as

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

- **Case I:** $\alpha \neq \beta$, that is, $b^2 + 4c \neq 0$.

Continuation

- By direct computation we obtain eigenvectors

$$\begin{pmatrix} \alpha \\ 1 \end{pmatrix}, \begin{pmatrix} \beta \\ 1 \end{pmatrix}$$

with eigenvalues α, β respectively.

- ▶ By direct computation we obtain eigenvectors

$$\begin{pmatrix} \alpha \\ 1 \end{pmatrix}, \begin{pmatrix} \beta \\ 1 \end{pmatrix}$$

with eigenvalues α, β respectively.

- ▶ Take

$$S = \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix}.$$

Continuation

- By direct computation we obtain eigenvectors

$$\begin{pmatrix} \alpha \\ 1 \end{pmatrix}, \begin{pmatrix} \beta \\ 1 \end{pmatrix}$$

with eigenvalues α, β respectively.

- Take

$$S = \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix}.$$

- Then

$$S^{-1} = \frac{1}{\alpha - \beta} \begin{bmatrix} 1 & -\beta \\ -1 & \alpha \end{bmatrix}.$$

Continuation

- From $A = SDS^{-1}$, we have

$$A = \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \cdot \frac{1}{\alpha - \beta} \begin{bmatrix} 1 & -\beta \\ -1 & \alpha \end{bmatrix}.$$

Continuation

- From $A = SDS^{-1}$, we have

$$A = \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \cdot \frac{1}{\alpha - \beta} \begin{bmatrix} 1 & -\beta \\ -1 & \alpha \end{bmatrix}.$$

- Hence for $n \geq 1$,

$$\begin{aligned} A^{n-1} &= \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha^{n-1} & 0 \\ 0 & \beta^{n-1} \end{bmatrix} \cdot \frac{1}{\alpha - \beta} \begin{bmatrix} 1 & -\beta \\ -1 & \alpha \end{bmatrix} \\ &= \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & \beta \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha^{n-1} & -\alpha^{n-1}\beta \\ -\beta^{n-1} & \alpha\beta^{n-1} \end{bmatrix} \\ &= \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha^n - \beta^n & \alpha\beta^n - \alpha^n\beta \\ \alpha^{n-1} - \beta^{n-1} & \alpha\beta^{n-1} - \alpha^{n-1}\beta \end{bmatrix}. \end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}a_n &= \frac{1}{\alpha - \beta}[(\alpha^n - \beta^n)v_1 + (\alpha\beta^n - \alpha^n\beta)v_0] \\&= \frac{1}{\alpha - \beta}[(v_1 - \beta v_0)\alpha^n + (\alpha v_0 - v_1)\beta^n].\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}a_n &= \frac{1}{\alpha - \beta}[(\alpha^n - \beta^n)v_1 + (\alpha\beta^n - \alpha^n\beta)v_0] \\&= \frac{1}{\alpha - \beta}[(v_1 - \beta v_0)\alpha^n + (\alpha v_0 - v_1)\beta^n].\end{aligned}$$

► Recall

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

Continuation

► Therefore,

$$\begin{aligned}a_n &= \frac{1}{\alpha - \beta}[(\alpha^n - \beta^n)v_1 + (\alpha\beta^n - \alpha^n\beta)v_0] \\&= \frac{1}{\alpha - \beta}[(v_1 - \beta v_0)\alpha^n + (\alpha v_0 - v_1)\beta^n].\end{aligned}$$

► Recall

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

► Therefore, $\alpha - \beta = \sqrt{b^2 + 4c}$.

Continuation

- Therefore,

$$\begin{aligned}a_n &= \frac{1}{\alpha - \beta}[(\alpha^n - \beta^n)v_1 + (\alpha\beta^n - \alpha^n\beta)v_0] \\&= \frac{1}{\alpha - \beta}[(v_1 - \beta v_0)\alpha^n + (\alpha v_0 - v_1)\beta^n].\end{aligned}$$

- Recall

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

- Therefore, $\alpha - \beta = \sqrt{b^2 + 4c}$.
- In particular, a_n has the form

$$a_n = s\alpha^n + t\beta^n$$

for some scalars s, t , where α, β are the two distinct roots of $x^2 - bx - c = 0$.

Fibonacci Sequence

- ▶ Consider $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2$$

Fibonacci Sequence

- ▶ Consider $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2$$

- ▶ So $v_0 = 0, v_1 = 1, b = c = 1$.

Fibonacci Sequence

- ▶ Consider $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2$$

- ▶ So $v_0 = 0, v_1 = 1, b = c = 1$.

- ▶ Hence

$$\alpha = \frac{1 + \sqrt{5}}{2}, \quad \beta = \frac{1 - \sqrt{5}}{2}$$

and $\alpha - \beta = \sqrt{5}$.

Fibonacci Sequence

- ▶ Consider $F_0 = 0, F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2}, \quad \forall n \geq 2$$

- ▶ So $v_0 = 0, v_1 = 1, b = c = 1$.

- ▶ Hence

$$\alpha = \frac{1 + \sqrt{5}}{2}, \quad \beta = \frac{1 - \sqrt{5}}{2}$$

and $\alpha - \beta = \sqrt{5}$.

- ▶ Therefore,

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right], \quad \forall n \geq 0.$$

Example

- ▶ Suppose $a_0 = 0$, $a_1 = 1$ and $a_n = a_{n-1} - a_{n-2}$, $\forall n \geq 2$.

Example

- ▶ Suppose $a_0 = 0, a_1 = 1$ and $a_n = a_{n-1} - a_{n-2}, \forall n \geq 2$.
- ▶ Here $b = 1, c = -1$. The roots α, β are not real as $b^2 + 4c < 0$.

Example

- ▶ Suppose $a_0 = 0, a_1 = 1$ and $a_n = a_{n-1} - a_{n-2}, \forall n \geq 2$.
- ▶ Here $b = 1, c = -1$. The roots α, β are not real as $b^2 + 4c < 0$.
- ▶ So the formula for a_n has the form

$$a_n = s\left(\frac{1 + \sqrt{3}i}{2}\right)^n + t\left(\frac{1 - \sqrt{3}i}{2}\right)^n.$$

Example

- ▶ Suppose $a_0 = 0, a_1 = 1$ and $a_n = a_{n-1} - a_{n-2}, \forall n \geq 2$.
- ▶ Here $b = 1, c = -1$. The roots α, β are not real as $b^2 + 4c < 0$.
- ▶ So the formula for a_n has the form

$$a_n = s\left(\frac{1 + \sqrt{3}i}{2}\right)^n + t\left(\frac{1 - \sqrt{3}i}{2}\right)^n.$$

- ▶ Though all the terms of the sequence are real, the formula for a_n requires complex terms!

Going further

- ▶ Case (ii): $\alpha = \beta$.

Going further

- ▶ Case (ii): $\alpha = \beta$.
- ▶ Here our matrix A is not diagonalizable. So our method fails!

Going further

- ▶ Case (ii): $\alpha = \beta$.
- ▶ Here our matrix A is not diagonalizable. So our method fails!
- ▶ It is still possible to get a formula for a_n . This is a challenge for you.

Going further

- ▶ Case (ii): $\alpha = \beta$.
- ▶ Here our matrix A is not diagonalizable. So our method fails!
- ▶ It is still possible to get a formula for a_n . This is a challenge for you.
- ▶ What about higher order recurrence relations such as **tribonacci numbers**:

0, 0, 1, 1, 2, 4, 7, 13, 24, 44, 81, 149, 274, ...

Going further

- ▶ Case (ii): $\alpha = \beta$.
- ▶ Here our matrix A is not diagonalizable. So our method fails!
- ▶ It is still possible to get a formula for a_n . This is a challenge for you.
- ▶ What about higher order recurrence relations such as **tribonacci numbers**:

$$0, 0, 1, 1, 2, 4, 7, 13, 24, 44, 81, 149, 274, \dots$$

- ▶ Here

$$t_0 = t_1 = 0, t_2 = 1, t_n = t_{n-1} + t_{n-2} + t_{n-3}, \quad \forall n \geq 3.$$

Continuation

- ▶ Sequences like Fibonacci sequences were studied earlier by Pingala and other Indian mathematicians.

Continuation

- ▶ Sequences like Fibonacci sequences were studied earlier by Pingala and other Indian mathematicians.
- ▶ Here is an excerpt from the Wikipedia:

Continuation

- ▶ Sequences like Fibonacci sequences were studied earlier by Pingala and other Indian mathematicians.
- ▶ Here is an excerpt from the Wikipedia:
- ▶ The Narayana's cows sequence is generated by the recurrence

$$N(k) = N(k - 1) + N(k - 3)$$

.

Continuation

- ▶ Sequences like Fibonacci sequences were studied earlier by Pingala and other Indian mathematicians.
- ▶ Here is an excerpt from the Wikipedia:
- ▶ The Narayana's cows sequence is generated by the recurrence

$$N(k) = N(k - 1) + N(k - 3)$$

.

- ▶ A random Fibonacci sequence can be defined by tossing a coin for each position n of the sequence and taking $F(n) = F(n - 1) + F(n - 2)$ if it lands heads and $F(n) = F(n - 1) - F(n - 2)$ if it lands tails. Work by Furstenberg and Kesten guarantees that this sequence almost surely grows exponentially at a constant rate:
- ▶ The constant is independent of the coin tosses and was computed in 1999 by Divakar Viswanath. It is now known as Viswanath's constant.

Continuation

- ▶ Sequences like Fibonacci sequences were studied earlier by Pingala and other Indian mathematicians.
- ▶ Here is an excerpt from the Wikipedia:
- ▶ The Narayana's cows sequence is generated by the recurrence

$$N(k) = N(k - 1) + N(k - 3)$$

- ▶ .
- ▶ A random Fibonacci sequence can be defined by tossing a coin for each position n of the sequence and taking $F(n) = F(n - 1) + F(n - 2)$ if it lands heads and $F(n) = F(n - 1) - F(n - 2)$ if it lands tails. Work by Furstenberg and Kesten guarantees that this sequence almost surely grows exponentially at a constant rate:
- ▶ The constant is independent of the coin tosses and was computed in 1999 by Divakar Viswanath. It is now known as Viswanath's constant.
- ▶ **END OF LECTURE 18.**

Lecture 19: Schur's upper triangularization theorem

► Recall:

Lecture 19: Schur's upper triangularization theorem

- ▶ Recall:
- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

Lecture 19: Schur's upper triangularization theorem

- ▶ Recall:
- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ The diagonal entries of D are eigenvalues of A and columns of S are corresponding eigenvectors.

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .
- ▶ There are matrices which are not diagonalizable. The next best would be to make the matrix 'triangular'.

Upper and lower triangular matrices

- **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq j < i \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq j < i \leq n.$$

- ▶ A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **lower triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq i < j \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \text{ for } 1 \leq j < i \leq n.$$

- ▶ A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **lower triangular** if

$$t_{ij} = 0, \text{ for } 1 \leq i < j \leq n.$$

- ▶ Upper triangular:

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}.$$

- ▶ Note that products of upper triangular matrices are upper triangular. If a matrix is both upper triangular and lower triangular then it is diagonal.

Upper triangularization

- **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

Upper triangularization

- **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Proof:** We will prove this by induction on n .

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Proof:** We will prove this by induction on n .
- ▶ For $n = 1$ there is nothing to prove, as every 1×1 matrix is upper triangular, we can take U as the 1×1 identity matrix.

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Proof:** We will prove this by induction on n .
- ▶ For $n = 1$ there is nothing to prove, as every 1×1 matrix is upper triangular, we can take U as the 1×1 identity matrix.
- ▶ Now take $n \geq 2$ and assume the result for all $(n - 1) \times (n - 1)$ matrices.

Continuation

- ▶ Consider $A = [a_{ij}]_{1 \leq i, j \leq n}$.

Continuation

- ▶ Consider $A = [a_{ij}]_{1 \leq i, j \leq n}$.
- ▶ Let a_1 be some eigenvalue of A .

Continuation

- ▶ Consider $A = [a_{ij}]_{1 \leq i, j \leq n}$.
- ▶ Let a_1 be some eigenvalue of A .
- ▶ Let v_1 be an eigenvector of A with eigenvalue a_1 . (Since $\det(A - a_1 I) = 0$, $A - a_1 I$ is singular and so there exists a non-zero vector v_1 such that $(A - a_1 I)v_1 = 0$.)

Continuation

- ▶ Consider $A = [a_{ij}]_{1 \leq i, j \leq n}$.
- ▶ Let a_1 be some eigenvalue of A .
- ▶ Let v_1 be an eigenvector of A with eigenvalue a_1 . (Since $\det(A - a_1 I) = 0$, $A - a_1 I$ is singular and so there exists a non-zero vector v_1 such that $(A - a_1 I)v_1 = 0$.)
- ▶ By dividing v_1 by its norm if necessary, we may assume that v_1 is a unit vector.

Continuation

- ▶ Extend $\{v_1\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$.
(Such an extension exists due to Gram-Schmidt orthogonalization process.)

Continuation

- ▶ Extend $\{v_1\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$. (Such an extension exists due to Gram-Schmidt orthogonalization process.)
- ▶ We have $Av_1 = a_1 v_1$ and for every j , expanding Av_j using the basis $\{v_1, \dots, v_n\}$,

$$Av_j = \sum_{i=1}^n \langle v_i, Av_j \rangle v_i.$$

Continuation

- ▶ Extend $\{v_1\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$. (Such an extension exists due to Gram-Schmidt orthogonalization process.)
- ▶ We have $Av_1 = a_1 v_1$ and for every j , expanding Av_j using the basis $\{v_1, \dots, v_n\}$,

$$Av_j = \sum_{i=1}^n \langle v_i, Av_j \rangle v_i.$$

- ▶ Let V be the matrix $V = [v_1, v_2, \dots, v_n]$. Then these linear equations can be written as:

$$AV = VS$$

where $S = [s_{ij}]$ is the matrix defined by

$$s_{ij} = \langle v_i, Av_j \rangle.$$

Continuation

- ▶ Extend $\{v_1\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$. (Such an extension exists due to Gram-Schmidt orthogonalization process.)
- ▶ We have $Av_1 = a_1 v_1$ and for every j , expanding Av_j using the basis $\{v_1, \dots, v_n\}$,

$$Av_j = \sum_{i=1}^n \langle v_i, Av_j \rangle v_i.$$

- ▶ Let V be the matrix $V = [v_1, v_2, \dots, v_n]$. Then these linear equations can be written as:

$$AV = VS$$

where $S = [s_{ij}]$ is the matrix defined by

$$s_{ij} = \langle v_i, Av_j \rangle.$$

- ▶ In other words, S is the matrix of the linear map $x \mapsto Ax$, on the basis $\{v_1, \dots, v_n\}$.

Continuation

- ▶ We have $AV = VS$.

Continuation

- ▶ We have $AV = VS$.
- ▶ Since columns of V form an orthonormal basis, V is a unitary, that is, $V^*V = VV^* = I$.

Continuation

- ▶ We have $AV = VS$.
- ▶ Since columns of V form an orthonormal basis, V is a unitary, that is, $V^*V = VV^* = I$.
- ▶ Hence we get,

$$A = VSV^*.$$

Continuation

- ▶ We have $AV = VS$.
- ▶ Since columns of V form an orthonormal basis, V is a unitary, that is, $V^*V = VV^* = I$.
- ▶ Hence we get,

$$A = VSV^*.$$

- ▶ Note that since $Av_1 = a_1v_1$ and $s_{ij} = \langle v_i, Av_j \rangle$, the matrix S is of the form:

$$S = \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix}$$

for some $1 \times (n-1)$ vector y and $(n-1) \times (n-1)$ matrix B .

Continuation

- ▶ We have $AV = VS$.
- ▶ Since columns of V form an orthonormal basis, V is a unitary, that is, $V^*V = VV^* = I$.
- ▶ Hence we get,

$$A = VSV^*.$$

- ▶ Note that since $Av_1 = a_1v_1$ and $s_{ij} = \langle v_i, Av_j \rangle$, the matrix S is of the form:

$$S = \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix}$$

for some $1 \times (n-1)$ vector y and $(n-1) \times (n-1)$ matrix B .

- ▶ By induction hypothesis, there exists an $(n-1) \times (n-1)$ unitary matrix U_1 and an upper triangular matrix T_1 such that

$$B = U_1 T_1 U_1^*.$$

Continuation

► So we get

$$\begin{aligned} A &= VSV^* \\ &= V \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix} V^* \\ &= V \begin{bmatrix} 1 & y \\ 0 & U_1 T_1 U_1^* \end{bmatrix} V^* \end{aligned}$$

Continuation

- So we get

$$\begin{aligned} A &= VSV^* \\ &= V \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix} V^* \\ &= V \begin{bmatrix} 1 & y \\ 0 & U_1 T_1 U_1^* \end{bmatrix} V^* \end{aligned}$$

- We may re-write this as:

$$A = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} a_1 & z \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} V^* = UTU^*,$$

Continuation

- So we get

$$\begin{aligned} A &= VSV^* \\ &= V \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix} V^* \\ &= V \begin{bmatrix} 1 & y \\ 0 & U_1 T_1 U_1^* \end{bmatrix} V^* \end{aligned}$$

- We may re-write this as:

$$A = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} a_1 & z \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} V^* = UTU^*,$$

- where $U = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}$ and $z = yU_1$.
- Now U being a product of two unitaries is a unitary and

Continuation

- So we get

$$\begin{aligned} A &= VSV^* \\ &= V \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix} V^* \\ &= V \begin{bmatrix} 1 & y \\ 0 & U_1 T_1 U_1^* \end{bmatrix} V^* \end{aligned}$$

- We may re-write this as:

$$A = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} a_1 & z \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} V^* = UTU^*,$$

- where $U = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}$ and $z = yU_1$.
- Now U being a product of two unitaries is a unitary and
- $\begin{bmatrix} a_1 & y \\ 0 & T_1 \end{bmatrix}$ with T_1 upper triangular is upper triangular.

Continuation

- So we get

$$\begin{aligned} A &= VSV^* \\ &= V \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix} V^* \\ &= V \begin{bmatrix} 1 & y \\ 0 & U_1 T_1 U_1^* \end{bmatrix} V^* \end{aligned}$$

- We may re-write this as:

$$A = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} a_1 & z \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} V^* = UTU^*,$$

- where $U = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}$ and $z = yU_1$.
- Now U being a product of two unitaries is a unitary and
- $\begin{bmatrix} a_1 & y \\ 0 & T_1 \end{bmatrix}$ with T_1 upper triangular is upper triangular.
- This completes the proof.

Diagonal entries

- ▶ **Remark 19.2:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .

Diagonal entries

- ▶ **Remark 19.2:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .
- ▶ As A and T are similar they have same characteristic polynomial.

Diagonal entries

- ▶ **Remark 19.2:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .
- ▶ AS A and T are similar they have same characteristic polynomial.
- ▶ The second part follows as determinant of any upper triangular matrix is product of its diagonal entries and hence

$$\det(I - A) = \det(xI - T) = (x - t_{11})(x - t_{22}) \cdots (x - t_{nn}).$$

Linear recurrence relations

- **Recall:** Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

Linear recurrence relations

- **Recall:** Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- We want to get a formula for a_n .

Linear recurrence relations

- **Recall:** Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- We want to get a formula for a_n .
- Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

Linear recurrence relations

- **Recall:** Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- We want to get a formula for a_n .
- Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

- We have

$$A \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}.$$

Linear recurrence relations

- **Recall:** Suppose $a_0, a_1, \dots, a_n, \dots$ is a sequence of real/complex numbers defined by

$$a_0 = v_0, a_1 = v_1$$

and

$$a_n = ba_{n-1} + ca_{n-2}, \quad \forall n \geq 2$$

where v_0, v_1, b, c are some complex numbers.

- We want to get a formula for a_n .
- Take

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

- We have

$$A \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}.$$

- Therefore,

$$\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = A^{n-1} \begin{pmatrix} v_1 \\ v_0 \end{pmatrix}.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

Eigenvalues and eigenvectors

- ▶ $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

- ▶ The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

► The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

► Solving $p(x) = 0$, we get the eigenvalues of A as

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

- The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

- Solving $p(x) = 0$, we get the eigenvalues of A as

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

- **Case I:** $\alpha \neq \beta$, that is, $b^2 + 4c \neq 0$. We have solved this case by diagonalization.

Eigenvalues and eigenvectors

► $A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$

- The characteristic polynomial of A is:

$$p(x) = (x - b)x - c = x^2 - bx - c.$$

- Solving $p(x) = 0$, we get the eigenvalues of A as

$$\alpha = \frac{b + \sqrt{b^2 + 4c}}{2}, \quad \beta = \frac{b - \sqrt{b^2 + 4c}}{2}.$$

- **Case I:** $\alpha \neq \beta$, that is, $b^2 + 4c \neq 0$. We have solved this case by diagonalization.
- **Case (ii):** $b^2 + 4c = 0$. So the two roots are equal to $\frac{b}{2}$.

Linear recurrence relation with repeated roots

- Consider the matrix

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

where $b^2 + 4c = 0$ and so the eigenvalues of A are $\frac{b}{2}$ and $\frac{b}{2}$.

Linear recurrence relation with repeated roots

- ▶ Consider the matrix

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

where $b^2 + 4c = 0$ and so the eigenvalues of A are $\frac{b}{2}$ and $\frac{b}{2}$.

- ▶ Now

$$\begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}$$

is an eigenvector with eigenvalue $\frac{b}{2}$. Up to scaling this is the only eigenvector.

Linear recurrence relation with repeated roots

- ▶ Consider the matrix

$$A = \begin{bmatrix} b & c \\ 1 & 0 \end{bmatrix}.$$

where $b^2 + 4c = 0$ and so the eigenvalues of A are $\frac{b}{2}$ and $\frac{b}{2}$.

- ▶ Now

$$\begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}$$

is an eigenvector with eigenvalue $\frac{b}{2}$. Up to scaling this is the only eigenvector.

- ▶ Further

$$\begin{pmatrix} -1 \\ \frac{b}{2} \end{pmatrix}$$

is a vector orthogonal to

$$\begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}.$$

Continuation

- ▶ Normalizing these vectors we get an orthonormal basis $\{u_1, u_2\}$ for $\mathbb{C}^2$ where

$$u_1 = \frac{1}{d} \begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}, \quad u_2 = \frac{1}{d} \begin{pmatrix} -1 \\ \bar{\frac{b}{2}} \end{pmatrix}$$

with $d = \sqrt{\frac{|b|^2}{4} + 1}$.

Continuation

- ▶ Normalizing these vectors we get an orthonormal basis $\{u_1, u_2\}$ for $\mathbb{C}^2$ where

$$u_1 = \frac{1}{d} \begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}, \quad u_2 = \frac{1}{d} \begin{pmatrix} -1 \\ \bar{\frac{b}{2}} \end{pmatrix}$$

with $d = \sqrt{\frac{|b|^2}{4} + 1}$.

- ▶ It follows that

$$A = UTU^*$$

for a unitary U with T upper-triangular.

Continuation

- ▶ Normalizing these vectors we get an orthonormal basis $\{u_1, u_2\}$ for $\mathbb{C}^2$ where

$$u_1 = \frac{1}{d} \begin{pmatrix} \frac{b}{2} \\ 1 \end{pmatrix}, \quad u_2 = \frac{1}{d} \begin{pmatrix} -1 \\ \bar{\frac{b}{2}} \end{pmatrix}$$

with $d = \sqrt{\frac{|b|^2}{4} + 1}$.

- ▶ It follows that

$$A = UTU^*$$

for a unitary U with T upper-triangular.

- ▶ By comparing eigenvalues of A and T ,

$$T = \begin{bmatrix} \frac{b}{2} & p \\ 0 & \frac{b}{2} \end{bmatrix}$$

for some p .

Continuation

- It is easy to see from induction that

$$T^n = \begin{bmatrix} (\frac{b}{2})^n & np(\frac{b}{2})^{n-1} \\ 0 & (\frac{b}{2})^n \end{bmatrix}$$

Continuation

- ▶ It is easy to see from induction that

$$T^n = \begin{bmatrix} (\frac{b}{2})^n & np(\frac{b}{2})^{n-1} \\ 0 & (\frac{b}{2})^n \end{bmatrix}$$

- ▶ Now the recurrence relations yields

$$a_n = s(\frac{b}{2})^n + tn(\frac{b}{2})^n, \quad \forall n \geq 0,$$

for some scalars s, t . (Do the necessary matrix computations to verify this.)

Continuation

- ▶ It is easy to see from induction that

$$T^n = \begin{bmatrix} (\frac{b}{2})^n & np(\frac{b}{2})^{n-1} \\ 0 & (\frac{b}{2})^n \end{bmatrix}$$

- ▶ Now the recurrence relations yields

$$a_n = s(\frac{b}{2})^n + tn(\frac{b}{2})^n, \quad \forall n \geq 0,$$

for some scalars s, t . (Do the necessary matrix computations to verify this.)

- ▶ The scalars can be determined using the initial conditions $a_0 = v_0$ and $a_1 = v_1$.

Example

- Example 19.3: Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

Example

- ▶ **Example 19.3:** Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

- ▶ The initial few terms are:

$$1, 2, 3, 0, -27, \dots$$

Example

- ▶ **Example 19.3:** Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

- ▶ The initial few terms are:

$$1, 2, 3, 0, -27, \dots$$

- ▶ We have $b^2 + 4c = 6^2 - 4 \cdot 9 = 0$. Hence the two roots are 3 and 3.

Example

- ▶ **Example 19.3:** Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

- ▶ The initial few terms are:

$$1, 2, 3, 0, -27, \dots$$

- ▶ We have $b^2 + 4c = 6^2 - 4 \cdot 9 = 0$. Hence the two roots are 3 and 3.
- ▶ So we must have

$$a_n = s3^n + tn3^n, \quad \forall n \geq 0.$$

Example

- ▶ **Example 19.3:** Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

- ▶ The initial few terms are:

$$1, 2, 3, 0, -27, \dots$$

- ▶ We have $b^2 + 4c = 6^2 - 4 \cdot 9 = 0$. Hence the two roots are 3 and 3.
- ▶ So we must have

$$a_n = s3^n + tn3^n, \quad \forall n \geq 0.$$

- ▶ Taking $n = 0, 1$ we get

$$s = 1, s \cdot 3 + t \cdot 3 = 2.$$

Hence $t = -\frac{1}{3}$.

Example

- ▶ **Example 19.3:** Suppose $a_0 = 1$, $a_1 = 2$ and

$$a_n = 6a_{n-1} - 9a_{n-2}, \quad n \geq 2.$$

- ▶ The initial few terms are:

$$1, 2, 3, 0, -27, \dots$$

- ▶ We have $b^2 + 4c = 6^2 - 4 \cdot 9 = 0$. Hence the two roots are 3 and 3.
- ▶ So we must have

$$a_n = s3^n + tn3^n, \quad \forall n \geq 0.$$

- ▶ Taking $n = 0, 1$ we get

$$s = 1, s \cdot 3 + t \cdot 3 = 2.$$

Hence $t = -\frac{1}{3}$.

- ▶ Therefore

$$a_n = 3^n \left(1 - \frac{n}{3}\right), \quad \forall n \geq 0.$$

- ▶ **END OF LECTURE 19.**

Lecture 20: Normal matrices

► Recall:

Lecture 20: Normal matrices

- ▶ Recall:
- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

Lecture 20: Normal matrices

- ▶ Recall:
- ▶ **Definition 17.1:** A matrix A is said to be **diagonalizable** if there exists an invertible matrix S and a diagonal matrix D such that that

$$A = SDS^{-1}.$$

- ▶ The diagonal entries of D are eigenvalues of A and columns of S are corresponding eigenvectors.

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .

Diagonalizability

- ▶ **Theorem 17.4:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is diagonalizable, that is, there exists an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}.$$

- ▶ (ii) There exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ (iii) The geometric multiplicity is same as the algebraic multiplicity for every eigenvalue of A .
- ▶ There are matrices which are not diagonalizable. The next best would be to make the matrix 'triangular'.

Upper and lower triangular matrices

- **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq j < i \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq j < i \leq n.$$

- ▶ A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **lower triangular** if

$$t_{ij} = 0, \quad \text{for } 1 \leq i < j \leq n.$$

Upper and lower triangular matrices

- ▶ **Definition 19.1:** A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **upper triangular** if

$$t_{ij} = 0, \text{ for } 1 \leq j < i \leq n.$$

- ▶ A matrix $T = [t_{ij}]_{1 \leq i, j \leq n}$ is said to be **lower triangular** if

$$t_{ij} = 0, \text{ for } 1 \leq i < j \leq n.$$

- ▶ Upper triangular:

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}.$$

- ▶ Note that products of upper triangular matrices are upper triangular. If a matrix is both upper triangular and lower triangular then it is diagonal.

Upper triangularization

- **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Remark 19.3:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Remark 19.3:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .
- ▶ As A and T are similar they have same characteristic polynomial.

Upper triangularization

- ▶ **Theorem 19.2 (Schur's upper triangularization theorem):** Let A be an $n \times n$ complex matrix. Then there exists a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ **Remark 19.3:** Suppose A is an $n \times n$ matrix, U is a unitary and T is an $n \times n$ upper triangular matrix such that $A = UTU^*$. Then the characteristic polynomials of A and T are same. Further, diagonal entries of T are eigenvalues of A .
- ▶ As A and T are similar they have same characteristic polynomial.
- ▶ The second part follows as determinant of any upper triangular matrix is product of its diagonal entries and hence

$$\det(xI - A) = \det(xI - T) = (x - t_{11})(x - t_{22}) \cdots (x - t_{nn}).$$

Motivation to study normal matrices

- ▶ One disadvantage with upper triangularization in comparison with diagonalization is that though products of upper triangular matrices is upper triangular there is no simple method to compute powers of upper triangular matrices.

Motivation to study normal matrices

- ▶ One disadvantage with upper triangularization in comparison with diagonalization is that though products of upper triangular matrices is upper triangular there is no simple method to compute powers of upper triangular matrices.
- ▶ We have seen that computations can be made easily with diagonal matrices.

Motivation to study normal matrices

- ▶ One disadvantage with upper triangularization in comparison with diagonalization is that though products of upper triangular matrices is upper triangular there is no simple method to compute powers of upper triangular matrices.
- ▶ We have seen that computations can be made easily with diagonal matrices.
- ▶ However, in general it is very difficult to check diagonalizability.

Motivation to study normal matrices

- ▶ One disadvantage with upper triangularization in comparison with diagonalization is that though products of upper triangular matrices is upper triangular there is no simple method to compute powers of upper triangular matrices.
- ▶ We have seen that computations can be made easily with diagonal matrices.
- ▶ However, in general it is very difficult to check diagonalizability.
- ▶ So we focus on a large class of matrices called normal matrices. Normality is easy to check and it ensures diagonalizability.

Self-adjoint and normal matrices

- **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 1 \end{bmatrix}.$$

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 1 \end{bmatrix}.$$

- ▶ Note that diagonal entry of every self-adjoint matrix is real as $\overline{a_{ii}} = a_{ii}$ for every i .

Examples of normal matrices

- Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.
- ▶ Example 20.3: Consider

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

Then C is not normal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

- ▶ We have

$$T^* = \begin{bmatrix} \overline{t_{11}} & 0 & 0 & \dots & 0 \\ \overline{t_{12}} & \overline{t_{22}} & 0 & \dots & 0 \\ \overline{t_{13}} & \overline{t_{23}} & \overline{t_{33}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{t_{1n}} & \overline{t_{2n}} & \overline{t_{3n}} & \dots & \overline{t_{nn}} \end{bmatrix}$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

- ▶ So we get

$$T = \begin{bmatrix} t_{11} & 0 & 0 & \cdots & 0 \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & 0 & t_{33} & \cdots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_{nn} \end{bmatrix}$$

Continuation

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0$, $\forall i \neq j$.

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0$, $\forall i \neq j$.
- ▶ In other words, T is diagonal. ■

Unitary equivalence

- **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).
- ▶ **Proof:** Suppose U is a unitary such that $B = UAU^*$. Then $B^*B = (UAU^*)^*(UAU^*) = UA^*UU^*AU = UA^*AU^*$. Similarly, $BB^* = UAA^*U^*$. Now the result follows easily.

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that A and T are unitarily equivalent. Consequently T is normal. Then by Theorem 20.4, as T is both upper triangular and normal it must be diagonal. Taking $D = T$, we have $A = UDU^*$ and we are done.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.
- ▶ Then as A is unitarily equivalent to D , A is also normal. ■.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.
- ▶ Then as A is unitarily equivalent to D , A is also normal. ■.
- ▶ END OF LECTURE 20

Lecture 21: More on Normal matrices

- ▶ Recall: We recall some definitions and the spectral theorem for normal matrices.

Self-adjoint and normal matrices

- **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 1 \end{bmatrix}.$$

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3 + 5i \\ 3 - 5i & 1 \end{bmatrix}.$$

- ▶ Note that diagonal entry of every self-adjoint matrix is real as $\overline{a_{ii}} = a_{ii}$ for every i .

Examples of normal matrices

- Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.
- ▶ Example 20.3: Consider

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

Then C is not normal.

Normal upper triangular matrices

- **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

- ▶ We have

$$T^* = \begin{bmatrix} \overline{t_{11}} & 0 & 0 & \dots & 0 \\ \overline{t_{12}} & \overline{t_{22}} & 0 & \dots & 0 \\ \overline{t_{13}} & \overline{t_{23}} & \overline{t_{33}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{t_{1n}} & \overline{t_{2n}} & \overline{t_{3n}} & \dots & \overline{t_{nn}} \end{bmatrix}$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

- ▶ So we get

$$T = \begin{bmatrix} t_{11} & 0 & 0 & \cdots & 0 \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & 0 & t_{33} & \cdots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_{nn} \end{bmatrix}$$

Continuation

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0$, $\forall i \neq j$.

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0$, $\forall i \neq j$.
- ▶ In other words, T is diagonal. ■

Unitary equivalence

- **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).
- ▶ **Proof:** Suppose U is a unitary such that $B = UAU^*$. Then $B^*B = (UAU^*)^*(UAU^*) = UA^*UU^*AU = UA^*AU^*$. Similarly, $BB^* = UAA^*U^*$. Now the result follows easily.

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that A and T are unitarily equivalent. Consequently T is normal. Then by Theorem 20.4, as T is both upper triangular and normal it must be diagonal. Taking $D = T$, we have $A = UDU^*$ and we are done.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.
- ▶ Then as A is unitarily equivalent to D , A is also normal. ■.

Consequences of the spectral theorem

- **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.
- ▶ Conversely, suppose $\{v_1, \dots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A , say $Av_j = d_j v_j, 1 \leq j \leq n$.

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.
- ▶ Conversely, suppose $\{v_1, \dots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A , say $Av_j = d_j v_j$, $1 \leq j \leq n$.
- ▶ Take $U = [v_1, \dots, v_n]$. Then U is a unitary and $AU = UD$. Hence $A = UDU^*$. Consequently A is normal. ■

Example

► Example 21.2: Suppose

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

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.
- ▶ Solving corresponding eigen equations we see that

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

- ▶ Alternatively,

$$A = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} U^*.$$

Terminology and notation

- **Definition 21.3:** Let A be a complex square matrix. Then

$$\sigma(A) = \{c \in \mathbb{C} : (cI - A) \text{ is not invertible}\}$$

is known as the **spectrum** of A .

Terminology and notation

- **Definition 21.3:** Let A be a complex square matrix. Then

$$\sigma(A) = \{c \in \mathbb{C} : (cI - A) \text{ is not invertible}\}$$

is known as the **spectrum** of A .

- Note that for a matrix A , if $a_1, a_2, \dots, a_n$ are eigenvalues of A , then

$$\sigma(A) = \{a_1, a_2, \dots, a_n\}$$

Some characterizations

- **Theorem 21.4:** Let A be a normal matrix. Then,
- (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

- ▶ If A is self-adjoint, then $A = A^*$. Hence,

$$UDU^* = UD^*U^*.$$

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

- ▶ If A is self-adjoint, then $A = A^*$. Hence,

$$UDU^* = UD^*U^*.$$

- ▶ Multiplication by U^* , U , yields $D = D^*$. Since D is diagonal, this means that all the diagonal entries are real. Hence

$$\sigma(A) \subset \mathbb{R}.$$

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

- ▶ However, clearly A is not self-adjoint.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

- ▶ However, clearly A is not self-adjoint.
- ▶ Similarly (ii) and (iii) of this theorem do not hold without the assumption of normality.

Another characterization of a normal matrix

- **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.
- ▶ Polarization identity yields,

$$\langle x, A^*Ay \rangle = \langle x, AA^*y \rangle, \quad \forall x, y \in \mathbb{C}^n.$$

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.
- ▶ Polarization identity yields,

$$\langle x, A^*Ay \rangle = \langle x, AA^*y \rangle, \quad \forall x, y \in \mathbb{C}^n.$$

- ▶ Hence $A^*A = AA^*$. ■

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.
- ▶ Polarization identity yields,

$$\langle x, A^*Ay \rangle = \langle x, AA^*y \rangle, \quad \forall x, y \in \mathbb{C}^n.$$

- ▶ Hence $A^*A = AA^*$. ■

- ▶ **END OF LECTURE 21**

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

- ▶ In particular, $\|(A - cI)x\| = 0$ if and only if $\|(A - cI)^*x\| = 0$.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

- ▶ In particular, $\|(A - cI)x\| = 0$ if and only if $\|(A - cI)^*x\| = 0$.
- ▶ As $(A - cI)^* = A^* - \bar{c}I$ this shows that x is an eigenvector for A with eigenvalue c if and only if it is an eigenvector for A^* with eigenvalue $\bar{c}$.

- ▶ **Exercise 21.7:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.

Exercises

- ▶ **Exercise 21.7:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 21.8:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 21.9:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.

Exercises

- ▶ **Exercise 21.7:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 21.8:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 21.9:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.
- ▶ **Exercise 21.10:** Suppose A is a normal matrix. Then show that a matrix B commutes with A if and only if it commutes with A^* .

Exercises

- ▶ **Exercise 21.7:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 21.8:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 21.9:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.
- ▶ **Exercise 21.10:** Suppose A is a normal matrix. Then show that a matrix B commutes with A if and only if it commutes with A^* .

Exercises

- ▶ **Exercise 21.7:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 21.8:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 21.9:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.
- ▶ **Exercise 21.10:** Suppose A is a normal matrix. Then show that a matrix B commutes with A if and only if it commutes with A^* .
- ▶ **END OF LECTURE 21**

Lecture 22: Path counting

- ▶ Recall: We recall some definitions and the spectral theorem for normal matrices.

Self-adjoint and normal matrices

- **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 1 \end{bmatrix}.$$

Self-adjoint and normal matrices

- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 1 \end{bmatrix}.$$

- ▶ Note that diagonal entry of every self-adjoint matrix is real as $\overline{a_{ii}} = a_{ii}$ for every i .

Examples of normal matrices

- Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.

Examples of normal matrices

- ▶ Remark 20.2: (i) If A is self-adjoint then A is normal as

$$A^*A = A.A = AA^*.$$

- ▶ (ii) If A is a unitary then A is normal as

$$AA^* = A^*A = I.$$

- ▶ (iii) If A is a projection then $A = A^* = A^2$. Hence A is self-adjoint and normal.
- ▶ (iv) Every diagonal matrix is normal. Every real diagonal matrix is self-adjoint.
- ▶ Example 20.3: Consider

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

Then C is not normal.

Normal upper triangular matrices

- **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

Normal upper triangular matrices

- ▶ **Theorem 20.4:** If a square matrix is upper triangular and normal then it is diagonal.
- ▶ **Proof:** Suppose

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & t_{33} & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{nn} \end{bmatrix}$$

is normal.

- ▶ We have

$$T^* = \begin{bmatrix} \overline{t_{11}} & 0 & 0 & \dots & 0 \\ \overline{t_{12}} & \overline{t_{22}} & 0 & \dots & 0 \\ \overline{t_{13}} & \overline{t_{23}} & \overline{t_{33}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{t_{1n}} & \overline{t_{2n}} & \overline{t_{3n}} & \dots & \overline{t_{nn}} \end{bmatrix}$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

Continuation

- ▶ Computing the first diagonal entries of T^*T and TT^* , as T is normal, we get

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2.$$

- ▶ Hence $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$ or

$$t_{12} = t_{13} = \cdots = t_{1n} = 0.$$

- ▶ So we get

$$T = \begin{bmatrix} t_{11} & 0 & 0 & \cdots & 0 \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & 0 & t_{33} & \cdots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_{nn} \end{bmatrix}$$

Continuation

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0$, $\forall i \neq j$.

Continuation

- ▶ Now from $(T^*T)_{22} = (TT^*)_{22}$ we get

$$|t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2.$$

- ▶ Consequently,

$$t_{23} = t_{24} = \cdots = t_{2n} = 0.$$

- ▶ Continuing this way (that is, by mathematical induction) we see that $t_{ij} = 0, \forall i \neq j$.
- ▶ In other words, T is diagonal. ■

Unitary equivalence

- **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).

Unitary equivalence

- ▶ **Definition 20.5:** A complex matrix B is said to be **unitarily equivalent** to a matrix A if there exists a unitary matrix U such that

$$B = UAU^*.$$

- ▶ It is easy to see that unitary equivalence is an equivalence relation in the space of matrices.
- ▶ B is unitarily equivalent to A means that A, B are matrices of same linear map on different orthonormal bases.
- ▶ **Proposition 20.6:** Suppose B is unitarily equivalent to A . Then B is normal (resp. self-adjoint, unitary, projection) if and only if A is normal (resp. self-adjoint, unitary, projection).
- ▶ **Proof:** Suppose U is a unitary such that $B = UAU^*$. Then $B^*B = (UAU^*)^*(UAU^*) = UA^*UU^*AU = UA^*AU^*$. Similarly, $BB^* = UAA^*U^*$. Now the result follows easily.

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ **Proof:** Suppose A is a normal matrix.
- ▶ By Schur's upper triangularization theorem (Theorem 19.2) there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that A and T are unitarily equivalent. Consequently T is normal. Then by Theorem 20.4, as T is both upper triangular and normal it must be diagonal. Taking $D = T$, we have $A = UDU^*$ and we are done.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.

Continuation

- ▶ Conversely, suppose $A = UDU^*$ where D is diagonal and U is a unitary.
- ▶ Since every diagonal matrix is normal, D is normal.
- ▶ Then as A is unitarily equivalent to D , A is also normal. ■.

Consequences of the spectral theorem

- **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.
- ▶ Conversely, suppose $\{v_1, \dots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A , say $Av_j = d_j v_j, 1 \leq j \leq n$.

Consequences of the spectral theorem

- ▶ **Corollary 21.1:** Let A be an $n \times n$ complex matrix. Then A is normal if and only if there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of A .
- ▶ **Proof:** Suppose A is normal. Then by the spectral theorem there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*.$$

- ▶ Writing this as $AU = UD$, we see that columns of U are eigenvectors for A . Since U is a unitary its columns form an orthonormal basis for $\mathbb{C}^n$.
- ▶ Conversely, suppose $\{v_1, \dots, v_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A , say $Av_j = d_j v_j$, $1 \leq j \leq n$.
- ▶ Take $U = [v_1, \dots, v_n]$. Then U is a unitary and $AU = UD$. Hence $A = UDU^*$. Consequently A is normal. ■

Example

► Example 21.2: Suppose

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

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.

Example

- ▶ Example 21.2: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.
- ▶ Solving corresponding eigen equations we see that

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

- ▶ Alternatively,

$$A = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} U^*.$$

Terminology and notation

- **Definition 21.3:** Let A be a complex square matrix. Then

$$\sigma(A) = \{c \in \mathbb{C} : (cI - A) \text{ is not invertible}\}$$

is known as the **spectrum** of A .

Terminology and notation

- **Definition 21.3:** Let A be a complex square matrix. Then

$$\sigma(A) = \{c \in \mathbb{C} : (cI - A) \text{ is not invertible}\}$$

is known as the **spectrum** of A .

- Note that for a matrix A , if $a_1, a_2, \dots, a_n$ are eigenvalues of A , then

$$\sigma(A) = \{a_1, a_2, \dots, a_n\}$$

Some characterizations

- **Theorem 21.4:** Let A be a normal matrix. Then,
- (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

- ▶ If A is self-adjoint, then $A = A^*$. Hence,

$$UDU^* = UD^*U^*.$$

Some characterizations

- ▶ **Theorem 21.4:** Let A be a normal matrix. Then,
 - (i) A is self-adjoint iff $\sigma(A) \subset \mathbb{R}$.
 - (ii) A is a unitary iff $\sigma(A) \subset \{z \in \mathbb{C} : |z| = 1\}$.
 - (iii) A is a projection iff $\sigma(A) \subseteq \{0, 1\}$.
- ▶ **Proof:** By spectral theorem

$$A = UDU^*$$

for some unitary U and a diagonal matrix D . By taking adjoint,

$$A^* = (UDU^*)^* = UD^*U^*.$$

- ▶ If A is self-adjoint, then $A = A^*$. Hence,

$$UDU^* = UD^*U^*.$$

- ▶ Multiplication by U^* , U , yields $D = D^*$. Since D is diagonal, this means that all the diagonal entries are real. Hence

$$\sigma(A) \subset \mathbb{R}.$$

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

- ▶ However, clearly A is not self-adjoint.

Continuation

- ▶ Conversely, if the eigenvalues of A are real, then the diagonal entries of D are real and hence $D = D^*$. Now $A = A^*$ is clear.
- ▶ Properties (ii) and (iii) also follow easily from the spectral theorem. ■
- ▶ **Remark 21.5:** In this theorem the hypothesis of A being normal is crucial. Just from knowing that eigenvalues of A are real one can't say that it is self-adjoint.
- ▶ For instance,

$$A = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

has $\{3, 7\}$ as its spectrum, which is a subset of the real line.

- ▶ However, clearly A is not self-adjoint.
- ▶ Similarly (ii) and (iii) of this theorem do not hold without the assumption of normality.

Another characterization of a normal matrix

- **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.
- ▶ Polarization identity yields,

$$\langle x, A^*Ay \rangle = \langle x, AA^*y \rangle, \quad \forall x, y \in \mathbb{C}^n.$$

Another characterization of a normal matrix

- ▶ **Theorem 21.6:** Let A be an $n \times n$ complex matrix. Then A is normal iff $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.
- ▶ **Proof:** Suppose A is normal. Then for $x \in \mathbb{C}^n$,

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle x, A^*Ax \rangle \\ &= \langle x, AA^*x \rangle \\ &= \langle A^*x, A^*x \rangle \\ &= \|A^*x\|^2.\end{aligned}$$

- ▶ Hence $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Conversely, suppose $\|Ax\| = \|A^*x\|$, $\forall x \in \mathbb{C}^n$.
- ▶ Then $\langle x, A^*Ax \rangle = \langle x, AA^*x \rangle$, $\forall x \in \mathbb{C}^n$.
- ▶ Polarization identity yields,

$$\langle x, A^*Ay \rangle = \langle x, AA^*y \rangle, \quad \forall x, y \in \mathbb{C}^n.$$

- ▶ Hence $A^*A = AA^*$. ■

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

- ▶ In particular, $\|(A - cI)x\| = 0$ if and only if $\|(A - cI)^*x\| = 0$.

Adjoint of a normal matrix

- ▶ Suppose A is a normal matrix. Then from the definition of normality, A^* is also a normal matrix.
- ▶ If $A = UDU^*$ is a diagonalization of A with unitary U , then $A^* = UD^*U^*$. In particular d is an eigenvalue of A if and only if $\bar{d}$ is an eigenvalue of A^* .
- ▶ For any $c \in \mathbb{C}$, $A - cI$ is also normal.
- ▶ Therefore

$$\|(A - cI)x\| = \|(A - cI)^*x\|, \quad \forall x \in \mathbb{C}^n.$$

- ▶ In particular, $\|(A - cI)x\| = 0$ if and only if $\|(A - cI)^*x\| = 0$.
- ▶ As $(A - cI)^* = A^* - \bar{c}I$ this shows that x is an eigenvector for A with eigenvalue c if and only if it is an eigenvector for A^* with eigenvalue $\bar{c}$.

- ▶ **Exercise 22.1:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.

Exercises

- ▶ **Exercise 22.1:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 22.2:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 22.3:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.

Exercises

- ▶ **Exercise 22.1:** Show that eigenvectors of a normal matrix corresponding to distinct eigenvalues are mutually orthogonal.
- ▶ **Exercise 22.2:** Let A be a normal matrix and q is a polynomial. Show that $q(A)$ is normal.
- ▶ **Exercise 22.3:** Suppose A is a normal matrix. Show that there exists a polynomial p (depending upon A) such that $A^* = p(A)$. Show that this result may not hold without the assumption that A is normal.
- ▶ **Exercise 22.4:** Suppose A is a normal matrix. Then show that a matrix B commutes with A if and only if it commutes with A^* .

Path counting

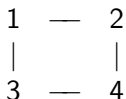
- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)

Path counting

- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)
- ▶ Given two vertices in the graph we wish to count the number of paths from one vertex to another with fixed number of steps.

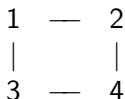
Path counting

- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)
- ▶ Given two vertices in the graph we wish to count the number of paths from one vertex to another with fixed number of steps.
- ▶ **Example 22.5:** Consider a 'square graph' with vertices $\{1, 2, 3, 4\}$ as in the picture:



Path counting

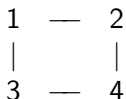
- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)
- ▶ Given two vertices in the graph we wish to count the number of paths from one vertex to another with fixed number of steps.
- ▶ **Example 22.5:** Consider a 'square graph' with vertices $\{1, 2, 3, 4\}$ as in the picture:



- ▶ We want to go from vertex 1 to 4 in **four** steps.

Path counting

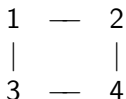
- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)
- ▶ Given two vertices in the graph we wish to count the number of paths from one vertex to another with fixed number of steps.
- ▶ **Example 22.5:** Consider a 'square graph' with vertices $\{1, 2, 3, 4\}$ as in the picture:



- ▶ We want to go from vertex 1 to 4 in **four** steps.
- ▶ Examples of such paths:

Path counting

- ▶ Consider a graph G . A graph consists of vertices and some edges between them. (A formal definition will be given later.)
- ▶ Given two vertices in the graph we wish to count the number of paths from one vertex to another with fixed number of steps.
- ▶ **Example 22.5:** Consider a 'square graph' with vertices $\{1, 2, 3, 4\}$ as in the picture:



- ▶ We want to go from vertex 1 to 4 in **four** steps.
- ▶ Examples of such paths:
- ▶ $1 \rightarrow 2 \rightarrow 4 \rightarrow 2 \rightarrow 4$, $1 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 4$.

- ▶ What is the total number of such paths?

Continuation

- ▶ What is the total number of such paths?
- ▶ What is the number of paths from 1 to 4 in 100 steps?

Continuation

- ▶ What is the total number of such paths?
- ▶ What is the number of paths from 1 to 4 in 100 steps?
- ▶ How do we count such paths?

Continuation

- ▶ A graph G is a pair (V, E) where V is a set called the set of vertices and E is a collection of pairs of the form $\{x, y\}$ where $x, y \in V$, with $x \neq y$, called the edges of the graph.

Continuation

- ▶ A graph G is a pair (V, E) where V is a set called the set of vertices and E is a collection of pairs of the form $\{x, y\}$ where $x, y \in V$, with $x \neq y$, called the edges of the graph.
- ▶ We are talking about simple graphs, which means no multiple edges between vertices and no loops and the edges are undirected. There are more general notions of graphs which we will not go into.

Continuation

- ▶ A graph G is a pair (V, E) where V is a set called the set of vertices and E is a collection of pairs of the form $\{x, y\}$ where $x, y \in V$, with $x \neq y$, called the edges of the graph.
- ▶ We are talking about simple graphs, which means no multiple edges between vertices and no loops and the edges are undirected. There are more general notions of graphs which we will not go into.
- ▶ For the square graph example above: $V = \{1, 2, 3, 4\}$ and

$$E = \{\{1, 2\}, \{2, 4\}, \{1, 3\}, \{3, 4\}\}.$$

- ▶ In the following, we take $V = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and so E is a collection of pairs of the form $\{i, j\}$ with $i, j \in V$.

Paths

- ▶ In the following, we take $V = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and so E is a collection of pairs of the form $\{i, j\}$ with $i, j \in V$.
- ▶ For $m \geq 2$, a path of length m from i to j , is a tuple of the form $(i, k_1, k_2, \dots, k_{m-1}, j)$ where $\{i, k_1\}, \{k_1, k_2\}, \dots, \{k_{m-1}, j\}$ are edges, i.e., they are elements of E .

Paths

- ▶ In the following, we take $V = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ and so E is a collection of pairs of the form $\{i, j\}$ with $i, j \in V$.
- ▶ For $m \geq 2$, a path of length m from i to j , is a tuple of the form $(i, k_1, k_2, \dots, k_{m-1}, j)$ where $\{i, k_1\}, \{k_1, k_2\}, \dots, \{k_{m-1}, j\}$ are edges, i.e., they are elements of E .
- ▶ Note that k_j 's need not be distinct.

Adjacency matrix

- **Definition 22.6:** The **adjacency** matrix $A = [a_{ij}]_{1 \leq i, j \leq n}$ of a graph $G = (V, E)$ where $V = \{1, 2, \dots, n\}$ is defined by taking

$$a_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Adjacency matrix

- **Definition 22.6:** The **adjacency** matrix $A = [a_{ij}]_{1 \leq i, j \leq n}$ of a graph $G = (V, E)$ where $V = \{1, 2, \dots, n\}$ is defined by taking

$$a_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

- Note that A is a real symmetric matrix and the diagonal entries are all equal to zero.

Two step paths

► We have

$$\begin{aligned}(A^2)_{ij} &= \sum_{k=1}^n a_{ik} a_{kj} \\&= \sum_{k: a_{ik} \cdot a_{kj} \neq 0} a_{ik} a_{kj} \\&= \#\{k : a_{ik} \cdot a_{kj} \neq 0\} \\&= \#\{k : a_{ik} \neq 0, \ a_{kj} \neq 0\} \\&= \#\{k : (i, k), (k, j) \in E\} \\&= \text{The number of two step paths from } i \text{ to } j.\end{aligned}$$

Two step paths

- ▶ We have

$$\begin{aligned}(A^2)_{ij} &= \sum_{k=1}^n a_{ik} a_{kj} \\&= \sum_{k: a_{ik} \cdot a_{kj} \neq 0} a_{ik} a_{kj} \\&= \#\{k : a_{ik} \cdot a_{kj} \neq 0\} \\&= \#\{k : a_{ik} \neq 0, \ a_{kj} \neq 0\} \\&= \#\{k : (i, k), (k, j) \in E\} \\&= \text{The number of two step paths from } i \text{ to } j.\end{aligned}$$

- ▶ Therefore, $(i, j)^{\text{th}}$ -entry of A^2 is the number of two step paths from (i, j) .

m -step paths

- ▶ For $m \geq 2$,

m -step paths

► For $m \geq 2$,



$$\begin{aligned} & (A^m)_{ij} \\ = & \sum_{k_1, k_2, \dots, k_{m-1}} a_{ik_1} a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_{m-2} k_{m-1}} a_{k_{m-1} j} \\ = & \#\{(k_1, k_2, \dots, k_{m-1}) : a_{ik_1} a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_{m-2} k_{m-1}} a_{k_{m-1} j} \neq 0\} \\ = & \#\{(i, k_1, \dots, k_{m-1}, j) : (i, k_1), (k_1, k_2), \dots, (k_{m-1}, j) \in E\} \\ = & \text{Number of paths of length } m \text{ from } i \text{ to } j. \end{aligned}$$

m -step paths

- ▶ For $m \geq 2$,



$$\begin{aligned} & (A^m)_{ij} \\ &= \sum_{k_1, k_2, \dots, k_{m-1}} a_{ik_1} a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_{m-2} k_{m-1}} a_{k_{m-1} j} \\ &= \#\{(k_1, k_2, \dots, k_{m-1}) : a_{ik_1} a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_{m-2} k_{m-1}} a_{k_{m-1} j} \neq 0\} \\ &= \#\{(i, k_1, \dots, k_{m-1}, j) : (i, k_1), (k_1, k_2), \dots, (k_{m-1}, j) \in E\} \\ &= \text{Number of paths of length } m \text{ from } i \text{ to } j. \end{aligned}$$

- ▶ In other words, $(i, j)^{\text{th}}$ -entry of A^m is exactly the number of paths of length m from i to j in the graph G , where A is the adjacency matrix of the graph G .

- ▶ Observe that the adjacency matrix of a graph G is a real symmetric matrix. Hence it is normal and also has real eigenvalues.

- ▶ Observe that the adjacency matrix of a graph G is a real symmetric matrix. Hence it is normal and also has real eigenvalues.
- ▶ By the spectral theorem

$$A = UDU^*$$

for some unitary U and real diagonal matrix D .

- ▶ Observe that the adjacency matrix of a graph G is a real symmetric matrix. Hence it is normal and also has real eigenvalues.
- ▶ By the spectral theorem

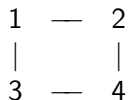
$$A = UDU^*$$

for some unitary U and real diagonal matrix D .

- ▶ Consequently $A^m = UD^mU^*$. This allows us to compute the number of paths of length m between any two vertices.

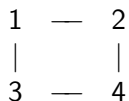
Example

- ▶ Consider the 'square graph' with vertices $\{1, 2, 3, 4\}$:



Example

- ▶ Consider the 'square graph' with vertices $\{1, 2, 3, 4\}$:

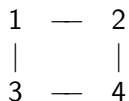


- ▶ The adjacency matrix of this graph is given by

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

Example

- ▶ Consider the 'square graph' with vertices $\{1, 2, 3, 4\}$:



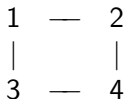
- ▶ The adjacency matrix of this graph is given by

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

- ▶ You may compute the eigenvalues of A and diagonalize it to compute the powers.

Example

- ▶ Consider the 'square graph' with vertices $\{1, 2, 3, 4\}$:



- ▶ The adjacency matrix of this graph is given by

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

- ▶ You may compute the eigenvalues of A and diagonalize it to compute the powers.
- ▶ Here we will take a different approach.

The powers of A

- By direct computation,

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

The powers of A

- By direct computation,

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

- and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

The powers of A

- By direct computation,

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

- and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

- In other words, $A^3 = 4A$.

The powers of A

- By direct computation,

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

- and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

- In other words, $A^3 = 4A$.
- Therefore, $A^4 = 4A^2$, $A^5 = 4A^3 = 16A$, $A^6 = 16A^2$.

The powers of A

- By direct computation,

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

- and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

- In other words, $A^3 = 4A$.
- Therefore, $A^4 = 4A^2$, $A^5 = 4A^3 = 16A$, $A^6 = 16A^2$.
- By induction, we get $A^{2m+1} = 4^m A$, $A^{2m+2} = 4^m A^2$ for all $m \geq 0$.

The powers of A

- By direct computation,

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

- and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

- In other words, $A^3 = 4A$.
- Therefore, $A^4 = 4A^2$, $A^5 = 4A^3 = 16A$, $A^6 = 16A^2$.
- By induction, we get $A^{2m+1} = 4^m A$, $A^{2m+2} = 4^m A^2$ for all $m \geq 0$.
- This we can write down the number of paths of given length between any two vertices of the square graph.

The powers of A

- ▶ By direct computation,

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

- ▶ and

$$A^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 4 & 0 & 0 & 4 \\ 4 & 0 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{bmatrix}.$$

- ▶ In other words, $A^3 = 4A$.
- ▶ Therefore, $A^4 = 4A^2$, $A^5 = 4A^3 = 16A$, $A^6 = 16A^2$.
- ▶ By induction, we get $A^{2m+1} = 4^m A$, $A^{2m+2} = 4^m A^2$ for all $m \geq 0$.
- ▶ This we can write down the number of paths of given length between any two vertices of the square graph.
- ▶ **END OF LECTURE 22**

Lecture 23: Spectral theorem -II and III

- ▶ Reference: Graphs and Matrices by R. Bapat

Lecture 23: Spectral theorem -II and III

- ▶ Reference: Graphs and Matrices by R. Bapat
- ▶ For further information on connections between graph theory and matrix theory.

Self-adjoint and normal matrices

- ▶ Recall:
- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.

Self-adjoint and normal matrices

- ▶ Recall:
- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.

Self-adjoint and normal matrices

- ▶ Recall:
- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3 + 5i \\ 3 - 5i & 1 \end{bmatrix}.$$

Self-adjoint and normal matrices

- ▶ Recall:
- ▶ **Definition 20.1:** (i) A complex square matrix A is said to be **self-adjoint** if $A^* = A$. (ii) A complex square matrix A is said to be **normal** if $A^*A = AA^*$.
- ▶ Recall that for $A = [a_{ij}]_{1 \leq i, j \leq n}$, we have $A^* = [\overline{a_{ji}}]_{1 \leq i, j \leq n}$.
- ▶ In particular, every real symmetric matrix is self-adjoint.
- ▶ Here is an example of a self-adjoint matrix which is not real and symmetric:

$$B = \begin{bmatrix} 2 & 3 + 5i \\ 3 - 5i & 1 \end{bmatrix}.$$

- ▶ Note that diagonal entry of every self-adjoint matrix is real as $\overline{a_{ii}} = a_{ii}$ for every i .

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .
- ▶ Recall that the diagonal entries of D are the eigenvalues of A , as the characteristic polynomial of A and D are same.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

- ▶ If I_{n_j} denotes the identity matrix of size $n_j \times n_j$, the matrix D can be written as:

$$D = \begin{bmatrix} a_1 I_{n_1} & 0 & 0 & \dots & 0 \\ 0 & a_2 I_{n_2} & 0 & \dots & 0 \\ 0 & 0 & a_3 I_{n_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_k I_{n_k} \end{bmatrix}$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

- ▶ Clearly $Q_1, Q_2, \dots, Q_k$ are projections, $Q_i Q_j = 0$, for $i \neq j$ (they are mutually orthogonal) and $Q_1 + Q_2 + \cdots + Q_k = I$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

- ▶ From $P_j = UQ_jU^*, 1 \leq j \leq k$, it is clear that $P_1, P_2, \dots, P_k$ are projections such that $P_iP_j = 0$ for $i \neq j$ and

$$P_1 + P_2 + \cdots + P_k = I.$$

Spectral Theorem -II

- ▶ We have now proved the following Theorem.

Spectral Theorem -II

- ▶ We have now proved the following Theorem.
- ▶ **Theorem 23.1 (Spectral Theorem -II):** Let A be a normal matrix and let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A . Then there exist mutually orthogonal projections $P_1, P_2, \dots, P_k$, such that

$$I = P_1 + P_2 + \cdots + P_k;$$

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k.$$

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

- Note that:

$$\langle y_1 \oplus y_2 \oplus \cdots \oplus y_k, z_1 \oplus z_2 \oplus \cdots \oplus z_k \rangle = \sum_{j=1}^k \langle y_j, z_j \rangle.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

- ▶ In other words, M_j is the eigenspace of A with respect to eigenvalue a_j .

Continuation

- ▶ In other words, we have proved the following statement:

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.
- ▶ Clearly given the normal matrix A , the decomposition of $\mathbb{C}^n$ as in this theorem is uniquely determined and so the corresponding projections are also uniquely determined. This also shows that the decomposition of A as in Spectral Theorem -II:

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k, I = P_1 + P_2 + \cdots + P_k$$

where $P_1, P_2, \dots, P_k$ are mutually orthogonal projections is unique up to permutation.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?
- ▶ Ans: No. If $A = I$, then $A = UIU^*$ for any unitary U . Hence U is not unique even up to scalar.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?
- ▶ Ans: No. If $A = I$, then $A = UIU^*$ for any unitary U . Hence U is not unique even up to scalar.
- ▶ **END OF LECTURE 23**

Lecture 24: Polynomial spectral mapping theorem

- ▶ We recall different versions of the spectral theorem.

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .
- ▶ Recall that the diagonal entries of D are the eigenvalues of A , as the characteristic polynomial of A and D are same.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

- ▶ If I_{n_j} denotes the identity matrix of size $n_j \times n_j$, the matrix D can be written as:

$$D = \begin{bmatrix} a_1 I_{n_1} & 0 & 0 & \dots & 0 \\ 0 & a_2 I_{n_2} & 0 & \dots & 0 \\ 0 & 0 & a_3 I_{n_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_k I_{n_k} \end{bmatrix}$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

- ▶ Clearly $Q_1, Q_2, \dots, Q_k$ are projections, $Q_i Q_j = 0$, for $i \neq j$ (they are mutually orthogonal) and $Q_1 + Q_2 + \cdots + Q_k = I$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

- ▶ From $P_j = UQ_jU^*, 1 \leq j \leq k$, it is clear that $P_1, P_2, \dots, P_k$ are projections such that $P_iP_j = 0$ for $i \neq j$ and

$$P_1 + P_2 + \cdots + P_k = I.$$

Spectral Theorem -II

- ▶ We have now proved the following Theorem.

Spectral Theorem -II

- ▶ We have now proved the following Theorem.
- ▶ **Theorem 23.1 (Spectral Theorem -II):** Let A be a normal matrix and let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A . Then there exist mutually orthogonal projections $P_1, P_2, \dots, P_k$, such that

$$I = P_1 + P_2 + \cdots + P_k;$$

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k.$$

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

- Note that:

$$\langle y_1 \oplus y_2 \oplus \cdots \oplus y_k, z_1 \oplus z_2 \oplus \cdots \oplus z_k \rangle = \sum_{j=1}^k \langle y_j, z_j \rangle.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

- ▶ In other words, M_j is the eigenspace of A with respect to eigenvalue a_j .

Continuation

- ▶ In other words, we have proved the following statement:

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.
- ▶ Clearly given the normal matrix A , the decomposition of $\mathbb{C}^n$ as in this theorem is uniquely determined and so the corresponding projections are also uniquely determined. This also shows that the decomposition of A as in Spectral Theorem -II:

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k, I = P_1 + P_2 + \cdots + P_k$$

where $P_1, P_2, \dots, P_k$ are mutually orthogonal projections is unique up to permutation.

Non-uniqueness in diagonalization

- Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?
- ▶ Ans: No. If $A = I$, then $A = UIU^*$ for any unitary U . Hence U is not unique even up to scalar.

Polynomial spectral mapping theorem

- **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.
- ▶ Now it is easy to see that the diagonal entries of T^2 are $d_1^2, d_2^2, \dots, d_n^2$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.
- ▶ Now it is easy to see that the diagonal entries of T^2 are $d_1^2, d_2^2, \dots, d_n^2$.
- ▶ More generally, for any $k \in \mathbb{N}$ the diagonal entries of T^k are $d_1^k, d_2^k, \dots, d_n^k$.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

- ▶ This proves the theorem.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

- ▶ This proves the theorem.
- ▶ **Exercise 24.2:** Find an alternative proof which does not use upper triangularization.

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

- and $I = UIU^*$. ■

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

- and $I = UIU^*$. ■

Polynomials of normal matrices -II

► **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

Polynomials of normal matrices -II

- **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

- Then for any complex polynomial q ,

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

Polynomials of normal matrices -II

- **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

- Then for any complex polynomial q ,

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

- **Proof:** We have

$$\begin{aligned} A^2 &= (a_1 P_1 + a_2 P_2 + \cdots + a_k P_k)(a_1 P_1 + a_2 P_2 + \cdots + a_k P_k) \\ &= a_1^2 P_1 + a_2^2 P_2 + \cdots + a_k^2 P_k \end{aligned}$$

as $P_i P_j = \delta_{ij} P_j$.

Continuation

- By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ Now the result follows by taking linear combinations of the powers of A . ■

Continuation

- ▶ By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ Now the result follows by taking linear combinations of the powers of A . ■
- ▶ **Remark 24.5:** It is to be noted that

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

may not be the spectral decomposition of $q(A)$ as $q(a_1), \dots, q(a_k)$ may not be distinct.

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

- ▶ For instance we can define $\sin(A)$, $\cos(A)$, e^A etc by this method.

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

- ▶ For instance we can define $\sin(A)$, $\cos(A)$, e^A etc by this method.
- ▶ At the moment this is only a definition. But it has many natural properties. Studying this concept not only for matrices but also for operators (infinite dimensional matrices) is the subject of Functional Calculus.

Examples

- ▶ Recall:
- ▶ Example 21.2 and 24.6: Suppose

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

Examples

- ▶ Recall:
- ▶ Example 21.2 and 24.6: Suppose

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

- ▶ Then as A is self-adjoint it is normal.

Examples

- ▶ Recall:
- ▶ Example 21.2 and 24.6: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

Examples

- ▶ Recall:
- ▶ Example 21.2 and 24.6: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.

Examples

- ▶ Recall:
- ▶ Example 21.2 and 24.6: Suppose

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

- ▶ Then as A is self-adjoint it is normal.
- ▶ The characteristic polynomial of A is

$$p(x) = \det(xI - A) = (x-2)^2 - 1 = x^2 - 4x + 3 = (x-3)(x-1).$$

- ▶ Hence the eigenvalues of A are 1 and 3.
- ▶ Solving corresponding eigen equations we see that

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Continuation

- ▶ Normalizing these eigenvectors, and taking them as columns we get a unitary,

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ satisfying,

$$AU = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

- ▶ Alternatively,

$$A = U \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} U^*.$$

Continuation

- ▶ We have the spectral decomposition of A as

$$A = 3P_1 + 1.P_2$$

where

$$P_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Continuation

- ▶ We have the spectral decomposition of A as

$$A = 3P_1 + 1.P_2$$

where

$$P_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

- ▶ and

$$P_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Continuation

- ▶ We have the spectral decomposition of A as

$$A = 3P_1 + 1.P_2$$

where

$$P_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

- ▶ and

$$P_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

- ▶ Note that P_1, P_2 are mutually orthogonal and $P_1 + P_2 = I$.

Continuation

- Example 24.7: Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Continuation

- ▶ Example 24.7: Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ We have seen that U is a unitary. So it must have eigenvalues on the unit circle.

Continuation

- ▶ Example 24.7: Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ We have seen that U is a unitary. So it must have eigenvalues on the unit circle.
- ▶ The characteristic polynomial of U is

$$p_U(x) = (x^2 - \frac{1}{2}) - \frac{1}{2} = x^2 - 1 = (x - 1)(x + 1)$$

Continuation

- ▶ Example 24.7: Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ We have seen that U is a unitary. So it must have eigenvalues on the unit circle.
- ▶ The characteristic polynomial of U is

$$p_U(x) = (x^2 - \frac{1}{2}) - \frac{1}{2} = x^2 - 1 = (x - 1)(x + 1)$$

- ▶ So the eigenvalues of U are 1 and -1 .

Continuation

- ▶ Example 24.7: Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ We have seen that U is a unitary. So it must have eigenvalues on the unit circle.
- ▶ The characteristic polynomial of U is

$$p_U(x) = (x^2 - \frac{1}{2}) - \frac{1}{2} = x^2 - 1 = (x - 1)(x + 1)$$

- ▶ So the eigenvalues of U are 1 and -1 .
- ▶ Compute the spectral decomposition of U .

Continuation

- ▶ **Example 24.7:** Consider

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- ▶ We have seen that U is a unitary. So it must have eigenvalues on the unit circle.
- ▶ The characteristic polynomial of U is

$$p_U(x) = (x^2 - \frac{1}{2}) - \frac{1}{2} = x^2 - 1 = (x - 1)(x + 1)$$

- ▶ So the eigenvalues of U are 1 and -1 .
- ▶ Compute the spectral decomposition of U .
- ▶ **Challenge Question:** Every time you diagonalize you get a unitary. Continue diagonalizing these unitaries. Does the process terminate or does it become cyclic?

Another example

- **Example 24.8:** Let B be the $n \times n$ matrix defined by

$$b_{ij} = \begin{cases} 0 & \text{if } i = j; \\ 1 & \text{otherwise.} \end{cases}$$

Another example

- ▶ **Example 24.8:** Let B be the $n \times n$ matrix defined by

$$b_{ij} = \begin{cases} 0 & \text{if } i = j; \\ 1 & \text{otherwise.} \end{cases}$$

- ▶ Compute powers of B .

Another example

- ▶ **Example 24.8:** Let B be the $n \times n$ matrix defined by

$$b_{ij} = \begin{cases} 0 & \text{if } i = j; \\ 1 & \text{otherwise.} \end{cases}$$

- ▶ Compute powers of B .
- ▶ Note that B is the adjacency matrix of the **complete graph** with n -vertices (usually denoted by K_n), where any two distinct vertex is connected by an edge.

Another example

- ▶ **Example 24.8:** Let B be the $n \times n$ matrix defined by

$$b_{ij} = \begin{cases} 0 & \text{if } i = j; \\ 1 & \text{otherwise.} \end{cases}$$

- ▶ Compute powers of B .
- ▶ Note that B is the adjacency matrix of the **complete graph** with n -vertices (usually denoted by K_n), where any two distinct vertex is connected by an edge.
- ▶ Draw K_1, K_2, K_3, K_4

Continuation

- ▶ We write B as $B = nP - I$ where P is the matrix with

$$p_{ij} = \frac{1}{n}, \quad \forall i, j.$$

Continuation

- ▶ We write B as $B = nP - I$ where P is the matrix with

$$p_{ij} = \frac{1}{n}, \quad \forall i, j.$$

- ▶ We have seen this P before. It is a projection. It is the projection to the one dimensional space:

$$\left\{ \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix} : a \in \mathbb{C} \right\}.$$

Continuation

- ▶ We write B as $B = nP - I$ where P is the matrix with

$$p_{ij} = \frac{1}{n}, \quad \forall i, j.$$

- ▶ We have seen this P before. It is a projection. It is the projection to the one dimensional space:

$$\left\{ \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix} : a \in \mathbb{C} \right\}.$$

- ▶ Take $Q = (I - P)$.

Continuation

- ▶ We write B as $B = nP - I$ where P is the matrix with

$$p_{ij} = \frac{1}{n}, \quad \forall i, j.$$

- ▶ We have seen this P before. It is a projection. It is the projection to the one dimensional space:

$$\left\{ \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix} : a \in \mathbb{C} \right\}.$$

- ▶ Take $Q = (I - P)$.
- ▶ Then $B = nP - (P + Q) = (n - 1)P + (-1)Q$

Continuation

- Observe that P, Q are mutually orthogonal projections whose sum is I . It follows that

$$B = (n - 1)P + (-1)Q$$

is the spectral decomposition of B .

Continuation

- ▶ Observe that P, Q are mutually orthogonal projections whose sum is I . It follows that

$$B = (n - 1)P + (-1)Q$$

is the spectral decomposition of B .

- ▶ So the eigenvalues of B are $(n - 1)$ with multiplicity 1 and (-1) with multiplicity $(n - 1)$.
- ▶ Moreover, $B^m = (n - 1)^m P + (-1)^m Q$ for all $m \geq 0$.

Continuation

- ▶ Observe that P, Q are mutually orthogonal projections whose sum is I . It follows that

$$B = (n - 1)P + (-1)Q$$

is the spectral decomposition of B .

- ▶ So the eigenvalues of B are $(n - 1)$ with multiplicity 1 and (-1) with multiplicity $(n - 1)$.
- ▶ Moreover, $B^m = (n - 1)^m P + (-1)^m Q$ for all $m \geq 0$.
- ▶ Use this to compute the number of paths in the complete graph and also compute it directly by considering the graph.

Continuation

- ▶ Observe that P, Q are mutually orthogonal projections whose sum is I . It follows that

$$B = (n - 1)P + (-1)Q$$

is the spectral decomposition of B .

- ▶ So the eigenvalues of B are $(n - 1)$ with multiplicity 1 and (-1) with multiplicity $(n - 1)$.
- ▶ Moreover, $B^m = (n - 1)^m P + (-1)^m Q$ for all $m \geq 0$.
- ▶ Use this to compute the number of paths in the complete graph and also compute it directly by considering the graph.
- ▶ **END OF LECTURE 24.**

Lecture 25: Hadamard matrices and circulant matrices

- ▶ We recall different versions of the spectral theorem.

Spectral theorem (Version -I)

- **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .

Spectral theorem (Version -I)

- ▶ **Theorem 20.7 (Spectral Theorem-I):** Let A be a complex matrix. Then there exists a unitary matrix U and a diagonal matrix D such that

$$A = UDU^*$$

if and only if A is normal.

- ▶ The theorem tells us that a matrix is diagonalizable through a unitary iff it is normal. In particular, every normal matrix is diagonalizable.
- ▶ Now we will present this Theorem in a different way.
- ▶ Consider the set up as above. Let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A .
- ▶ Recall that the diagonal entries of D are the eigenvalues of A , as the characteristic polynomial of A and D are same.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

Continuation

- ▶ This means that every diagonal entry of D is one of the a_j 's.
- ▶ Suppose that a_j appears n_j times in the diagonal with $1 \leq n_j$ and $n_1 + n_2 + \cdots + n_k = n$.
- ▶ Without loss of generality, we may assume that repeated entries are clubbed together, that is, the diagonal entries of D are equal to

$$(a_1, a_1, \dots, a_1, a_2, a_2, \dots, a_2, a_3, a_3, \dots, a_k, a_k)$$

where a_j appears n_j times.

- ▶ If I_{n_j} denotes the identity matrix of size $n_j \times n_j$, the matrix D can be written as:

$$D = \begin{bmatrix} a_1 I_{n_1} & 0 & 0 & \dots & 0 \\ 0 & a_2 I_{n_2} & 0 & \dots & 0 \\ 0 & 0 & a_3 I_{n_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_k I_{n_k} \end{bmatrix}$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

Continuation

- ▶ We may then re-write D as $D = a_1 Q_1 + a_2 Q_2 + \cdots + a_k Q_k$
- ▶ where,

$$Q_1 = \begin{bmatrix} I_{n_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, Q_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

- ▶ and so on up to

$$Q_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_k} \end{bmatrix}.$$

- ▶ Clearly $Q_1, Q_2, \dots, Q_k$ are projections, $Q_i Q_j = 0$, for $i \neq j$ (they are mutually orthogonal) and $Q_1 + Q_2 + \cdots + Q_k = I$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

Continuation

- ▶ Now take $P_j = UQ_jU^*, 1 \leq j \leq k$.
- ▶ Then we have

$$\begin{aligned} A &= UDU^* \\ &= U(a_1Q_1 + a_2Q_2 + \cdots + a_kQ_k)U^* \\ &= a_1UQ_1U^* + a_2UQ_2U^* + \cdots + a_kUQ_kU^* \\ &= a_1P_1 + a_2P_2 + \cdots + a_kP_k. \end{aligned}$$

- ▶ From $P_j = UQ_jU^*, 1 \leq j \leq k$, it is clear that $P_1, P_2, \dots, P_k$ are projections such that $P_iP_j = 0$ for $i \neq j$ and

$$P_1 + P_2 + \cdots + P_k = I.$$

Spectral Theorem -II

- ▶ We have now proved the following Theorem.

Spectral Theorem -II

- ▶ We have now proved the following Theorem.
- ▶ **Theorem 23.1 (Spectral Theorem -II):** Let A be a normal matrix and let $a_1, a_2, \dots, a_k$ be the distinct eigenvalues of A . Then there exist mutually orthogonal projections $P_1, P_2, \dots, P_k$, such that

$$I = P_1 + P_2 + \cdots + P_k;$$

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k.$$

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

Orthogonal Direct sums

- **Definition 23.2:** Suppose $M_1, M_2, \dots, M_k$ are mutually orthogonal subspaces of a finite dimensional inner product space V such that every vector x in V decomposes uniquely as

$$x = y_1 + y_2 + \cdots + y_k$$

where $y_j \in M_j$, $1 \leq j \leq k$, then V is said to be an **orthogonal direct sum** of $M_1, M_2, \dots, M_k$.

- (Notation) Sometimes this is denoted by

$$x = y_1 \oplus y_2 \oplus \cdots \oplus y_k.$$

We may also write $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$.

- Note that:

$$\langle y_1 \oplus y_2 \oplus \cdots \oplus y_k, z_1 \oplus z_2 \oplus \cdots \oplus z_k \rangle = \sum_{j=1}^k \langle y_j, z_j \rangle.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

Continuation

- ▶ Now in Spectral theorem-II, taking $M_j = P_j(\mathbb{C}^n) = \{P_j x : x \in \mathbb{C}^n\}$, we see that $\mathbb{C}^n$ is a direct sum of $M_1, M_2, \dots, M_k$.
- ▶ That is, every vector x in $\mathbb{C}^n$ decomposes uniquely as $x = (P_1 + P_2 + \dots + P_k)x = P_1 x + P_2 x + \dots + P_k x$ with $P_j x \in M_j$.
- ▶ We also note that

$$Ax = (a_1 P_1 + a_2 P_2 + \dots + a_k P_k)x = a_1 P_1 x + a_2 P_2 x + \dots + a_k P_k x$$

is the unique decomposition of Ax .

- ▶ What is this subspace $M_j = P_j(\mathbb{C}^n)$?
- ▶ **Exercise 23.3:** Show that

$$M_j = \{y \in \mathbb{C}^n : Ay = a_j y\}.$$

- ▶ In other words, M_j is the eigenspace of A with respect to eigenvalue a_j .

Continuation

- ▶ In other words, we have proved the following statement:

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.

Continuation

- ▶ In other words, we have proved the following statement:
- ▶ **Theorem 23.4 (Spectral theorem -III):** Let A be a normal matrix. Then the eigenspaces of distinct eigenvalues of A are mutually orthogonal and $\mathbb{C}^n$ is their direct sum.
- ▶ Clearly given the normal matrix A , the decomposition of $\mathbb{C}^n$ as in this theorem is uniquely determined and so the corresponding projections are also uniquely determined. This also shows that the decomposition of A as in Spectral Theorem -II:

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k, I = P_1 + P_2 + \cdots + P_k$$

where $P_1, P_2, \dots, P_k$ are mutually orthogonal projections is unique up to permutation.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?

Non-uniqueness in diagonalization

- ▶ Recall that in Spectral Theorem -I, we had

$$A = UDU^*$$

where U is a unitary and D is diagonal. Here D is unique up to permutation of the diagonal entries.

- ▶ However, U is not unique. We can always replace U by zU where $|z| = 1$. Then zU is also a unitary and $A = (zU)D(zU)^*$.
- ▶ Is U unique up to multiplication by scalar when D is fixed?
- ▶ Ans: No. If $A = I$, then $A = UIU^*$ for any unitary U . Hence U is not unique even up to scalar.

Polynomial spectral mapping theorem

- **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.
- ▶ Now it is easy to see that the diagonal entries of T^2 are $d_1^2, d_2^2, \dots, d_n^2$.

Polynomial spectral mapping theorem

- ▶ **Theorem 24.1:** Let A be an $n \times n$ matrix and let $d_1, d_2, \dots, d_n$ be the eigenvalues of A . Then for any complex polynomial q , the eigenvalues of $q(A)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ **Proof:** By Schur's upper triangularization theorem there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

- ▶ Note that the diagonal entries of T are the eigenvalues of A .
- ▶ Without loss of generality, we may denote the diagonal entries of T as $d_1, d_2, \dots, d_n$.
- ▶ Now it is easy to see that the diagonal entries of T^2 are $d_1^2, d_2^2, \dots, d_n^2$.
- ▶ More generally, for any $k \in \mathbb{N}$ the diagonal entries of T^k are $d_1^k, d_2^k, \dots, d_n^k$.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

- ▶ This proves the theorem.

Continuation

- ▶ Now suppose $q(x) = c_0 + c_1x + c_2x^2 + \cdots + c_mx^m$, then the diagonal entries of $q(T)$ are $q(d_1), q(d_2), \dots, q(d_n)$.
- ▶ We have

$$q(A) = Uq(T)U^*$$

and $q(T)$ is upper triangular, Therefore $q(A)$ and $q(T)$ have same characteristic polynomial and hence same set of eigenvalues.

- ▶ This proves the theorem.
- ▶ **Exercise 24.2:** Find an alternative proof which does not use upper triangularization.

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

Polynomials of normal matrices first version

- ▶ **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- ▶ Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- ▶ **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

- and $I = UIU^*$. ■

Polynomials of normal matrices first version

- **Theorem 24.3:** Let A be a normal matrix. Suppose U is a unitary and D is a diagonal matrix such that

$$A = UDU^*.$$

- Then for any complex polynomial q ,

$$q(A) = Uq(D)U^* = U \begin{bmatrix} q(d_1) & 0 & \dots & 0 \\ 0 & q(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(d_n) \end{bmatrix} U^*.$$

- **Proof:** This is clear as

$$A^k = UD^kU^* = U \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} U^*,$$

for every $k \geq 1$,

- and $I = UIU^*$. ■

Polynomials of normal matrices -II

► **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

Polynomials of normal matrices -II

- **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

- Then for any complex polynomial q ,

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

Polynomials of normal matrices -II

- **Theorem 24.4:** Let A be a normal matrix. Suppose

$$A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k,$$

is the spectral decomposition of A (This means that $a_1, a_2, \dots, a_k$ are distinct eigenvalues of A and $P_1, P_2, \dots, P_k$ are mutually orthogonal projections such that $P_1 + P_2 + \cdots + P_k = I$.)

- Then for any complex polynomial q ,

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

- **Proof:** We have

$$\begin{aligned} A^2 &= (a_1 P_1 + a_2 P_2 + \cdots + a_k P_k)(a_1 P_1 + a_2 P_2 + \cdots + a_k P_k) \\ &= a_1^2 P_1 + a_2^2 P_2 + \cdots + a_k^2 P_k \end{aligned}$$

as $P_i P_j = \delta_{ij} P_j$.

Continuation

- By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ Now the result follows by taking linear combinations of the powers of A . ■

Continuation

- ▶ By induction,

$$A^m = a_1^m P_1 + a_2^m P_2 + \cdots + a_k^m P_k$$

for all $m \geq 1$ and for $m = 0$, $A^0 = I = P_1 + P_2 + \cdots + P_k$.

- ▶ Now the result follows by taking linear combinations of the powers of A . ■
- ▶ **Remark 24.5:** It is to be noted that

$$q(A) = q(a_1)P_1 + q(a_2)P_2 + \cdots + q(a_k)P_k.$$

may not be the spectral decomposition of $q(A)$ as $q(a_1), \dots, q(a_k)$ may not be distinct.

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

- ▶ For instance we can define $\sin(A)$, $\cos(A)$, e^A etc by this method.

Functional Calculus

- ▶ The last two theorems suggest that for a normal matrix A , if f is a function defined on $\sigma(A)$ (the spectrum of A) we may define $f(A)$ by taking

$$\begin{aligned} f(A) &:= U \begin{bmatrix} f(d_1) & 0 & \dots & 0 \\ 0 & f(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{bmatrix} U^* \\ &= f(a_1)P_1 + f(a_2)P_2 + \dots + f(a_k)P_k. \end{aligned}$$

- ▶ For instance we can define $\sin(A)$, $\cos(A)$, e^A etc by this method.
- ▶ At the moment this is only a definition. But it has many natural properties. Studying this concept not only for matrices but also for operators (infinite dimensional matrices) is the subject of Functional Calculus.

Hadamard matrices

- **Definition 25.1:** A square matrix is said to be a **Hadamard matrix** if every entry of it is ± 1 and its rows are mutually orthogonal.

Hadamard matrices

- ▶ **Definition 25.1:** A square matrix is said to be a **Hadamard matrix** if every entry of it is ± 1 and its rows are mutually orthogonal.
- ▶ **Example 25.2:** The matrices $H_1 = [1]$ and

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

are Hadamard matrices.

Hadamard matrices

- ▶ **Definition 25.1:** A square matrix is said to be a **Hadamard matrix** if every entry of it is ± 1 and its rows are mutually orthogonal.
- ▶ **Example 25.2:** The matrices $H_1 = [1]$ and

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

are Hadamard matrices.

- ▶ Note that if an $n \times n$ matrix H is a Hadamard matrix, clearly

$$HH^t = nI.$$

Therefore $\frac{1}{\sqrt{n}}H^t$ is the inverse of $\frac{1}{\sqrt{n}}H$. Alternatively, $\frac{1}{\sqrt{n}}H$ is an orthogonal matrix. Consequently we also have $H^tH = nI$. Therefore columns of H are also mutually orthogonal.

A construction

- **Proposition 25.3:** If H is an $n \times n$ Hadamard matrix then the block matrix

$$K = \begin{bmatrix} H & H \\ H & -H \end{bmatrix}$$

is a $2n \times 2n$ Hadamard matrix.

A construction

- **Proposition 25.3:** If H is an $n \times n$ Hadamard matrix then the block matrix

$$K = \begin{bmatrix} H & H \\ H & -H \end{bmatrix}$$

is a $2n \times 2n$ Hadamard matrix.

- **Proof** By block matrix multiplication,

$$\begin{aligned} KK^t &= \begin{bmatrix} H & H \\ H & -H \end{bmatrix} \cdot \begin{bmatrix} H^t & H^t \\ H^t & -H^t \end{bmatrix} \\ &= \begin{bmatrix} nl + nl & nl - nl \\ nl - nl & nl + nl \end{bmatrix} \\ &= 2n \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \end{aligned}$$

A construction

- **Proposition 25.3:** If H is an $n \times n$ Hadamard matrix then the block matrix

$$K = \begin{bmatrix} H & H \\ H & -H \end{bmatrix}$$

is a $2n \times 2n$ Hadamard matrix.

- **Proof** By block matrix multiplication,

$$\begin{aligned} KK^t &= \begin{bmatrix} H & H \\ H & -H \end{bmatrix} \cdot \begin{bmatrix} H^t & H^t \\ H^t & -H^t \end{bmatrix} \\ &= \begin{bmatrix} nl + nl & nl - nl \\ nl - nl & nl + nl \end{bmatrix} \\ &= 2n \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \end{aligned}$$

- This proves the claim. ■

Continuation

- ▶ Applying this construction to H_2 defined above we get a 4×4 Hadamard matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Continuation

- ▶ Applying this construction to H_2 defined above we get a 4×4 Hadamard matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

- ▶ By induction, we can construct a Hadamard matrix of order 2^n for every $n \geq 0$.

Multiples of 4

- ▶ **Proposition 25.4:** Suppose H is an $n \times n$ Hadamard matrix with $n > 2$. Then n is a multiple of 4.

Multiples of 4

- ▶ **Proposition 25.4:** Suppose H is an $n \times n$ Hadamard matrix with $n > 2$. Then n is a multiple of 4.
- ▶ **Proof:** Note that entries are ± 1 and with $n > 1$, any two rows are orthogonal means that n has to be even.

Multiples of 4

- ▶ **Proposition 25.4:** Suppose H is an $n \times n$ Hadamard matrix with $n > 2$. Then n is a multiple of 4.
- ▶ **Proof:** Note that entries are ± 1 and with $n > 1$, any two rows are orthogonal means that n has to be even.
- ▶ If H is a Hadamard matrix, and if we replace a row with its negative, clearly it stays as Hadamard matrix.

Multiples of 4

- ▶ **Proposition 25.4:** Suppose H is an $n \times n$ Hadamard matrix with $n > 2$. Then n is a multiple of 4.
- ▶ **Proof:** Note that entries are ± 1 and with $n > 1$, any two rows are orthogonal means that n has to be even.
- ▶ If H is a Hadamard matrix, and if we replace a row with its negative, clearly it stays as Hadamard matrix.
- ▶ Therefore we may assume that the first entry of every row is $+1$. (That is $h_{j1} = +1, \forall j$.)

Multiples of 4

- ▶ **Proposition 25.4:** Suppose H is an $n \times n$ Hadamard matrix with $n > 2$. Then n is a multiple of 4.
- ▶ **Proof:** Note that entries are ± 1 and with $n > 1$, any two rows are orthogonal means that n has to be even.
- ▶ If H is a Hadamard matrix, and if we replace a row with its negative, clearly it stays as Hadamard matrix.
- ▶ Therefore we may assume that the first entry of every row is $+1$. (That is $h_{j1} = +1, \forall j$.)
- ▶ In other words now the first column has only $+1$'s. This forces that every other column has equal number of positive and negative entries. In particular n must be even.

Continuation

- ▶ Take $n = 2k$. So that every column except the first one has exactly k positive signs and k negative signs.

Continuation

- ▶ Take $n = 2k$. So that every column except the first one has exactly k positive signs and k negative signs.
- ▶ If H is a Hadamard matrix and if we permute the rows it would stay as a Hadamard matrix. So we may assume that first k entries of second column are positive and the next k are negative. That is:

$$h_{j2} = \begin{cases} +1 & \text{if } 1 \leq j \leq k; \\ -1 & \text{if } (k+1) \leq j \leq 2k. \end{cases}$$

Continuation

- ▶ Take $n = 2k$. So that every column except the first one has exactly k positive signs and k negative signs.
- ▶ If H is a Hadamard matrix and if we permute the rows it would stay as a Hadamard matrix. So we may assume that first k entries of second column are positive and the next k are negative. That is:

$$h_{j2} = \begin{cases} +1 & \text{if } 1 \leq j \leq k; \\ -1 & \text{if } (k+1) \leq j \leq 2k. \end{cases}$$

- ▶ Since $n > 2$, and n is even we can consider third and fourth columns of H .

Continuation

- ▶ Suppose among the first k entries of the third column r entries are positive and $(k - r)$ are negative and among the remaining k entries s are positive and $(k - s)$ are negative.

Continuation

- ▶ Suppose among the first k entries of the third column r entries are positive and $(k - r)$ are negative and among the remaining k entries s are positive and $(k - s)$ are negative.
- ▶ Since the total number of positive entries in a column has to be k , we get $r + s = k$.

Continuation

- ▶ Orthogonality of second and third column shows that $h_{j2}.h_{j3} = +1$ at k many j 's.

Continuation

- ▶ Orthogonality of second and third column shows that $h_{j2}.h_{j3} = +1$ at k many j 's.
- ▶ Counting the positions where both the entry of second and third column have same sign we get $r + (k - s) = k$ or equivalently $r = s$.

Continuation

- ▶ Orthogonality of second and third column shows that $h_{j2}.h_{j3} = +1$ at k many j 's.
- ▶ Counting the positions where both the entry of second and third column have same sign we get $r + (k - s) = k$ or equivalently $r = s$.
- ▶ Then $k = r + s = 2r$ is even.

Continuation

- ▶ Orthogonality of second and third column shows that $h_{j2}.h_{j3} = +1$ at k many j 's.
- ▶ Counting the positions where both the entry of second and third column have same sign we get $r + (k - s) = k$ or equivalently $r = s$.
- ▶ Then $k = r + s = 2r$ is even.
- ▶ Consequently $n = 2k$ is a multiple of 4.

Hadamard's Conjecture

- ▶ **Hadamard's Conjecture:** There exists a Hadamard matrix of order $4k$ for every natural number k .

Hadamard's Conjecture

- ▶ **Hadamard's Conjecture:** There exists a Hadamard matrix of order $4k$ for every natural number k .
- ▶ The conjecture is still open.

Hadamard's Conjecture

- ▶ **Hadamard's Conjecture:** There exists a Hadamard matrix of order $4k$ for every natural number k .
- ▶ The conjecture is still open.
- ▶ As per Wikipedia currently 668 is the smallest number for which we don't know the existence of a Hadamard matrix.

Complex Hadamard matrices

- **Definition 25.5:** A square matrix $A = [a_{ij}]$ is said to be a **complex Hadamard** matrix if its rows are orthogonal and $|a_{ij}| = 1$ for all i, j .

Complex Hadamard matrices

- ▶ **Definition 25.5:** A square matrix $A = [a_{ij}]$ is said to be a **complex Hadamard** matrix if its rows are orthogonal and $|a_{ij}| = 1$ for all i, j .
- ▶ If A is a complex Hadamard matrix of order $n \times n$, then

$$AA^* = A^*A = nI.$$

Complex Hadamard matrices

- ▶ **Definition 25.5:** A square matrix $A = [a_{ij}]$ is said to be a **complex Hadamard** matrix if its rows are orthogonal and $|a_{ij}| = 1$ for all i, j .
- ▶ If A is a complex Hadamard matrix of order $n \times n$, then

$$AA^* = A^*A = nI.$$

- ▶ In the following Example we find it convenient to index the rows and columns of the matrix from 0 to $(n - 1)$ instead of 1 to n .
- ▶ **Example 25.6:** For $n \geq 1$, consider the matrix $W = [w_{jk}]_{0 \leq j, k \leq (n-1)}$ defined by

$$w_{jk} = e^{\frac{2\pi ijk}{n}}.$$

Then W is a complex Hadamard matrix (Prove it.)

Finite Fourier transform

- ▶ This matrix W of Example 25.6 is known as the **Finite Fourier Transform** matrix or Discrete Fourier Transform matrix of order n and it appears in various contexts.

Finite Fourier transform

- ▶ This matrix W of Example 25.6 is known as the **Finite Fourier Transform** matrix or Discrete Fourier Transform matrix of order n and it appears in various contexts.
- ▶ It has several practical applications.

Finite Fourier transform

- ▶ This matrix W of Example 25.6 is known as the **Finite Fourier Transform** matrix or Discrete Fourier Transform matrix of order n and it appears in various contexts.
- ▶ It has several practical applications.
- ▶ For instance, for $n = 3$ we have

$$W = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix},$$

where $\omega = e^{\frac{2\pi i}{3}}$.

Circulant matrices

- **Definition 25.6:** For $n \geq 2$, an $n \times n$ matrix C is said to be a **circulant matrix** if

$$C = \begin{bmatrix} c_0 & c_{n-1} & c_{n-2} & \dots & c_1 \\ c_1 & c_0 & c_{n-1} & \dots & c_2 \\ c_2 & c_1 & c_0 & \dots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \dots & c_0 \end{bmatrix}$$

for some $c_0, c_1, \dots, c_{n-1} \in \mathbb{C}$.

Circulant matrices

- **Definition 25.6:** For $n \geq 2$, an $n \times n$ matrix C is said to be a **circulant matrix** if

$$C = \begin{bmatrix} c_0 & c_{n-1} & c_{n-2} & \dots & c_1 \\ c_1 & c_0 & c_{n-1} & \dots & c_2 \\ c_2 & c_1 & c_0 & \dots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \dots & c_0 \end{bmatrix}$$

for some $c_0, c_1, \dots, c_{n-1} \in \mathbb{C}$.

- Suppose A is as above. Consider the matrix

$$S = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

- ▶ Then it is easily seen that

$$C = c_0 + c_1 S + c_2 S^2 + \cdots + c_{n-1} S^{n-1}.$$

- ▶ Then it is easily seen that

$$C = c_0 + c_1 S + c_2 S^2 + \cdots + c_{n-1} S^{n-1}.$$

- ▶ So C is a polynomial in S .

- ▶ Then it is easily seen that

$$C = c_0 + c_1 S + c_2 S^2 + \cdots + c_{n-1} S^{n-1}.$$

- ▶ So C is a polynomial in S .
- ▶ Note that S is a permutation matrix and in particular it is a unitary.

- ▶ The characteristic polynomial of S is

$$p(x) = x^n - 1.$$

- ▶ The characteristic polynomial of S is

$$p(x) = x^n - 1.$$

- ▶ Therefore the eigenvalues of S are the n -th roots of unity:
 $\sigma(S) = \{e^{\frac{2\pi ij}{n}} : 0 \leq j \leq (n-1)\}.$

- ▶ The characteristic polynomial of S is

$$p(x) = x^n - 1.$$

- ▶ Therefore the eigenvalues of S are the n -th roots of unity:
 $\sigma(S) = \{e^{\frac{2\pi ij}{n}} : 0 \leq j \leq (n-1)\}.$
- ▶ Taking $\omega = e^{\frac{2\pi i}{n}}$. $\sigma(S) = \{1, \omega, \omega^2, \omega^{n-1}\}.$

- ▶ The characteristic polynomial of S is

$$p(x) = x^n - 1.$$

- ▶ Therefore the eigenvalues of S are the n -th roots of unity:
 $\sigma(S) = \{e^{\frac{2\pi ij}{n}} : 0 \leq j \leq (n-1)\}.$
- ▶ Taking $\omega = e^{\frac{2\pi i}{n}}$. $\sigma(S) = \{1, \omega, \omega^2, \omega^{n-1}\}.$
- ▶ Let D be the diagonal matrix with diagonal entries $1, \omega, \dots, \omega^{n-1}.$

- ▶ The characteristic polynomial of S is

$$p(x) = x^n - 1.$$

- ▶ Therefore the eigenvalues of S are the n -th roots of unity:
 $\sigma(S) = \{e^{\frac{2\pi ij}{n}} : 0 \leq j \leq (n-1)\}.$
- ▶ Taking $\omega = e^{\frac{2\pi i}{n}}$. $\sigma(S) = \{1, \omega, \omega^2, \omega^{n-1}\}.$
- ▶ Let D be the diagonal matrix with diagonal entries $1, \omega, \dots, \omega^{n-1}.$
- ▶ So we have $d_{jk} = \delta_{jk}\omega^j, 0 \leq j, k \leq (n-1).$

Continuation

- Note that on indexing the rows and columns of S from 0 to $(n - 1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l + 1) \\ 0 & \text{otherwise} \end{cases}$$

Continuation

- ▶ Note that on indexing the rows and columns of S from 0 to $(n - 1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l + 1) \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Here 0 is taken as $(n - 1) + 1$ (The addition is modulo n).

Continuation

- ▶ Note that on indexing the rows and columns of S from 0 to $(n - 1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l + 1) \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Here 0 is taken as $(n - 1) + 1$ (The addition is modulo n).
- ▶ Let W be the finite Fourier transform matrix of order n .

Continuation

- ▶ Note that on indexing the rows and columns of S from 0 to $(n - 1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l + 1) \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Here 0 is taken as $(n - 1) + 1$ (The addition is modulo n).
- ▶ Let W be the finite Fourier transform matrix of order n .
- ▶ Then

$$(WS)_{jl} = \sum_{k=0}^{n-1} \omega^{jk} \cdot s_{kl} = \omega^{j(l+1)}.$$

Continuation

- ▶ Note that on indexing the rows and columns of S from 0 to $(n-1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l+1) \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Here 0 is taken as $(n-1) + 1$ (The addition is modulo n).
- ▶ Let W be the finite Fourier transform matrix of order n .
- ▶ Then

$$(WS)_{jl} = \sum_{k=0}^{n-1} \omega^{jk} \cdot s_{kl} = \omega^{j(l+1)}.$$

- ▶ Also

$$(DW)_{jl} = \sum_{k=0}^{n-1} d_{jk} \omega^{kl} = \omega^j \cdot \omega^{jl} = \omega^{j(l+1)}$$

Continuation

- ▶ Note that on indexing the rows and columns of S from 0 to $(n-1)$, we have

$$s_{kl} = \begin{cases} 1 & \text{if } k = (l+1) \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Here 0 is taken as $(n-1) + 1$ (The addition is modulo n).
- ▶ Let W be the finite Fourier transform matrix of order n .
- ▶ Then

$$(WS)_{jl} = \sum_{k=0}^{n-1} \omega^{jk} \cdot s_{kl} = \omega^{j(l+1)}.$$

- ▶ Also

$$(DW)_{jl} = \sum_{k=0}^{n-1} d_{jk} \omega^{kl} = \omega^j \cdot \omega^{jl} = \omega^{j(l+1)}$$

- ▶ As $W^*W = nI$, $\frac{1}{\sqrt{n}}$ is a unitary. Therefore $\frac{1}{\sqrt{n}}WS = D \cdot \frac{1}{\sqrt{n}}W$, or $S = \frac{1}{n}W^*DW$ is the diagonalization of S .

- ▶ Recall that the Circulant matrix

$$C = c_0 + c_1 S + \cdots + c_{n-1} S^{n-1}.$$

Continuation

- ▶ Recall that the Circulant matrix

$$C = c_0 + c_1 S + \cdots + c_{n-1} S^{n-1}.$$

- ▶ Consequently the circulant matrix $C = \frac{1}{n} W^* q(D) W$. where $q(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$.

- ▶ Recall that the Circulant matrix

$$C = c_0 + c_1 S + \cdots + c_{n-1} S^{n-1}.$$

- ▶ Consequently the circulant matrix $C = \frac{1}{n} W^* q(D) W$. where $q(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$.
- ▶ In particular, the spectrum of C is given by

$$\{c_0 + c_1 \omega^k + c_2 \omega^{2k} + \cdots + c_{n-1} \omega^{(n-1)k} : 0 \leq k \leq (n-1)\}.$$

Continuation

- ▶ Recall that the Circulant matrix

$$C = c_0 + c_1 S + \cdots + c_{n-1} S^{n-1}.$$

- ▶ Consequently the circulant matrix $C = \frac{1}{n} W^* q(D) W$. where $q(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$.
- ▶ In particular, the spectrum of C is given by

$$\{c_0 + c_1 \omega^k + c_2 \omega^{2k} + \cdots + c_{n-1} \omega^{(n-1)k} : 0 \leq k \leq (n-1)\}.$$

- ▶ **END OF LECTURE 25.**

Lecture 26: Positive matrices

- ▶ Now we study a very important class of matrices. The significance of this class can't be over emphasized.

Lecture 26: Positive matrices

- ▶ Now we study a very important class of matrices. The significance of this class can't be over emphasized.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .

Lecture 26: Positive matrices

- ▶ Now we study a very important class of matrices. The significance of this class can't be over emphasized.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.

Lecture 26: Positive matrices

- ▶ Now we study a very important class of matrices. The significance of this class can't be over emphasized.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.

Lecture 26: Positive matrices

- ▶ Now we study a very important class of matrices. The significance of this class can't be over emphasized.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.
- ▶ Matrices whose entries are positive would be called as **entrywise positive** matrices. That is also an important class, but we will not be studying them now.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.
 - ▶ (vi) $A = S^2$ for some self-adjoint $n \times n$ matrix S .

Continuation

► **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.
- ▶ Now for any $x \in \mathbb{C}^n$:

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \overline{x_i} (Ax)_i \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n a_{ij} x_j \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n \langle v_i, v_j \rangle \cdot x_j\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

► where $y = \sum_{i=1}^n x_i v_i$.

- Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

- where $y = \sum_{i=1}^n x_i v_i$.
- Hence

$$\langle x, Ax \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

- ▶ implies that $a \geq 0$ as $\langle x, x \rangle \neq 0$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

- ▶ Then clearly S is self-adjoint and $A = S^2$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ Example 26.3: Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.
- ▶ Find all self-adjoint operators S such that $R = S^2$. (Exercise)

Cartesian decomposition

- **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

Cartesian decomposition

- **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.
- ▶ **END OF LECTURE 26**

Lecture 27: Sums, products and square roots of positive matrices

- ▶ First we recall the notion of positivity and some characterizations.

Lecture 27: Sums, products and square roots of positive matrices

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .

Lecture 27: Sums, products and square roots of positive matrices

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.

Lecture 27: Sums, products and square roots of positive matrices

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.

Lecture 27: Sums, products and square roots of positive matrices

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.
- ▶ Matrices whose entries are positive would be called as **entrywise positive** matrices. That is also an important class, but we will not be studying them now.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.
 - ▶ (vi) $A = S^2$ for some self-adjoint $n \times n$ matrix S .

Continuation

► **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.
- ▶ Now for any $x \in \mathbb{C}^n$:

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \overline{x_i} (Ax)_i \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n a_{ij} x_j \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n \langle v_i, v_j \rangle \cdot x_j\end{aligned}$$

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

► where $y = \sum_{i=1}^n x_i v_i$.

- Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

- where $y = \sum_{i=1}^n x_i v_i$.
- Hence

$$\langle x, Ax \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

- ▶ implies that $a \geq 0$ as $\langle x, x \rangle \neq 0$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

- ▶ Then clearly S is self-adjoint and $A = S^2$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ Example 26.3: Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.
- ▶ Find all self-adjoint operators S such that $R = S^2$. (Exercise)

Cartesian decomposition

- **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.
- ▶ This theorem shows that the set of $n \times n$ positive matrices has 'cone' structure: It is closed under taking sums and it is closed under multiplication by positive scalar.

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .
- ▶ Now, $B^*AB = B^*D^*DB = (DB)^*(DB)$. Hence B^*AB is positive from the definition of positivity. We may also see this from looking at the quadratic form. ■

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.
- ▶ **Proof:** The first part is clear as the trace and determinant of a matrix are respectively the sum and the product of its eigenvalues and a positive matrix has non-negative eigenvalues. The second claim follows from $a_{ii} = \langle v_i, v_i \rangle$ in part (iv) of the characterization.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- ▶ We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.
- In other words, we have the Cauchy-Schwarz inequality:

$$|\langle x, y \rangle|^2 \leq \|x\|^2 \cdot \|y\|^2.$$

Uniqueness of positive square root

- **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.
- ▶ Let $b_1, b_2, \dots, b_k$ the distinct eigenvalues of B and $B = b_1 Q_1 + b_2 Q_2 + \dots + b_k Q_k$ be the spectral decomposition of B .

► We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then $B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k$.

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then
$$B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k.$$
- ▶ This proves uniqueness. ■

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

- ▶ For any projection P , the unitary $2P - I = P - P^\perp$ is a square root of I . This shows that I has infinitely many square roots (in dimension bigger than 1) if we do not insist on positivity.

Notation

- **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?
- ▶ **END OF LECTURE 27.**

Lecture 28: Polar decomposition theorem

- ▶ First we recall the notion of positivity and some characterizations.

Lecture 28: Polar decomposition theorem

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .

Lecture 28: Polar decomposition theorem

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.

Lecture 28: Polar decomposition theorem

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.

Lecture 28: Polar decomposition theorem

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.
- ▶ Matrices whose entries are positive would be called as **entrywise positive** matrices. That is also an important class, but we will not be studying them now.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.
 - ▶ (vi) $A = S^2$ for some self-adjoint $n \times n$ matrix S .

Continuation

► **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.
- ▶ Now for any $x \in \mathbb{C}^n$:

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \overline{x_i} (Ax)_i \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n a_{ij} x_j \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n \langle v_i, v_j \rangle \cdot x_j\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

► where $y = \sum_{i=1}^n x_i v_i$.

- Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

- where $y = \sum_{i=1}^n x_i v_i$.
- Hence

$$\langle x, Ax \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

- ▶ implies that $a \geq 0$ as $\langle x, x \rangle \neq 0$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

- ▶ Then clearly S is self-adjoint and $A = S^2$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ Example 26.3: Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.
- ▶ Find all self-adjoint operators S such that $R = S^2$. (Exercise)

Cartesian decomposition

- **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.
- ▶ This theorem shows that the set of $n \times n$ positive matrices has 'cone' structure: It is closed under taking sums and it is closed under multiplication by positive scalar.

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .
- ▶ Now, $B^*AB = B^*D^*DB = (DB)^*(DB)$. Hence B^*AB is positive from the definition of positivity. We may also see this from looking at the quadratic form. ■

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.
- ▶ **Proof:** The first part is clear as the trace and determinant of a matrix are respectively the sum and the product of its eigenvalues and a positive matrix has non-negative eigenvalues. The second claim follows from $a_{ii} = \langle v_i, v_i \rangle$ in part (iv) of the characterization.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- ▶ We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.
- In other words, we have the Cauchy-Schwarz inequality:

$$|\langle x, y \rangle|^2 \leq \|x\|^2 \cdot \|y\|^2.$$

Uniqueness of positive square root

- **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.
- ▶ Let $b_1, b_2, \dots, b_k$ the distinct eigenvalues of B and $B = b_1 Q_1 + b_2 Q_2 + \dots + b_k Q_k$ be the spectral decomposition of B .

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then $B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k$.

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then $B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k$.
- ▶ This proves uniqueness. ■

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

- ▶ For any projection P , the unitary $2P - I = P - P^\perp$ is a square root of I . This shows that I has infinitely many square roots (in dimension bigger than 1) if we do not insist on positivity.

Notation

- **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?
- ▶ **END OF REVIEW**

Polar decomposition theorem

- **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.
- ▶ Let $\mathcal{M}$ be the range of $|B|$, that is,

$$\mathcal{M} = \{|B|x : x \in \mathbb{C}^n\}.$$

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.
- ▶ Let $\mathcal{M}$ be the range of $|B|$, that is,

$$\mathcal{M} = \{|B|x : x \in \mathbb{C}^n\}.$$

- ▶ Clearly $\mathcal{M}$ is a subspace of $\mathbb{C}^n$.

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.
- ▶ Observe that, for $u, v \in \mathbb{C}^n$,

$$\begin{aligned}\langle Bu, Bv \rangle &= \langle u, B^* Bv \rangle \\ &= \langle u, |B|^2 v \rangle \\ &= \langle |B|u, |B|v \rangle\end{aligned}$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.
- ▶ Observe that, for $u, v \in \mathbb{C}^n$,

$$\begin{aligned}\langle Bu, Bv \rangle &= \langle u, B^* Bv \rangle \\ &= \langle u, |B|^2 v \rangle \\ &= \langle |B|u, |B|v \rangle\end{aligned}$$

- ▶ In particular, if $|B|x = |B|y$, then $|B|(x - y) = 0$. Hence,

$$\langle |B|(x - y), |B|(x - y) \rangle = 0.$$

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

- ▶ Hence U_0 is isometric.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y), B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

- ▶ Hence U_0 is isometric.
- ▶ From the definition of U_0 it is clear that U_0 maps $\mathcal{M}$ onto $\mathcal{N}$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.
- ▶ Extend $\{v_1, v_2, \dots, v_k\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.
- ▶ Extend $\{v_1, v_2, \dots, v_k\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$.
- ▶ Similarly, extend $\{U_0 v_1, U_0 v_2, \dots, U_0 v_k\}$ to an orthonormal basis $\{U_0 v_1, U_0 v_2, \dots, U_0 v_k, w_1, w_2, \dots, w_{n-k}\}$ of $\mathbb{C}^n$.

Continuation

- Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.
- ▶ If B is invertible, and $B = UR$, then as $\det(B) = \det(U) \cdot \det(R) \neq 0$, R is also invertible and we get

$$U = BR^{-1} = B(|B|)^{-1}.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.
- ▶ If B is invertible, and $B = UR$, then as $\det(B) = \det(U) \cdot \det(R) \neq 0$, R is also invertible and we get

$$U = BR^{-1} = B(|B|)^{-1}.$$

- ▶ Therefore, U is uniquely determined. ■

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

- We can write D as

$$\begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} \cdot \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix}.$$

using the polar decompositions of $d_1, d_2, \dots, d_n$.

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

- We can write D as

$$\begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} \cdot \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix}.$$

using the polar decompositions of $d_1, d_2, \dots, d_n$.

- Note that if $d_j = 0$, then when we write $d_j = e^{i\theta_j}|d_j|$, $e^{i\theta_j}$ is not unique.

Continuation

- Now $A = V|A|$ where

$$V = U \begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} U^*$$

is a unitary and

Continuation

- ▶ Now $A = V|A|$ where

$$V = U \begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} U^*$$

is a unitary and



$$|A| = U \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix} U^*$$

is positive.

- ▶ Alternatively, if $A = a_1P_1 + a_2P_2 + \cdots + a_kP_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .
- ▶ Take $V = e^{i\theta_1} P_1 + e^{i\theta_2} P_2 + \cdots + e^{i\theta_k} P_k$ and $|A| = |a_1| P_1 + |a_2| P_2 + \cdots + |a_k| P_k$.

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .
- ▶ Take $V = e^{i\theta_1} P_1 + e^{i\theta_2} P_2 + \cdots + e^{i\theta_k} P_k$ and $|A| = |a_1| P_1 + |a_2| P_2 + \cdots + |a_k| P_k$.
- ▶ Then $A = V|A|$ is the polar decomposition of A .

Example

- Example 28.3: Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Example

- Example 28.3: Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

Example

- ▶ Example 28.3: Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

- ▶ Recall that an $n \times n$ matrix with all entries equal to $\frac{1}{n}$ is a projection.

Example

- ▶ Example 28.3: Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

- ▶ Recall that an $n \times n$ matrix with all entries equal to $\frac{1}{n}$ is a projection.
- ▶ Therefore

$$B^*B = 12 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} + 0 \cdot \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}$$

is the spectral decomposition of B^*B .

► Therefore,

$$\begin{aligned}|B| &= \sqrt{12} \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \\ &= \begin{bmatrix} 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \end{bmatrix}.\end{aligned}$$

Continuation

► Take

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

Continuation

- Take

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

- We may now verify that

$$B = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix} \cdot \begin{bmatrix} 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \end{bmatrix}.$$

is a polar decomposition of B .

Continuation

- Take

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

- We may now verify that

$$B = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix} \cdot \begin{bmatrix} 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \end{bmatrix}.$$

is a polar decomposition of B .

- **END OF LECTURE 28.**

Lecture 29: Singular value decomposition

- ▶ First we recall the notion of positivity and some characterizations.

Lecture 29: Singular value decomposition

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .

Lecture 29: Singular value decomposition

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.

Lecture 29: Singular value decomposition

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.

Lecture 29: Singular value decomposition

- ▶ First we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.
- ▶ Matrices whose entries are positive would be called as **entrywise positive** matrices. That is also an important class, but we will not be studying them now.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.
 - ▶ (vi) $A = S^2$ for some self-adjoint $n \times n$ matrix S .

Continuation

► **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.

Continuation

- ▶ **Proof:** (i) $\Rightarrow$ (ii). Take $m = n$ and $C = B$.
- ▶ (ii) $\Rightarrow$ (iii). Let $v_1, v_2, \dots, v_n$ be the columns of C . Then $v_j \in \mathbb{C}^m$ for every j and $A = C^*C$ implies $a_{ij} = \langle v_i, v_j \rangle$.
- ▶ (iii) $\Rightarrow$ (iv). We have $a_{ij} = \langle v_i, v_j \rangle, \quad \forall i, j$.
- ▶ Now for any $x \in \mathbb{C}^n$:

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \overline{x_i} (Ax)_i \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n a_{ij} x_j \\ &= \sum_{i=1}^n \overline{x_i} \cdot \sum_{j=1}^n \langle v_i, v_j \rangle \cdot x_j\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

Continuation

► Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

► where $y = \sum_{i=1}^n x_i v_i$.

- Therefore,

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i=1}^n \sum_{j=1}^n \langle x_i v_i, x_j v_j \rangle \\ &= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n x_j v_j \right\rangle \\ &= \langle y, y \rangle\end{aligned}$$

- where $y = \sum_{i=1}^n x_i v_i$.
- Hence

$$\langle x, Ax \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

Continuation

- ▶ (iv) $\Rightarrow$ (v). It is given that $\langle x, Ax \rangle \geq 0$ for every $x \in \mathbb{C}^n$.
- ▶ First we want to show that $A = A^*$. Here we use the polarization identity and the fact that if $\langle v, w \rangle$ is real then $\langle v, w \rangle = \langle w, v \rangle$. For all x, y ,

$$\begin{aligned}\langle x, Ay \rangle &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle (x + i^j y), A(x + i^j y) \rangle \\ &= \frac{1}{4} \sum_{j=0}^3 i^{-j} \langle A(x + i^j y), (x + i^j y) \rangle \\ &= \langle Ax, y \rangle\end{aligned}$$

- ▶ This proves $A^* = A$ from the defining condition of the adjoint.
- ▶ Now suppose a is an eigenvalue of A . Choose an eigenvector x with a as the eigenvalue. Then

$$\langle x, Ax \rangle = a \langle x, x \rangle \geq 0.$$

- ▶ implies that $a \geq 0$ as $\langle x, x \rangle \neq 0$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

Continuation

- ▶ (v) $\Rightarrow$ (vi). We assume $A = A^*$ and the eigenvalues of A are non-negative.
- ▶ By spectral theorem there exists a unitary U and a diagonal matrix D , such that

$$A = UDU^*.$$

- ▶ Since the eigenvalues of A are non-negative, the diagonal entries of D are non-negative. We denote the diagonal entries by $d_1, d_2, \dots, d_n$.
- ▶ Take

$$S = U \begin{bmatrix} \sqrt{d_1} & 0 & \dots & 0 \\ 0 & \sqrt{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_n} \end{bmatrix} U^*.$$

- ▶ Then clearly S is self-adjoint and $A = S^2$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ Example 26.3: Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.

Continuation

- ▶ (vi) $\Rightarrow$ (i). Assume $A = S^2$ where $S = S^*$.
- ▶ In particular, $A = S^*S$. This proves (i). ■
- ▶ **Example 26.3:** Take

$$R = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

- ▶ Clearly R is self-adjoint. We have the characteristic polynomial of R , as

$$p(x) = (x - 2)^2 - 1 = x^2 - 4x + 3 = (x - 1)(x - 3).$$

- ▶ Therefore, the eigenvalues of R are $\{1, 3\}$, which are non-negative. Then by part (v) of the previous Theorem, R is positive.
- ▶ Note that though R is positive as per our definition, some of its entries are negative.
- ▶ Find all self-adjoint operators S such that $R = S^2$. (Exercise)

Cartesian decomposition

- **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■

Cartesian decomposition

- ▶ **Theorem 26.3 (Cartesian decomposition):** Let A be a complex square matrix. Then A decomposes uniquely as

$$A = B + iC$$

where B, C are self-adjoint.

- ▶ **Proof:** Take $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ Then it is easily verified that B, C are self-adjoint and $A = B + iC$.
- ▶ Conversely, suppose $A = B + iC$, with B, C self-adjoint. We see directly that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.
- ▶ This proves uniqueness. ■
- ▶ This is known as **Cartesian decomposition**.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.

Constructing positive matrices out of positive matrices

- ▶ **Theorem 27.1** Suppose A, B are positive matrices and a, b are positive real numbers. Then $aA + bB$ is positive.
- ▶ **Proof:** Suppose A, B are $n \times n$ matrices. Consider arbitrary $x \in \mathbb{C}^n$.
- ▶ Then, $\langle x, (aA + bB)x \rangle = a\langle x, Ax \rangle + b\langle x, Bx \rangle \geq 0$
- ▶ as the quadratic forms $\langle x, Ax \rangle$, and $\langle x, Bx \rangle$ are positive.
- ▶ Then by part (iv) of the characterization theorem, $aA + bB$ is positive. ■
- ▶ Note that this theorem does not follow directly from the definition of positivity or from the eigenvalue criterion.
- ▶ This theorem shows that the set of $n \times n$ positive matrices has 'cone' structure: It is closed under taking sums and it is closed under multiplication by positive scalar.

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .

Products

- ▶ **Remark 27.2:** Suppose A, B are $n \times n$ positive matrices. Then AB need not be positive as it may not be self-adjoint.
- ▶ **Theorem 27.3:** Let A be an $n \times n$ positive matrix. Suppose B is an $n \times m$ matrix. Then B^*AB is positive.
- ▶ **Proof:** As A is positive, $A = D^*D$ for some matrix D .
- ▶ Now, $B^*AB = B^*D^*DB = (DB)^*(DB)$. Hence B^*AB is positive from the definition of positivity. We may also see this from looking at the quadratic form. ■

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.

Trace and Determinant

- ▶ **Theorem 27.4:** Let A be a positive matrix. Then $\text{trace}(A) \geq 0$ and $\det(A) \geq 0$. Diagonal entries of A are positive.
- ▶ **Proof:** The first part is clear as the trace and determinant of a matrix are respectively the sum and the product of its eigenvalues and a positive matrix has non-negative eigenvalues. The second claim follows from $a_{ii} = \langle v_i, v_i \rangle$ in part (iv) of the characterization.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

Gram matrices

- ▶ **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- ▶ We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- ▶ Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- ▶ We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.

Gram matrices

- **Definition 27.5:** Let $v_1, v_2, \dots, v_n$ be vectors in an inner product space V . Then their Gram matrix is defined as the matrix

$$A = [\langle v_i, v_j \rangle]_{1 \leq i, j \leq n}.$$

- We have seen that Gram matrices are positive and conversely all positive matrices can be written as Gram matrices. In Probability theory Gram matrices appear as 'covariance matrices'.
- Suppose x, y are vectors in an inner product space V . Consider their Gram matrix:

$$G = \begin{bmatrix} \langle x, x \rangle & \langle x, y \rangle \\ \langle y, x \rangle & \langle y, y \rangle \end{bmatrix}.$$

- We know that G is positive. Hence its determinant is positive. So we get $\langle x, x \rangle \cdot \langle y, y \rangle - \langle x, y \rangle \cdot \langle y, x \rangle \geq 0$.
- In other words, we have the Cauchy-Schwarz inequality:

$$|\langle x, y \rangle|^2 \leq \|x\|^2 \cdot \|y\|^2.$$

Uniqueness of positive square root

- **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.

Uniqueness of positive square root

- ▶ **Theorem 27.5:** Let A be a positive matrix. Then there exists unique positive S such that $A = S^2$.
- ▶ **Proof:** In the proof the main characterization theorem for positive matrices we have seen that if A is positive then there exists positive S such that $A = S^2$.
- ▶ Now suppose B is positive and $A = B^2$.
- ▶ Let $b_1, b_2, \dots, b_k$ the distinct eigenvalues of B and $B = b_1 Q_1 + b_2 Q_2 + \dots + b_k Q_k$ be the spectral decomposition of B .

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then $B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k$.

Continuation

- ▶ We have,

$$A = B^2 = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k.$$

- ▶ Note that as $b_1, b_2, \dots, b_k$ are positive and distinct, $b_1^2, b_2^2, \dots, b_k^2$ are also positive and distinct.
- ▶ Hence $A = b_1^2 Q_1 + b_2^2 Q_2 + \cdots + b_k^2 Q_k$ is the unique (up to permutation) spectral decomposition of A .
- ▶ In other words, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , then
$$B = \sqrt{a_1} P_1 + \sqrt{a_2} P_2 + \cdots + \sqrt{a_k} P_k.$$
- ▶ This proves uniqueness. ■

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

Continuation

- ▶ Note that if we do not insist on positivity of the square root there can be many square roots.
- ▶ For instance, for A positive, if $A = UDU^*$ is the diagonalization of A ,

$$U \begin{bmatrix} \pm\sqrt{d_1} & 0 & \cdots & 0 \\ 0 & \pm\sqrt{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm\sqrt{d_n} \end{bmatrix} U^*$$

with any choice of sign on the diagonal is a square root of A .

- ▶ For any projection P , the unitary $2P - I = P - P^\perp$ is a square root of I . This shows that I has infinitely many square roots (in dimension bigger than 1) if we do not insist on positivity.

Notation

- **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.

Notation

- ▶ **Notation:** For a positive matrix A , $A^{\frac{1}{2}}$ denotes the square root of A . For any matrix B , $|B|$ denotes $(B^*B)^{\frac{1}{2}}$.
- ▶ Note that for a positive matrix A , $A^{\frac{1}{2}} = A$ if and only if A is a projection.
- ▶ **Question:** Do we have $|A + B| \leq |A| + |B|$? In other words can we say that $|A| + |B| - |A + B|$ is positive?

Polar decomposition theorem

- **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

Polar decomposition theorem

- **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.
- ▶ Let $\mathcal{M}$ be the range of $|B|$, that is,

$$\mathcal{M} = \{|B|x : x \in \mathbb{C}^n\}.$$

Polar decomposition theorem

- ▶ **Theorem 28.1:** Let B be an $n \times n$ matrix. Then B factorizes as

$$B = UR$$

where U is a unitary and R is positive. Here $R = |B|$ and U is uniquely determined if B is invertible.

- ▶ **Proof:** We want to find U, R satisfying

$$B = UR$$

as above.

- ▶ As suggested by the second line we take $R = |B| = (B^*B)^{\frac{1}{2}}$.
- ▶ Let $\mathcal{M}$ be the range of $|B|$, that is,

$$\mathcal{M} = \{|B|x : x \in \mathbb{C}^n\}.$$

- ▶ Clearly $\mathcal{M}$ is a subspace of $\mathbb{C}^n$.

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.
- ▶ Observe that, for $u, v \in \mathbb{C}^n$,

$$\begin{aligned}\langle Bu, Bv \rangle &= \langle u, B^* Bv \rangle \\ &= \langle u, |B|^2 v \rangle \\ &= \langle |B|u, |B|v \rangle\end{aligned}$$

Continuation

- ▶ Similarly let $\mathcal{N}$ be the range of B :

$$\mathcal{N} = \{Bx : x \in \mathbb{C}^n\}.$$

- ▶ Define $U_0 : \mathcal{M} \rightarrow \mathcal{N}$ by

$$U_0|B|x = Bx, x \in \mathbb{C}^n.$$

- ▶ We claim that U_0 is a well-defined unitary from $\mathcal{M}$ to $\mathcal{N}$.
- ▶ Observe that, for $u, v \in \mathbb{C}^n$,

$$\begin{aligned}\langle Bu, Bv \rangle &= \langle u, B^* Bv \rangle \\ &= \langle u, |B|^2 v \rangle \\ &= \langle |B|u, |B|v \rangle\end{aligned}$$

- ▶ In particular, if $|B|x = |B|y$, then $|B|(x - y) = 0$. Hence,

$$\langle |B|(x - y), |B|(x - y) \rangle = 0.$$

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

- ▶ Hence U_0 is isometric.

Continuation

- ▶ Taking $u = v = (x - y)$, in the previous equation we get $\langle B(x - y)B(x - y) \rangle = 0$ or $B(x - y) = 0$, that is, $Bx = By$.
- ▶ In other words if $|B|x = |B|y$ then $Bx = By$. This shows that U_0 is well defined.
- ▶ Clearly U_0 as defined is linear.
- ▶ Further,

$$\langle |B|u, |B|v \rangle = \langle Bu, Bv \rangle, \quad \forall u, v \in \mathbb{C}^n$$

shows that

$$\langle U_0(|B|u), U_0(|B|v) \rangle = \langle |B|u, |B|v \rangle, \quad \forall u, v \in \mathbb{C}^n$$

- ▶ Hence U_0 is isometric.
- ▶ From the definition of U_0 it is clear that U_0 maps $\mathcal{M}$ onto $\mathcal{N}$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.
- ▶ Extend $\{v_1, v_2, \dots, v_k\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$.

Continuation

- ▶ We can extend U_0 to a unitary from $\mathbb{C}^n$ to $\mathbb{C}^n$ as the dimensions of $\mathcal{M}^\perp$ and $\mathcal{N}^\perp$ are equal. More explicitly we can take the following steps:
- ▶ Let $v_1, v_2, \dots, v_k$ be an ortho-normal basis of $\mathcal{M}$. Then $U_0 v_1, U_0 v_2, \dots, U_0 v_k$ is an orthonormal basis of $\mathcal{N}$.
- ▶ Extend $\{v_1, v_2, \dots, v_k\}$ to an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{C}^n$.
- ▶ Similarly, extend $\{U_0 v_1, U_0 v_2, \dots, U_0 v_k\}$ to an orthonormal basis $\{U_0 v_1, U_0 v_2, \dots, U_0 v_k, w_1, w_2, \dots, w_{n-k}\}$ of $\mathbb{C}^n$.

Continuation

- Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.
- ▶ If B is invertible, and $B = UR$, then as $\det(B) = \det(U) \cdot \det(R) \neq 0$, R is also invertible and we get

$$U = BR^{-1} = B(|B|)^{-1}.$$

Continuation

- ▶ Now we extend U to a unitary of $\mathbb{C}^n$ by defining it on the orthonormal basis $\{v_1, \dots, v_n\}$ (and extending linearly) by setting

$$Uv_j = \begin{cases} U_0 v_j & \text{for } 1 \leq j \leq k; \\ w_{j-k} & \text{for } k+1 \leq j \leq n. \end{cases}.$$

- ▶ Clearly U is a unitary and $Uv = U_0 v$ for $v \in \mathcal{M}$.
- ▶ Now for $x \in \mathbb{C}^n$,

$$U|B|x = URx = U_0 Rx = Bx, \quad \forall x \in \mathbb{C}^n.$$

- ▶ This proves the existence. Now if $B = UR$ for some unitary U and positive R , $B^*B = R^*U^*UR = R^2$. Therefore R is the unique positive square root of B^*B , that is, $R = |B|$.
- ▶ If B is invertible, and $B = UR$, then as $\det(B) = \det(U) \cdot \det(R) \neq 0$, R is also invertible and we get

$$U = BR^{-1} = B(|B|)^{-1}.$$

- ▶ Therefore, U is uniquely determined. ■

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

- We can write D as

$$\begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} \cdot \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix}.$$

using the polar decompositions of $d_1, d_2, \dots, d_n$.

Polar decomposition for normal matrices

- **Example 28.2:** Suppose A is a normal matrix and let

$$A = UDU^*$$

be the diagonalization of A with unitary U and diagonal matrix D . Let $d_1, d_2, \dots, d_n$ be the diagonal entries of D .

- We can write D as

$$\begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} \cdot \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix}.$$

using the polar decompositions of $d_1, d_2, \dots, d_n$.

- Note that if $d_j = 0$, then when we write $d_j = e^{i\theta_j}|d_j|$, $e^{i\theta_j}$ is not unique.

Continuation

- Now $A = V|A|$ where

$$V = U \begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} U^*$$

is a unitary and

Continuation

- ▶ Now $A = V|A|$ where

$$V = U \begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix} U^*$$

is a unitary and



$$|A| = U \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix} U^*$$

is positive.

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .
- ▶ Take $V = e^{i\theta_1} P_1 + e^{i\theta_2} P_2 + \cdots + e^{i\theta_k} P_k$ and $|A| = |a_1| P_1 + |a_2| P_2 + \cdots + |a_k| P_k$.

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .
- ▶ Take $V = e^{i\theta_1} P_1 + e^{i\theta_2} P_2 + \cdots + e^{i\theta_k} P_k$ and $|A| = |a_1| P_1 + |a_2| P_2 + \cdots + |a_k| P_k$.
- ▶ Then $A = V|A|$ is the polar decomposition of A .

Continuation

- ▶ Alternatively, if $A = a_1 P_1 + a_2 P_2 + \cdots + a_k P_k$ is the spectral decomposition of A , we can get the polar decomposition as follows.
- ▶ Suppose $a_j = e^{i\theta_j} |a_j|$, $1 \leq j \leq k$ is the polar decomposition of a_j .
- ▶ Take $V = e^{i\theta_1} P_1 + e^{i\theta_2} P_2 + \cdots + e^{i\theta_k} P_k$ and $|A| = |a_1| P_1 + |a_2| P_2 + \cdots + |a_k| P_k$.
- ▶ Then $A = V|A|$ is the polar decomposition of A .
- ▶ **END OF REVIEW**

Singular value decomposition and singular values

- **Theorem 29.1:** Let B be an $n \times n$ complex matrix. Then there exist two unitaries V, W and a diagonal matrix S with non-negative entries such that

$$B = VSW.$$

Singular value decomposition and singular values

- ▶ **Theorem 29.1:** Let B be an $n \times n$ complex matrix. Then there exist two unitaries V, W and a diagonal matrix S with non-negative entries such that

$$B = VSW.$$

- ▶ **Proof:** By polar decomposition theorem,

$$B = U|B|$$

for some unitary U with $|B| = (B^*B)^{\frac{1}{2}}$.

Singular value decomposition and singular values

- ▶ **Theorem 29.1:** Let B be an $n \times n$ complex matrix. Then there exist two unitaries V, W and a diagonal matrix S with non-negative entries such that

$$B = VSW.$$

- ▶ **Proof:** By polar decomposition theorem,

$$B = U|B|$$

for some unitary U with $|B| = (B^*B)^{\frac{1}{2}}$.

- ▶ Now as $|B|$ is a normal operator there exists a unitary operator W and a diagonal matrix S such that

$$|B| = W^*SW.$$

Singular value decomposition and singular values

- ▶ **Theorem 29.1:** Let B be an $n \times n$ complex matrix. Then there exist two unitaries V, W and a diagonal matrix S with non-negative entries such that

$$B = VSW.$$

- ▶ **Proof:** By polar decomposition theorem,

$$B = U|B|$$

for some unitary U with $|B| = (B^*B)^{\frac{1}{2}}$.

- ▶ Now as $|B|$ is a normal operator there exists a unitary operator W and a diagonal matrix S such that

$$|B| = W^*SW.$$

- ▶ Taking, $V = UW^*$, V is a unitary and

$$B = VSW.$$

Singular value decomposition and singular values

- ▶ **Theorem 29.1:** Let B be an $n \times n$ complex matrix. Then there exist two unitaries V, W and a diagonal matrix S with non-negative entries such that

$$B = VSW.$$

- ▶ **Proof:** By polar decomposition theorem,

$$B = U|B|$$

for some unitary U with $|B| = (B^*B)^{\frac{1}{2}}$.

- ▶ Now as $|B|$ is a normal operator there exists a unitary operator W and a diagonal matrix S such that

$$|B| = W^*SW.$$

- ▶ Taking, $V = UW^*$, V is a unitary and

$$B = VSW.$$

- ▶ Note that as $|B|$ is positive, the diagonal entries of S are non-negative. ■

Continuation

- Definition 29.2: The decomposition of a matrix B as

$$B = VSW$$

where V, W are unitaries and S is diagonal with non-negative entries, is known as **singular value decomposition**. The diagonal entries of S are known as **singular values** of B .

Continuation

- Definition 29.2: The decomposition of a matrix B as

$$B = VSW$$

where V, W are unitaries and S is diagonal with non-negative entries, is known as **singular value decomposition**. The diagonal entries of S are known as **singular values** of B .

- The unitaries V, W in the singular value decomposition are not unique. However, as

$$B^*B = W^*S^*SW = W^*S^2W$$

the diagonal entries of S are eigenvalues of $|B|$. In other words, singular values of B are just eigenvalues of $|B|$.

Continuation

- Definition 29.2: The decomposition of a matrix B as

$$B = VSW$$

where V, W are unitaries and S is diagonal with non-negative entries, is known as **singular value decomposition**. The diagonal entries of S are known as **singular values** of B .

- The unitaries V, W in the singular value decomposition are not unique. However, as

$$B^*B = W^*S^*SW = W^*S^2W$$

the diagonal entries of S are eigenvalues of $|B|$. In other words, singular values of B are just eigenvalues of $|B|$.

- Note that,

$$BB^* = VSS^*V^* = VS^2V^*.$$

Continuation

- Definition 29.2: The decomposition of a matrix B as

$$B = VSW$$

where V, W are unitaries and S is diagonal with non-negative entries, is known as **singular value decomposition**. The diagonal entries of S are known as **singular values** of B .

- The unitaries V, W in the singular value decomposition are not unique. However, as

$$B^*B = W^*S^*SW = W^*S^2W$$

the diagonal entries of S are eigenvalues of $|B|$. In other words, singular values of B are just eigenvalues of $|B|$.

- Note that,

$$BB^* = VSS^*V^* = VS^2V^*.$$

- Therefore singular values of B are also eigenvalues of $(BB^*)^{\frac{1}{2}}$.

► Further,

$$BB^* = VS^2V^* = VW(B^*B)W^*V^*.$$

- ▶ Further,

$$BB^* = VS^2V^* = VW(B^*B)W^*V^*.$$

- ▶ Hence B^*B and BB^* are unitarily equivalent.

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

- ▶ Let $w_1, w_2, \dots, w_n$ be the columns of W^* and $v_1, v_2, \dots, v_n$ be the columns of V . Let $s_1, s_2, \dots, s_n$ be the diagonal entries of S .

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

- ▶ Let $w_1, w_2, \dots, w_n$ be the columns of W^* and $v_1, v_2, \dots, v_n$ be the columns of V . Let $s_1, s_2, \dots, s_n$ be the diagonal entries of S .
- ▶ Then the matrix equations above can be written as:

$$Bw_j = s_jv_j, \quad B^*v_j = s_jw_j, \quad 1 \leq j \leq n.$$

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

- ▶ Let $w_1, w_2, \dots, w_n$ be the columns of W^* and $v_1, v_2, \dots, v_n$ be the columns of V . Let $s_1, s_2, \dots, s_n$ be the diagonal entries of S .
- ▶ Then the matrix equations above can be written as:

$$Bw_j = s_jv_j, \quad B^*v_j = s_jw_j, \quad 1 \leq j \leq n.$$

- ▶ So we have pairs of vectors, which behave somewhat like eigenvectors.

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

- ▶ Let $w_1, w_2, \dots, w_n$ be the columns of W^* and $v_1, v_2, \dots, v_n$ be the columns of V . Let $s_1, s_2, \dots, s_n$ be the diagonal entries of S .
- ▶ Then the matrix equations above can be written as:

$$Bw_j = s_jv_j, \quad B^*v_j = s_jw_j, \quad 1 \leq j \leq n.$$

- ▶ So we have pairs of vectors, which behave somewhat like eigenvectors.
- ▶ It is possible to have singular value decomposition for non-square matrices. But we will not be considering that.

Eigenvector type equations

- ▶ Suppose $B = VSW$ is a singular value decomposition of B .
- ▶ Then, $B^* = W^*SV^*$ and so

$$BW^* = VS, \quad B^*V = W^*S.$$

- ▶ Let $w_1, w_2, \dots, w_n$ be the columns of W^* and $v_1, v_2, \dots, v_n$ be the columns of V . Let $s_1, s_2, \dots, s_n$ be the diagonal entries of S .
- ▶ Then the matrix equations above can be written as:

$$Bw_j = s_jv_j, \quad B^*v_j = s_jw_j, \quad 1 \leq j \leq n.$$

- ▶ So we have pairs of vectors, which behave somewhat like eigenvectors.
- ▶ It is possible to have singular value decomposition for non-square matrices. But we will not be considering that.
- ▶ The notion of SVD has many practical applications.

Normal matrices

- ▶ If A is normal, then by diagonalization,

$$A = UDU^*$$

for some unitary U and diagonal matrix D .

Normal matrices

- ▶ If A is normal, then by diagonalization,

$$A = UDU^*$$

for some unitary U and diagonal matrix D .

- ▶ Let $d_j = e^{i\theta_j}|d_j|$ be the diagonal entries of D , $1 \leq j \leq n$.

Normal matrices

- ▶ If A is normal, then by diagonalization,

$$A = UDU^*$$

for some unitary U and diagonal matrix D .

- ▶ Let $d_j = e^{i\theta_j}|d_j|$ be the diagonal entries of D , $1 \leq j \leq n$.
- ▶ Then the singular value decomposition of A is given by

$$A = V_1 \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix} U^*,$$

Normal matrices

- ▶ If A is normal, then by diagonalization,

$$A = UDU^*$$

for some unitary U and diagonal matrix D .

- ▶ Let $d_j = e^{i\theta_j}|d_j|$ be the diagonal entries of D , $1 \leq j \leq n$.
- ▶ Then the singular value decomposition of A is given by

$$A = V_1 \begin{bmatrix} |d_1| & 0 & \dots & 0 \\ 0 & |d_2| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |d_n| \end{bmatrix} U^*,$$

- ▶ where

$$V_1 = U \begin{bmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{i\theta_n} \end{bmatrix}.$$

Example

- ▶ We revisit an example we saw before.

Example

- ▶ We revisit an example we saw before.
- ▶ Example 28.3 and 29.4: Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

Example

- ▶ We revisit an example we saw before.
- ▶ **Example 28.3 and 29.4:** Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

- ▶ Recall that an $n \times n$ matrix with all entries equal to $\frac{1}{n}$ is a projection.

Example

- ▶ We revisit an example we saw before.
- ▶ **Example 28.3 and 29.4:** Consider

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Then

$$B^*B = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

- ▶ Recall that an $n \times n$ matrix with all entries equal to $\frac{1}{n}$ is a projection.
- ▶ Therefore

$$B^*B = 12 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} + 0 \cdot \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}$$

is the spectral decomposition of B^*B .

► Therefore,

$$\begin{aligned} |B| &= \sqrt{12} \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \\ &= \begin{bmatrix} 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \end{bmatrix}. \end{aligned}$$

Continuation

► Take

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

Continuation

- ▶ Take

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

- ▶ We may now verify that

$$B = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix} \cdot \begin{bmatrix} 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 2/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \end{bmatrix}.$$

is a polar decomposition of B .

Continuation



$$B = I \cdot \begin{bmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix},$$

is a singular value decomposition of B .

Continuation



$$B = I \cdot \begin{bmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix},$$

is a singular value decomposition of B .



Continuation



$$B = I. \begin{bmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix},$$

is a singular value decomposition of B .



▶ END OF LECTURE 29.

Lecture 30: Balanced incomplete block designs (BIBD)

- ▶ Consider an agricultural experiment where we wish to determine the best combination of fertilizers and other inputs.

Lecture 30: Balanced incomplete block designs (BIBD)

- ▶ Consider an agricultural experiment where we wish to determine the best combination of fertilizers and other inputs.
- ▶ Suppose out of 5 inputs at a time we can try out only 3 at a time.

Lecture 30: Balanced incomplete block designs (BIBD)

- ▶ Consider an agricultural experiment where we wish to determine the best combination of fertilizers and other inputs.
- ▶ Suppose out of 5 inputs at a time we can try out only 3 at a time.
- ▶ Two different inputs may interact with each other.

Lecture 30: Balanced incomplete block designs (BIBD)

- ▶ Consider an agricultural experiment where we wish to determine the best combination of fertilizers and other inputs.
- ▶ Suppose out of 5 inputs at a time we can try out only 3 at a time.
- ▶ Two different inputs may interact with each other.
- ▶ In such situation it becomes helpful to use some combinatorial structures called balanced incomplete block designs.

- **Definition 30.1:** An (v, b, r, k, λ) design is an arrangement of v symbols (called varieties or treatments or factors) into b blocks as follows:

- ▶ **Definition 30.1:** An (v, b, r, k, λ) design is an arrangement of v symbols (called varieties or treatments or factors) into b blocks as follows:
- ▶ (i) Each block has exactly k symbols (with no repetitions).

- ▶ **Definition 30.1:** An (v, b, r, k, λ) design is an arrangement of v symbols (called varieties or treatments or factors) into b blocks as follows:
 - ▶ (i) Each block has exactly k symbols (with no repetitions).
 - ▶ (ii) No symbol is repeated in any block.
 - ▶ (iii) Each symbol appears in exactly r blocks.

- ▶ **Definition 30.1:** An (v, b, r, k, λ) design is an arrangement of v symbols (called varieties or treatments or factors) into b blocks as follows:
 - ▶ (i) Each block has exactly k symbols (with no repetitions).
 - ▶ (ii) No symbol is repeated in any block.
 - ▶ (iii) Each symbol appears in exactly r blocks.
 - ▶ (iv) Each pair of symbols appear in exactly λ blocks.

- ▶ **Definition 30.1:** An (v, b, r, k, λ) design is an arrangement of v symbols (called varieties or treatments or factors) into b blocks as follows:
 - ▶ (i) Each block has exactly k symbols (with no repetitions).
 - ▶ (ii) No symbol is repeated in any block.
 - ▶ (iii) Each symbol appears in exactly r blocks.
 - ▶ (iv) Each pair of symbols appear in exactly λ blocks.
- ▶ Here v, b, r, k, λ are natural numbers. It is also assumed $k < v$ (No block contains all the treatments). For this reason they are called incomplete designs.

Example 1

- ▶ Example 30.1:

Example 1

- ▶ Example 30.1:
- ▶ Here is a $(4, 4, 3, 3, 2)$ BIBD with symbol set $V = \{a, b, c, d\}$.

Example 1

- ▶ Example 30.1:
- ▶ Here is a $(4, 4, 3, 3, 2)$ BIBD with symbol set $V = \{a, b, c, d\}$.
- ▶ Block 1: $\{a, b, c\}$ Block 2: $\{b, c, d\}$ Block 3: $\{c, d, a\}$ Block 4: $\{d, a, b\}$.

Example 1

- ▶ Example 30.1:
- ▶ Here is a $(4, 4, 3, 3, 2)$ BIBD with symbol set $V = \{a, b, c, d\}$.
- ▶ Block 1: $\{a, b, c\}$ Block 2: $\{b, c, d\}$ Block 3: $\{c, d, a\}$ Block 4: $\{d, a, b\}$.
- ▶ Observe: There are 4 (v) symbols and 4 (b) blocks.

Example 1

- ▶ Example 30.1:
- ▶ Here is a $(4, 4, 3, 3, 2)$ BIBD with symbol set $V = \{a, b, c, d\}$.
- ▶ Block 1: $\{a, b, c\}$ Block 2: $\{b, c, d\}$ Block 3: $\{c, d, a\}$ Block 4: $\{d, a, b\}$.
- ▶ Observe: There are 4 (v) symbols and 4 (b) blocks.
- ▶ Each symbol appears in 3 (r) blocks and each block has size 3 (k).

Example 1

- ▶ Example 30.1:
- ▶ Here is a $(4, 4, 3, 3, 2)$ BIBD with symbol set $V = \{a, b, c, d\}$.
- ▶ Block 1: $\{a, b, c\}$ Block 2: $\{b, c, d\}$ Block 3: $\{c, d, a\}$ Block 4: $\{d, a, b\}$.
- ▶ Observe: There are 4 (v) symbols and 4 (b) blocks.
- ▶ Each symbol appears in 3 (r) blocks and each block has size 3 (k).
- ▶ Each pair of symbols appears in 2 (λ) blocks.

Example 2

- ▶ Example 30.2:

Example 2

- ▶ Example 30.2:
- ▶ Here is a $(8, 14, 7, 4, 3)$ BIBD with symbol set $V = \{1, 2, \dots, 8\}$:

Example 2

- ▶ Example 30.2:
- ▶ Here is a $(8, 14, 7, 4, 3)$ BIBD with symbol set $V = \{1, 2, \dots, 8\}$:
- ▶ Blocks:

$\{1, 3, 7, 8\}, \{1, 2, 4, 8\}, \{2, 3, 5, 8\}, \{3, 4, 6, 8\}, \{4, 5, 7, 8\},$

$\{2, 6, 7, 8\}, \{1, 2, 3, 6\}, \{1, 2, 5, 7\}, \{1, 3, 4, 5\}, \{1, 4, 6, 7\},$

$\{2, 4, 5, 6\}, \{3, 5, 6, 7\}, \{1, 5, 6, 8\}, \{2, 3, 4, 7\}.$

Basic relations between parameters

- **Theorem 30.4:** Suppose there exists a (v, b, r, k, λ) BIBD.
Then

$$vr = bk, \quad r(k - 1) = \lambda(v - 1).$$

Basic relations between parameters

- **Theorem 30.4:** Suppose there exists a (v, b, r, k, λ) BIBD.
Then

$$vr = bk, \quad r(k - 1) = \lambda(v - 1).$$

- (i) $vr = bk$. This is clear by counting the number of elements in the whole design in two different ways. There are v symbols and each one appears r times. So there are vr elements in the design. Alternatively, there are b blocks with each of size k and hence the total number of elements is bk . So we get the equality.

Basic relations between parameters

- ▶ **Theorem 30.4:** Suppose there exists a (v, b, r, k, λ) BIBD. Then

$$vr = bk, \quad r(k - 1) = \lambda(v - 1).$$

- ▶ (i) $vr = bk$. This is clear by counting the number of elements in the whole design in two different ways. There are v symbols and each one appears r times. So there are vr elements in the design. Alternatively, there are b blocks with each of size k and hence the total number of elements is bk . So we get the equality.
- ▶ (ii) Say the symbols are $\{1, 2, \dots, v\}$. Consider how many times $(1, 2)$ appears in a block. It appears λ times. Similarly $(1, 3)$ appears λ times. So $(1, j)$ appears for some j , a total of $\lambda(v - 1)$ times.

Basic relations between parameters

- ▶ **Theorem 30.4:** Suppose there exists a (v, b, r, k, λ) BIBD. Then

$$vr = bk, \quad r(k - 1) = \lambda(v - 1).$$

- ▶ (i) $vr = bk$. This is clear by counting the number of elements in the whole design in two different ways. There are v symbols and each one appears r times. So there are vr elements in the design. Alternatively, there are b blocks with each of size k and hence the total number of elements is bk . So we get the equality.
- ▶ (ii) Say the symbols are $\{1, 2, \dots, v\}$. Consider how many times $(1, 2)$ appears in a block. It appears λ times. Similarly $(1, 3)$ appears λ times. So $(1, j)$ appears for some j , a total of $\lambda(v - 1)$ times.
- ▶ Since 1 has appeared in exactly r blocks, and each block has $(k - 1)$ other elements, we have $(1, j)$ for some j appearing $r(k - 1)$ times. This gives $r(k - 1) = \lambda(v - 1)$.

Fisher's inequality

- ▶ **Theorem (Fisher's inequality) 30.5:** Suppose there exists an (v, b, k, r, λ) design. Then $b \geq v$.

Fisher's inequality

- ▶ **Theorem (Fisher's inequality) 30.5:** Suppose there exists an (v, b, k, r, λ) design. Then $b \geq v$.
- ▶ This is an important inequality and it is non-trivial!

Incidence matrix

- ▶ Consider an (v, b, r, k, λ) BIBD.

Incidence matrix

- ▶ Consider an (v, b, r, k, λ) BIBD.
- ▶ Let the symbols be $\{1, 2, \dots, v\}$. Let the blocks be $B_1, B_2, \dots, B_b$.

Incidence matrix

- ▶ Consider an (v, b, r, k, λ) BIBD.
- ▶ Let the symbols be $\{1, 2, \dots, v\}$. Let the blocks be $B_1, B_2, \dots, B_b$.
- ▶ The **incidence matrix** of the design is a matrix N of order $v \times b$. Put

$$n_{ij} = \begin{cases} 1 & \text{if } i \in B_j \\ 0 & \text{otherwise} \end{cases}$$

Incidence matrix

- ▶ Consider an (v, b, r, k, λ) BIBD.
- ▶ Let the symbols be $\{1, 2, \dots, v\}$. Let the blocks be $B_1, B_2, \dots, B_b$.
- ▶ The **incidence matrix** of the design is a matrix N of order $v \times b$. Put

$$n_{ij} = \begin{cases} 1 & \text{if } i \in B_j \\ 0 & \text{otherwise} \end{cases}$$

- ▶ In other words, $n_{ij} = 1$ if i appears in the block B_j and it is zero otherwise.

Example

- ▶ Replacing symbols a, b, c, d by 1, 2, 3, 4 respectively in the first example:

Example

- ▶ Replacing symbols a, b, c, d by $1, 2, 3, 4$ respectively in the first example:
- ▶ $B_1 = \{1, 2, 3\}, B_2 = \{2, 3, 4\}, B_3 = \{1, 3, 4\}, B_4 = \{1, 2, 4\}$.

Example

- ▶ Replacing symbols a, b, c, d by $1, 2, 3, 4$ respectively in the first example:
- ▶ $B_1 = \{1, 2, 3\}, B_2 = \{2, 3, 4\}, B_3 = \{1, 3, 4\}, B_4 = \{1, 2, 4\}$.
- ▶ The corresponding incidence matrix is given by:

Example

- ▶ Replacing symbols a, b, c, d by 1, 2, 3, 4 respectively in the first example:
- ▶ $B_1 = \{1, 2, 3\}, B_2 = \{2, 3, 4\}, B_3 = \{1, 3, 4\}, B_4 = \{1, 2, 4\}$.
- ▶ The corresponding incidence matrix is given by:
- ▶

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

Example

- ▶ Replacing symbols a, b, c, d by 1, 2, 3, 4 respectively in the first example:
- ▶ $B_1 = \{1, 2, 3\}, B_2 = \{2, 3, 4\}, B_3 = \{1, 3, 4\}, B_4 = \{1, 2, 4\}$.
- ▶ The corresponding incidence matrix is given by:



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

- ▶ Note that

$$N^t = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}, \quad NN^t = \begin{bmatrix} 3 & 2 & 2 & 2 \\ 2 & 3 & 2 & 2 \\ 2 & 2 & 3 & 2 \\ 2 & 2 & 2 & 3 \end{bmatrix}.$$

Continuation

- **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

Continuation

- ▶ **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

- ▶ where J is the matrix of all entries equal to 1 and of order $v \times v$.

Continuation

- ▶ **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

- ▶ where J is the matrix of all entries equal to 1 and of order $v \times v$.
- ▶ **Proof:** The $(i, j)^{\text{th}}$ entry of NN^t is the inner product between row i and row j .

Continuation

- ▶ **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

- ▶ where J is the matrix of all entries equal to 1 and of order $v \times v$.
- ▶ **Proof:** The $(i, j)^{\text{th}}$ entry of NN^t is the inner product between row i and row j .
- ▶ Since every entry is 0 or 1, when $i = j$, $(NN^t)_{ii}$ is just the number of times the treatment i appears in the design. So

$$(NN^t)_{ii} = r.$$

Continuation

- ▶ **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

- ▶ where J is the matrix of all entries equal to 1 and of order $v \times v$.
- ▶ **Proof:** The $(i, j)^{\text{th}}$ entry of NN^t is the inner product between row i and row j .
- ▶ Since every entry is 0 or 1, when $i = j$, $(NN^t)_{ii}$ is just the number of times the treatment i appears in the design. So

$$(NN^t)_{ii} = r.$$

- ▶ Similarly, when $i \neq j$, $(NN^t)_{ij}$ is the number of times the pair of treatments $\{i, j\}$ appears in the design. Hence

$$(NN^t)_{ij} = \lambda, \quad i \neq j.$$

Continuation

- ▶ **Theorem 30.6:** Suppose N is the incidence matrix of a (v, b, r, k, λ) BIBD. Then

$$(NN^t) = (r - \lambda)I + \lambda J$$

- ▶ where J is the matrix of all entries equal to 1 and of order $v \times v$.
- ▶ **Proof:** The $(i, j)^{\text{th}}$ entry of NN^t is the inner product between row i and row j .
- ▶ Since every entry is 0 or 1, when $i = j$, $(NN^t)_{ii}$ is just the number of times the treatment i appears in the design. So

$$(NN^t)_{ii} = r.$$

- ▶ Similarly, when $i \neq j$, $(NN^t)_{ij}$ is the number of times the pair of treatments $\{i, j\}$ appears in the design. Hence

$$(NN^t)_{ij} = \lambda, \quad i \neq j.$$

- ▶ Hence $(NN^t) = (r - \lambda)I + \lambda J$ ■

Continuation

- **Proof of Fisher's inequality:** Let P be the projection $\frac{1}{v}J$. Then

$$\begin{aligned} NN^t &= (r - \lambda)I + \lambda J \\ &= (r - \lambda)I + v\lambda P \\ &= (r - \lambda)(P + P^\perp) + v\lambda P \\ &= (r - \lambda)P^\perp + (r - \lambda + v\lambda)P. \end{aligned}$$

Continuation

- **Proof of Fisher's inequality:** Let P be the projection $\frac{1}{v}J$. Then

$$\begin{aligned} NN^t &= (r - \lambda)I + \lambda J \\ &= (r - \lambda)I + v\lambda P \\ &= (r - \lambda)(P + P^\perp) + v\lambda P \\ &= (r - \lambda)P^\perp + (r - \lambda + v\lambda)P. \end{aligned}$$

- In particular the eigenvalues of NN^t are $(r - \lambda)$ and $r - \lambda + v\lambda$.

Continuation

- **Proof of Fisher's inequality:** Let P be the projection $\frac{1}{v}J$. Then

$$\begin{aligned} NN^t &= (r - \lambda)I + \lambda J \\ &= (r - \lambda)I + v\lambda P \\ &= (r - \lambda)(P + P^\perp) + v\lambda P \\ &= (r - \lambda)P^\perp + (r - \lambda + v\lambda)P. \end{aligned}$$

- In particular the eigenvalues of NN^t are $(r - \lambda)$ and $r - \lambda + v\lambda$.
- Recall that r is the number of times a particular treatment appears in the design and λ is the number of times a pair of treatments appears.

Continuation

- **Proof of Fisher's inequality:** Let P be the projection $\frac{1}{v}J$. Then

$$\begin{aligned} NN^t &= (r - \lambda)I + \lambda J \\ &= (r - \lambda)I + v\lambda P \\ &= (r - \lambda)(P + P^\perp) + v\lambda P \\ &= (r - \lambda)P^\perp + (r - \lambda + v\lambda)P. \end{aligned}$$

- In particular the eigenvalues of NN^t are $(r - \lambda)$ and $r - \lambda + v\lambda$.
- Recall that r is the number of times a particular treatment appears in the design and λ is the number of times a pair of treatments appears.
- Further, a block can't have all the treatments ($k < v$). So we clearly have $r > \lambda$.

Continuation

- **Proof of Fisher's inequality:** Let P be the projection $\frac{1}{v}J$. Then

$$\begin{aligned} NN^t &= (r - \lambda)I + \lambda J \\ &= (r - \lambda)I + v\lambda P \\ &= (r - \lambda)(P + P^\perp) + v\lambda P \\ &= (r - \lambda)P^\perp + (r - \lambda + v\lambda)P. \end{aligned}$$

- In particular the eigenvalues of NN^t are $(r - \lambda)$ and $r - \lambda + v\lambda$.
- Recall that r is the number of times a particular treatment appears in the design and λ is the number of times a pair of treatments appears.
- Further, a block can't have all the treatments ($k < v$). So we clearly have $r > \lambda$.
- Consequently $(r - \lambda) > 0$ and $(r - \lambda + v\lambda) > 0$.

- ▶ This in particular means that the eigenvalues of NN^t are non-zero.

Continuation

- ▶ This in particular means that the eigenvalues of NN^t are non-zero.
- ▶ Therefore the rank of NN^t is ν . Hence the rank of N is at least ν .

Continuation

- ▶ This in particular means that the eigenvalues of NN^t are non-zero.
- ▶ Therefore the rank of NN^t is ν . Hence the rank of N is at least ν .
- ▶ Since N is of order $\nu \times b$, we have $\text{rank}(N) \leq \min\{\nu, b\}$.

Continuation

- ▶ This in particular means that the eigenvalues of NN^t are non-zero.
- ▶ Therefore the rank of NN^t is v . Hence the rank of N is at least v .
- ▶ Since N is of order $v \times b$, we have $\text{rank}(N) \leq \min\{v, b\}$.
- ▶ This proves $b \geq v$. ■.

Continuation

- ▶ This in particular means that the eigenvalues of NN^t are non-zero.
- ▶ Therefore the rank of NN^t is v . Hence the rank of N is at least v .
- ▶ Since N is of order $v \times b$, we have $\text{rank}(N) \leq \min\{v, b\}$.
- ▶ This proves $b \geq v$. ■.
- ▶ **END OF LECTURE 30.**

Lecture 31: Determining positivity

- ▶ Once again we recall the notion of positivity and some characterizations.

Lecture 31: Determining positivity

- ▶ Once again we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .

Lecture 31: Determining positivity

- ▶ Once again we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.

Lecture 31: Determining positivity

- ▶ Once again we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.

Lecture 31: Determining positivity

- ▶ Once again we recall the notion of positivity and some characterizations.
- ▶ **Definition 26.1:** An $n \times n$ matrix A is said to be a **positive** matrix if $A = B^*B$ for some $n \times n$ matrix B .
- ▶ Some authors may call these as non-negative definite matrices and invertible matrices of the form B^*B as positive definite matrices.
- ▶ **Warning:** A positive matrix need not have positive entries. It can have negative entries and also complex entries.
- ▶ Matrices whose entries are positive would be called as **entrywise positive** matrices. That is also an important class, but we will not be studying them now.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^* B$ for some $n \times n$ matrix B .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.

Characterizations of positivity

- ▶ **Theorem 26.2:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive, that is, $A = B^*B$ for some $n \times n$ matrix B .
 - ▶ (ii) $A = C^*C$ for some $m \times n$ matrix for some m .
 - ▶ (iii) $a_{ij} = \langle v_i, v_j \rangle$, $1 \leq i, j \leq n$ for vectors $v_1, v_2, \dots, v_n$ in some inner product space V .
 - ▶ (iv) $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^n$.
 - ▶ (v) $A = A^*$ and eigenvalues of A are non-negative.
 - ▶ (vi) $A = S^2$ for some self-adjoint $n \times n$ matrix S .

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.
- ▶ We have already seen that if A is positive then $A = A^*$. Self-adjointness is a necessity for positivity and it is easy to check.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.
- ▶ We have already seen that if A is positive then $A = A^*$. Self-adjointness is a necessity for positivity and it is easy to check.
- ▶ **Theorem 31.1:** Suppose A is a positive matrix and $a_{ii} = 0$ for some i . Then $a_{ij} = 0 = a_{ji}$ for all j , that is, if a diagonal entry is zero then all entries in corresponding row and column are zeros.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.
- ▶ We have already seen that if A is positive then $A = A^*$. Self-adjointness is a necessity for positivity and it is easy to check.
- ▶ **Theorem 31.1:** Suppose A is a positive matrix and $a_{ii} = 0$ for some i . Then $a_{ij} = 0 = a_{ji}$ for all j , that is, if a diagonal entry is zero then all entries in corresponding row and column are zeros.
- ▶ **Proof:** We write A as a Gram matrix of some vectors $v_1, v_2, \dots, v_n$.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.
- ▶ We have already seen that if A is positive then $A = A^*$. Self-adjointness is a necessity for positivity and it is easy to check.
- ▶ **Theorem 31.1:** Suppose A is a positive matrix and $a_{ii} = 0$ for some i . Then $a_{ij} = 0 = a_{ji}$ for all j , that is, if a diagonal entry is zero then all entries in corresponding row and column are zeros.
- ▶ **Proof:** We write A as a Gram matrix of some vectors $v_1, v_2, \dots, v_n$.
- ▶ Now $a_{ii} = 0$ implies $\langle v_i, v_i \rangle = 0$ and hence $v_i = 0$.

Diagonal entries

- ▶ It is difficult to determine positivity of a matrix using the characterizations given above.
- ▶ Let us first look at some necessary conditions for positivity which can be checked easily.
- ▶ We have already seen that if A is positive then $A = A^*$. Self-adjointness is a necessity for positivity and it is easy to check.
- ▶ **Theorem 31.1:** Suppose A is a positive matrix and $a_{ii} = 0$ for some i . Then $a_{ij} = 0 = a_{ji}$ for all j , that is, if a diagonal entry is zero then all entries in corresponding row and column are zeros.
- ▶ **Proof:** We write A as a Gram matrix of some vectors $v_1, v_2, \dots, v_n$.
- ▶ Now $a_{ii} = 0$ implies $\langle v_i, v_i \rangle = 0$ and hence $v_i = 0$.
- ▶ Consequently, $a_{ij} = \langle v_i, v_j \rangle = 0$ for any j . Similarly $a_{ji} = 0$ for any j .

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
- ▶ (i) A is positive definite;

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;
 - ▶ (iii) A is self-adjoint and all its eigenvalues are strictly positive.

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;
 - ▶ (iii) A is self-adjoint and all its eigenvalues are strictly positive.

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;
 - ▶ (iii) A is self-adjoint and all its eigenvalues are strictly positive.
 - ▶ (iv) A is a Gram matrix of linearly independent vectors in some inner product space.

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;
 - ▶ (iii) A is self-adjoint and all its eigenvalues are strictly positive.
 - ▶ (iv) A is a Gram matrix of linearly independent vectors in some inner product space.
 - ▶ (v) $\langle x, Ax \rangle > 0$ for all $x \neq 0$.

Strict positivity

- ▶ Recall that a matrix A is said to be **positive definite** or **strictly positive** if A is positive and invertible.
- ▶ **Theorem 31.2:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive definite;
 - ▶ (ii) A is positive and $\det(A) \neq 0$;
 - ▶ (iii) A is self-adjoint and all its eigenvalues are strictly positive.
 - ▶ (iv) A is a Gram matrix of linearly independent vectors in some inner product space.
 - ▶ (v) $\langle x, Ax \rangle > 0$ for all $x \neq 0$.
- ▶ **Proof:** Easy.

Principal submatrices and principal minors

- **Definition 31.3:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then for any non-empty subset S of $\{1, 2, \dots, n\}$, the matrix $A_S := [a_{ij}]_{i, j \in S}$ is called the **principal submatrix** of A corresponding to S . The determinant of A_S is called the **principal minor** of A corresponding to the subset S . The principal minors corresponding to sets of the form $\{1, 2, \dots, k\}$ for $1 \leq k \leq n$ are known as **leading principal minors**.

Principal submatrices and principal minors

- ▶ **Definition 31.3:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then for any non-empty subset S of $\{1, 2, \dots, n\}$, the matrix $A_S := [a_{ij}]_{i, j \in S}$ is called the **principal submatrix** of A corresponding to S . The determinant of A_S is called the **principal minor** of A corresponding to the subset S . The principal minors corresponding to sets of the form $\{1, 2, \dots, k\}$ for $1 \leq k \leq n$ are known as **leading principal minors**.
- ▶ **Theorem 31.4:** Let A be a positive matrix. Then all its principal minors are positive. If A is strictly positive then all its principal minors are strictly positive.

Principal submatrices and principal minors

- ▶ **Definition 31.3:** Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a complex matrix. Then for any non-empty subset S of $\{1, 2, \dots, n\}$, the matrix $A_S := [a_{ij}]_{i, j \in S}$ is called the **principal submatrix** of A corresponding to S . The determinant of A_S is called the **principal minor** of A corresponding to the subset S . The principal minors corresponding to sets of the form $\{1, 2, \dots, k\}$ for $1 \leq k \leq n$ are known as **leading principal minors**.
- ▶ **Theorem 31.4:** Let A be a positive matrix. Then all its principal minors are positive. If A is strictly positive then all its principal minors are strictly positive.
- ▶ **Proof:** Exercise. (Hint: Use Gram matrices.)

Leading principal minors

- ▶ **Theorem 31.5:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Leading principal minors

- ▶ **Theorem 31.5:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is strictly positive;

Leading principal minors

- ▶ **Theorem 31.5:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is strictly positive;
 - ▶ (ii) A is self-adjoint and all leading principal minors of A are strictly positive.

Leading principal minors

- ▶ **Theorem 31.5:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is strictly positive;
 - ▶ (ii) A is self-adjoint and all leading principal minors of A are strictly positive.
- ▶ **Proof:** We have already seen $(i) \Rightarrow (ii)$.

Leading principal minors

- ▶ **Theorem 31.5:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is strictly positive;
 - ▶ (ii) A is self-adjoint and all leading principal minors of A are strictly positive.
- ▶ **Proof:** We have already seen $(i) \Rightarrow (ii)$.
- ▶ To see the converse, we use induction on n .

Continuation

- ▶ There is nothing to verify when $n = 1$.

Continuation

- ▶ There is nothing to verify when $n = 1$.
- ▶ Now for $n \geq 2$, assume the result for $(n - 1)$ and we will prove it for n .

Continuation

- ▶ There is nothing to verify when $n = 1$.
- ▶ Now for $n \geq 2$, assume the result for $(n - 1)$ and we will prove it for n .
- ▶ Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a matrix with all its leading principal minors strictly positive.

Continuation

- ▶ There is nothing to verify when $n = 1$.
- ▶ Now for $n \geq 2$, assume the result for $(n - 1)$ and we will prove it for n .
- ▶ Let $A = [a_{ij}]_{1 \leq i, j \leq n}$ be a matrix with all its leading principal minors strictly positive.
- ▶ In particular $a_{11} > 0$.

Continuation

- ▶ We consider a block decomposition of A as

$$A = \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix}$$

Continuation

- ▶ We consider a block decomposition of A as

$$A = \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix}$$

- ▶ where,

$$y = \begin{pmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad B = \begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}.$$

Continuation

- ▶ We consider a block decomposition of A as

$$A = \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix}$$

- ▶ where,

$$y = \begin{pmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad B = \begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}.$$

- ▶ Using the fact that $a_{11} > 0$, we have

$$\begin{bmatrix} 1 & 0 \\ -\frac{y}{a_{11}} & I \end{bmatrix} \cdot \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{y^*}{a_{11}} \\ 0 & I \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix},$$

Continuation

- ▶ We consider a block decomposition of A as

$$A = \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix}$$

- ▶ where,

$$y = \begin{pmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad B = \begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}.$$

- ▶ Using the fact that $a_{11} > 0$, we have

$$\begin{bmatrix} 1 & 0 \\ -\frac{y}{a_{11}} & I \end{bmatrix} \cdot \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{y^*}{a_{11}} \\ 0 & I \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix},$$

- ▶ where, $C = B - \frac{1}{a_{11}}yy^*$. (Recall 'Schur-complement'.)

Continuation

- ▶ We consider a block decomposition of A as

$$A = \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix}$$

- ▶ where,

$$y = \begin{pmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad B = \begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}.$$

- ▶ Using the fact that $a_{11} > 0$, we have

$$\begin{bmatrix} 1 & 0 \\ -\frac{y}{a_{11}} & I \end{bmatrix} \cdot \begin{bmatrix} a_{11} & y^* \\ y & B \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{y^*}{a_{11}} \\ 0 & I \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix},$$

- ▶ where, $C = B - \frac{1}{a_{11}}yy^*$. (Recall 'Schur-complement'.)
- ▶ Taking determinant, we get

$$\det(A) = a_{11} \cdot \det(C).$$

Continuation

- ▶ In particular $\det(C) > 0$.
- ▶ Instead of starting with A , if we start with leading principal submatrices of A and do similar computation we see that leading principal minors of C are all strictly positive.

Continuation

- ▶ In particular $\det(C) > 0$.
- ▶ Instead of starting with A , if we start with leading principal submatrices of A and do similar computation we see that leading principal minors of C are all strictly positive.
- ▶ Clearly C is self-adjoint. Hence by induction hypothesis C is strictly positive.

Continuation

- ▶ In particular $\det(C) > 0$.
- ▶ Instead of starting with A , if we start with leading principal submatrices of A and do similar computation we see that leading principal minors of C are all strictly positive.
- ▶ Clearly C is self-adjoint. Hence by induction hypothesis C is strictly positive.
- ▶ Then it is easy to see that,

$$\begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix}$$

is also strictly positive.

Continuation

- ▶ In particular $\det(C) > 0$.
- ▶ Instead of starting with A , if we start with leading principal submatrices of A and do similar computation we see that leading principal minors of C are all strictly positive.
- ▶ Clearly C is self-adjoint. Hence by induction hypothesis C is strictly positive.
- ▶ Then it is easy to see that,

$$\begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix}$$

is also strictly positive.

- ▶ Further, as

$$A = \begin{bmatrix} 1 & 0 \\ \frac{y}{a_{11}} & I \end{bmatrix} \cdot \begin{bmatrix} a_{11} & 0 \\ 0 & C \end{bmatrix} \cdot \begin{bmatrix} 1 & \frac{y^*}{a_{11}} \\ 0 & I \end{bmatrix},$$

A is positive. Since its determinant is non-zero it is strictly positive. ■

- Example 31.6: Consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$$

- ▶ Example 31.6: Consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$$

- ▶ Here A is self-adjoint and all its leading principal minors are non-negative.

- ▶ Example 31.6: Consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$$

- ▶ Here A is self-adjoint and all its leading principal minors are non-negative.
- ▶ However A is not positive as a diagonal entry is negative.

- ▶ Example 31.6: Consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$$

- ▶ Here A is self-adjoint and all its leading principal minors are non-negative.
- ▶ However A is not positive as a diagonal entry is negative.
- ▶ In other words, the previous theorem does not hold if strict positivity is replaced by positivity.

- ▶ Example 31.6: Consider

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}$$

- ▶ Here A is self-adjoint and all its leading principal minors are non-negative.
- ▶ However A is not positive as a diagonal entry is negative.
- ▶ In other words, the previous theorem does not hold if strict positivity is replaced by positivity.
- ▶ To get the correct result for positivity we need to consider all principal minors instead of just leading principal minors.

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive;

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive;
 - ▶ (ii) A is self-adjoint and all principal minors of A are positive.

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive;
 - ▶ (ii) A is self-adjoint and all principal minors of A are positive.
- ▶ **Proof:** We have already seen $(i) \Rightarrow (ii)$.

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive;
 - ▶ (ii) A is self-adjoint and all principal minors of A are positive.
- ▶ **Proof:** We have already seen $(i) \Rightarrow (ii)$.
- ▶ Like before to show $(ii) \Rightarrow (i)$, we use induction.

Principal minor condition for positivity

- ▶ **Theorem 31.7:** Let A be an $n \times n$ complex matrix. Then the following are equivalent:
 - ▶ (i) A is positive;
 - ▶ (ii) A is self-adjoint and all principal minors of A are positive.
- ▶ **Proof:** We have already seen $(i) \Rightarrow (ii)$.
- ▶ Like before to show $(ii) \Rightarrow (i)$, we use induction.
- ▶ There is nothing to show for $n = 1$.

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.
- ▶ If $a_{11} > 0$, we can essentially repeat the argument of previous theorem to deduce that A is positive.

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.
- ▶ If $a_{11} > 0$, we can essentially repeat the argument of previous theorem to deduce that A is positive.
- ▶ If $a_{11} = 0$, but $a_{jj} > 0$ for some $j > 1$, we can interchange first row and column with j -th row and column by pre and post multiplying by a permutation matrix and repeat the argument.

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.
- ▶ If $a_{11} > 0$, we can essentially repeat the argument of previous theorem to deduce that A is positive.
- ▶ If $a_{11} = 0$, but $a_{jj} > 0$ for some $j > 1$, we can interchange first row and column with j -th row and column by pre and post multiplying by a permutation matrix and repeat the argument.
- ▶ If $a_{jj} = 0$ for every j , then by considering positivity of 2×2 principal minors we see that $a_{ij} = 0$ for all i, j . In other words, in such a case A is the zero matrix and hence it is positive. ■

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.
- ▶ If $a_{11} > 0$, we can essentially repeat the argument of previous theorem to deduce that A is positive.
- ▶ If $a_{11} = 0$, but $a_{jj} > 0$ for some $j > 1$, we can interchange first row and column with j -th row and column by pre and post multiplying by a permutation matrix and repeat the argument.
- ▶ If $a_{jj} = 0$ for every j , then by considering positivity of 2×2 principal minors we see that $a_{ij} = 0$ for all i, j . In other words, in such a case A is the zero matrix and hence it is positive. ■
- ▶ These criteria are very useful to check positivity as in general it is difficult to compute eigenvalues.

Continuation

- ▶ Now consider a self-adjoint $A = [a_{ij}]_{1 \leq i, j \leq n}$ with all its principal minors positive.
- ▶ If $a_{11} > 0$, we can essentially repeat the argument of previous theorem to deduce that A is positive.
- ▶ If $a_{11} = 0$, but $a_{jj} > 0$ for some $j > 1$, we can interchange first row and column with j -th row and column by pre and post multiplying by a permutation matrix and repeat the argument.
- ▶ If $a_{jj} = 0$ for every j , then by considering positivity of 2×2 principal minors we see that $a_{ij} = 0$ for all i, j . In other words, in such a case A is the zero matrix and hence it is positive. ■
- ▶ These criteria are very useful to check positivity as in general it is difficult to compute eigenvalues.
- ▶ **END OF LECTURE 31.**

Lecture 32: Polynomials of a matrix

- ▶ Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

Lecture 32: Polynomials of a matrix

- ▶ Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

Lecture 32: Polynomials of a matrix

- ▶ Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.

Lecture 32: Polynomials of a matrix

- ▶ Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .

Lecture 32: Polynomials of a matrix

- ▶ Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .
- ▶ In particular, $I, A, A^2, \dots, A^{n^2}$ are linearly dependent.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.
- ▶ We may scale such a polynomial to make the leading coefficient one, i. e. we may take it to be monic.

Annihilating polynomials and division algorithm

- **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

- ▶ **Proof:** This is clear from the division algorithm for polynomials. As $f(A) = 0 = g(A).s(A)$, we get $r(A) = 0$.

Minimal polynomial

- **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.
- ▶ We may scale it suitably to make it monic. This contradicts minimality of q_1, q_2 . ■

Factorization

- **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.
- ▶ **Proof:** This is clear from the minimality of q and the division algorithm on dividing f by q . ■

Example-I

- Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Example-I

- Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.
- ▶ In particular, the unique minimal polynomial of C is given by $q(x) = (x - 2)(x - 3) = x^2 - 5x + 6$.

Example -II

► Example 32.7: Consider

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

Example -II

- ▶ Example 32.7: Consider

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

- ▶ Now the unique minimal polynomial of D is given by $q(x) = (x - 2)^2(x - 3)$.

Eigenvalues and annihilating polynomials

- **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■
- ▶ Now we may guess the following result.

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- **Corollary 32.9:** For any square matrix, the minimal polynomial is a factor of the characteristic polynomial.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (1)$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (1)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (1)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (1)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.
- ▶ We will see a correct proof in the next lecture.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (1)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.
- ▶ We will see a correct proof in the next lecture.
- ▶ **END OF LECTURE 32.**

Lecture 33: Cayley Hamilton theorem

- Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

Lecture 33: Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

Lecture 33: Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.

Lecture 33: Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .

Lecture 33: Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .
- ▶ In particular, $I, A, A^2, \dots, A^{n^2}$ are linearly dependent.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then
$$A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)}).$$

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.
- ▶ We may scale such a polynomial to make the leading coefficient one, i. e. we may take it to be monic.

Annihilating polynomials and division algorithm

- **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

- ▶ **Proof:** This is clear from the division algorithm for polynomials. As $f(A) = 0 = g(A).s(A)$, we get $r(A) = 0$.

Minimal polynomial

- **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.
- ▶ We may scale it suitably to make it monic. This contradicts minimality of q_1, q_2 . ■

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.
- ▶ **Proof:** This is clear from the minimality of q and the division algorithm on dividing f by q . ■

Example-I

- Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Example-I

- Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.
- ▶ In particular, the unique minimal polynomial of C is given by $q(x) = (x - 2)(x - 3) = x^2 - 5x + 6$.

Example -II

► Example 32.7: Consider

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

Example -II

- ▶ Example 32.7: Consider

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

- ▶ Now the unique minimal polynomial of D is given by $q(x) = (x - 2)^2(x - 3)$.

Eigenvalues and annihilating polynomials

- **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■
- ▶ Now we may guess the following result.

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- **Corollary 32.9:** For any square matrix, the minimal polynomial is a factor of the characteristic polynomial.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (2)$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (2)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (2)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (2)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.
- ▶ **END OF REVIEW.**

Similarity instead of unitary equivalence

- Example 33.1: Consider

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

Similarity instead of unitary equivalence

- ▶ Example 33.1: Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

- ▶ This shows that some times it maybe more prudent not to insist on unitary equivalence. We may try to simplify A through similarity instead of unitary equivalence. This is done either when there is no underlying inner product or when we have a prescribed inner product but we choose to ignore it.

Upper triangular form

- **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

Upper triangular form

- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ **Proof:** We may consider the standard inner product on $\mathbb{C}^n$. Then by Schur's upper triangularization theorem, there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Take $S = U$. Since $U^* = S^{-1}$, the proof is complete.

Upper triangular form

- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ **Proof:** We may consider the standard inner product on $\mathbb{C}^n$. Then by Schur's upper triangularization theorem, there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Take $S = U$. Since $U^* = S^{-1}$, the proof is complete.

- ▶ Alternatively, we may imitate the proof of Schur's upper triangularization theorem. Choose an eigenvector v_1 corresponding to some eigenvalue a_1 of A , extend $\{v_1\}$ to a basis of $\mathbb{C}^n$.

- ▶ In the new basis, the linear map A will have the form:

$$A = \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix}$$

for some $1 \times (n-1)$ row vector y and $(n-1) \times (n-1)$ matrix B . Now use induction. ■.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.
► Then $M_k = \text{span}\{e_1, e_2, \dots, e_k\}$.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.
► Then $M_k = \text{span}\{e_1, e_2, \dots, e_k\}$.
► Since T is upper triangular $(T - d_k I)$ is also upper triangular with k -th diagonal entry equal to 0.

- ▶ In particular, the j -th column of $(T - d_k I)$ is in the span of $\{e_1, e_2, \dots, e_{k-1}\}$ for $1 \leq j \leq k$.

- ▶ In particular, the j -th column of $(T - d_k I)$ is in the span of $\{e_1, e_2, \dots, e_{k-1}\}$ for $1 \leq j \leq k$.
- ▶ In other words $(T - d_k I)e_j \in M_{k-1}$ for $1 \leq k \leq n$. ■

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

- ▶ Note that for any polynomial f ,

$$f(A) = Sf(T)S^{-1}.$$

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

- ▶ Note that for any polynomial f ,

$$f(A) = Sf(T)S^{-1}.$$

- ▶ Let p be the characteristic polynomial of A and let $d_1, \dots, d_n$ be the diagonal entries of T .

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.
- ▶ By the lemma

$$(T - d_n I)x \in M_{n-1}.$$

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.
- ▶ By the lemma

$$(T - d_n I)x \in M_{n-1}.$$

- ▶ As $(T - d_{n-1} I)M_{n-1} \subseteq M_{n-2}$ we get

$$(T - d_{n-1} I)(T - d_n I)x \in M_{n-2}.$$

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ In other words, $p(T)x = 0$ for every $x \in \mathbb{C}^n$.

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ In other words, $p(T)x = 0$ for every $x \in \mathbb{C}^n$.
- ▶ This proves the claim. ■

Example

- **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .
- ▶ **Proof:** This is now clear, as the Cayley Hamilton theorem tells us that A^n is a linear combination of $\{I, A, \dots, A^{n-1}\}$.

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .
- ▶ **Proof:** This is now clear, as the Cayley Hamilton theorem tells us that A^n is a linear combination of $\{I, A, \dots, A^{n-1}\}$.
- ▶ It is then easy to see that A^m for $m \geq n$ are also in the span of $\{I, A, \dots, A^{n-1}\}$.

Multivariable case

- ▶ **Remark 33.6:** Suppose A, B are commuting $n \times n$ matrices. What is the maximum possible dimension of $\{f(A, B) : f \text{ is a two variable polynomial}\}$?

Multivariable case

- ▶ **Remark 33.6:** Suppose A, B are commuting $n \times n$ matrices. What is the maximum possible dimension of $\{f(A, B) : f \text{ is a two variable polynomial}\}$?
- ▶ Surprisingly the answer is still n . This is known as Gerstenhaber's theorem.

Multivariable case

- ▶ **Remark 33.6:** Suppose A, B are commuting $n \times n$ matrices. What is the maximum possible dimension of $\{f(A, B) : f \text{ is a two variable polynomial}\}$?
- ▶ Surprisingly the answer is still n . This is known as Gerstenhaber's theorem.
- ▶ **Question:** Suppose A, B, C are commuting $n \times n$ matrices. Is the maximum possible dimension of

$$\{f(A, B, C) : f \text{ is a three variable polynomial}\}$$

Multivariable case

- ▶ **Remark 33.6:** Suppose A, B are commuting $n \times n$ matrices. What is the maximum possible dimension of $\{f(A, B) : f \text{ is a two variable polynomial}\}$?
- ▶ Surprisingly the answer is still n . This is known as Gerstenhaber's theorem.
- ▶ **Question:** Suppose A, B, C are commuting $n \times n$ matrices. Is the maximum possible dimension of

$$\{f(A, B, C) : f \text{ is a three variable polynomial}\}$$

- ▶ equal to n ?

Multivariable case

- ▶ **Remark 33.6:** Suppose A, B are commuting $n \times n$ matrices. What is the maximum possible dimension of $\{f(A, B) : f \text{ is a two variable polynomial}\}$?
- ▶ Surprisingly the answer is still n . This is known as Gerstenhaber's theorem.
- ▶ **Question:** Suppose A, B, C are commuting $n \times n$ matrices. Is the maximum possible dimension of

$$\{f(A, B, C) : f \text{ is a three variable polynomial}\}$$

- ▶ equal to n ?
- ▶ Ans: Not known.

References (Optional)

- ▶ (1) B. Sethuraman, The algebra generated by three commuting matrices, Ramanujan Math. Soc. Newsletter, 21 number 2, September 2011, 26-31.

References (Optional)

- ▶ (1) B. Sethuraman, The algebra generated by three commuting matrices, Ramanujan Math. Soc. Newsletter, 21 number 2, September 2011, 26-31.
- ▶ (2) Holbrook and C.O'Meara, Some thoughts on Gerstenhaber's theorem, Linear Algebra and its Applications Volume 466, 1 February 2015, Pages 267-295.

References (Optional)

- ▶ (1) B. Sethuraman, The algebra generated by three commuting matrices, Ramanujan Math. Soc. Newsletter, 21 number 2, September 2011, 26-31.
- ▶ (2) Holbrook and C.O'Meara, Some thoughts on Gerstenhaber's theorem, Linear Algebra and its Applications Volume 466, 1 February 2015, Pages 267-295.
- ▶ It is known that the dimension of the algebra generated can be more than n , when there are four or more commuting matrices.

References (Optional)

- ▶ (1) B. Sethuraman, The algebra generated by three commuting matrices, Ramanujan Math. Soc. Newsletter, 21 number 2, September 2011, 26-31.
- ▶ (2) Holbrook and C.O'Meara, Some thoughts on Gerstenhaber's theorem, Linear Algebra and its Applications Volume 466, 1 February 2015, Pages 267-295.
- ▶ It is known that the dimension of the algebra generated can be more than n , when there are four or more commuting matrices.
- ▶ **END OF LECTURE 33.**

Lecture 34: Some applications of Cayley Hamilton theorem

- Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

Lecture 34: Some applications of Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

Lecture 34: Some applications of Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.

Lecture 34: Some applications of Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .

Lecture 34: Some applications of Cayley Hamilton theorem

- ▶ Recall: Let A be an $n \times n$ complex matrix. For any complex polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, by definition,

$$f(A) = a_0I + a_1A + \cdots + a_mA^m.$$

- ▶ Consider

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}.$$

- ▶ Clearly this is a subspace of $M_n(\mathbb{C})$. Actually, $\mathcal{A}$ is a 'sub-algebra' of $M_n(\mathbb{C})$, that is, it is also closed under taking products.
- ▶ Note that $M_n(\mathbb{C})$ is a vector space of dimension n^2 . Therefore the dimension of $\mathcal{A}$ can't be more than n^2 .
- ▶ In particular, $I, A, A^2, \dots, A^{n^2}$ are linearly dependent.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.

Continuation

- ▶ In other words, there exists a non-zero polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ of degree at most n^2 such that

$$q(A) = b_0I + b_1A + b_2A^2 + \cdots + b_mA^m = 0.$$

- ▶ Assume $b_m \neq 0$. Then $A^m = -\frac{1}{b_m}(b_0I + b_1A + \cdots + b_{m-1}A^{(m-1)})$.
- ▶ This may help us to compute higher powers of A or to simplify higher degree polynomials in A .
- ▶ So we would look for a non-zero polynomial q of lowest degree satisfying $q(A) = 0$.
- ▶ We may scale such a polynomial to make the leading coefficient one, i. e. we may take it to be monic.

Annihilating polynomials and division algorithm

- **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

Annihilating polynomials and division algorithm

- ▶ **Definition 32.1:** A polynomial f is said to be **annihilating** for a matrix A if $f(A) = 0$.
- ▶ **Theorem 32.2:** Let f, g be non-zero annihilating polynomials of a matrix A and suppose $\text{degree}(g) \leq \text{degree}(f)$. Then

$$f(x) = g(x)s(x) + r(x)$$

for some polynomials s, r , where either $r = 0$ or $\text{degree}(r) < \text{degree}(g)$ and $r(A) = 0$.

- ▶ **Proof:** This is clear from the division algorithm for polynomials. As $f(A) = 0 = g(A).s(A)$, we get $r(A) = 0$.

Minimal polynomial

- **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.

Minimal polynomial

- ▶ **Theorem 32.3:** Let A be an $n \times n$ complex matrix. Then there exists a unique monic polynomial q of lowest degree such that $q(A) = 0$.
- ▶ **Proof:** Suppose q_1, q_2 are two distinct non-zero monic polynomials of lowest degree such that $q_1(A) = q_2(A) = 0$.
- ▶ Then clearly $q_1 - q_2$ is a lower degree polynomial with $(q_1 - q_2)(A) = 0$.
- ▶ We may scale it suitably to make it monic. This contradicts minimality of q_1, q_2 . ■

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.

Factorization

- ▶ **Definition 32.4:** Given a matrix A , the unique monic polynomial of lowest degree q , satisfying $q(A) = 0$ is defined as the **minimal polynomial** of A .
- ▶ **Theorem 32.5:** Let A be an $n \times n$ complex matrix. Let q be the minimal polynomial of A . Suppose f is an annihilating polynomial of A , then there exists a polynomial s such that $f(x) = q(x)s(x)$. In other words, the minimal polynomial is a factor of every annihilating polynomial.
- ▶ **Proof:** This is clear from the minimality of q and the division algorithm on dividing f by q . ■

Example-I

- Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.

Example-I

- ▶ Example 32.6: Consider

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ Then for any polynomial f ,

$$f(C) = \begin{bmatrix} f(2) & 0 & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(3) \end{bmatrix}.$$

- ▶ Therefore, f is an annihilating polynomial for C if and only if $f(2) = f(3) = 0$.
- ▶ In particular, the unique minimal polynomial of C is given by $q(x) = (x - 2)(x - 3) = x^2 - 5x + 6$.

Example -II

► Example 32.7: Consider

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

Example -II

- ▶ Example 32.7: Consider

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

- ▶ Now the unique minimal polynomial of D is given by $q(x) = (x - 2)^2(x - 3)$.

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■

Eigenvalues and annihilating polynomials

- ▶ **Theorem 32.8:** Suppose A is a complex matrix and a is an eigenvalue of A . If f is an annihilating polynomial of A then $f(a) = 0$. In particular, every eigenvalue is a root of the minimal polynomial.
- ▶ **Proof:** Suppose v is an eigenvector of A with eigenvalue a .
- ▶ Clearly, $A^k v = a^k v$ for every k .
- ▶ Hence for any polynomial f ,

$$f(A)v = f(a)v.$$

- ▶ Since $v \neq 0$, if $f(A)v = 0$ then $f(a) = 0$. Now the result is immediate. ■
- ▶ Now we may guess the following result.

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

Cayley Hamilton theorem

- **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- **Corollary 32.9:** For any square matrix, the minimal polynomial is a factor of the characteristic polynomial.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (3)$$

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(AI - A) = \det(A - A) = \det(0) = 0. \quad (3)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.

A wrong proof

- ▶ **Wrong proof:** By the definition of the characteristic polynomial:

$$p(x) = \det(xI - A).$$

- ▶ Taking $x = A$,

$$p(A) = \det(A.I - A) = \det(A - A) = \det(0) = 0. \quad (3)$$

- ▶ This is a wrong proof, as in the equation above, on the left we have a matrix, where as, on the right we have a scalar.
- ▶ We can't blindly substitute $x = A$ and do determinant computations.

Similarity instead of unitary equivalence

- Example 33.1: Consider

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

Similarity instead of unitary equivalence

- ▶ Example 33.1: Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

Similarity instead of unitary equivalence

- ▶ **Example 33.1:** Consider

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

- ▶ Then clearly A is not normal. Hence A can not be diagonalized using unitary equivalence.
- ▶ However, $\sigma(A) = \{1, 3\}$ and since the corresponding geometric multiplicities are at least 1, we can get a basis of eigenvectors of A . In other words, there exists an invertible matrix S such that

$$A = S \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} S^{-1}.$$

- ▶ This shows that some times it maybe more prudent not to insist on unitary equivalence. We may try to simplify A through similarity instead of unitary equivalence. This is done either when there is no underlying inner product or when we have a prescribed inner product but we choose to ignore it.

Upper triangular form

- **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

Upper triangular form

- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ **Proof:** We may consider the standard inner product on $\mathbb{C}^n$. Then by Schur's upper triangularization theorem, there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Take $S = U$. Since $U^* = S^{-1}$, the proof is complete.

Upper triangular form

- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ **Proof:** We may consider the standard inner product on $\mathbb{C}^n$. Then by Schur's upper triangularization theorem, there exists a unitary U and an upper triangular matrix T such that

$$A = UTU^*.$$

Take $S = U$. Since $U^* = S^{-1}$, the proof is complete.

- ▶ Alternatively, we may imitate the proof of Schur's upper triangularization theorem. Choose an eigenvector v_1 corresponding to some eigenvalue a_1 of A , extend $\{v_1\}$ to a basis of $\mathbb{C}^n$.

- ▶ In the new basis, the linear map A will have the form:

$$A = \begin{bmatrix} a_1 & y \\ 0 & B \end{bmatrix}$$

for some $1 \times (n-1)$ row vector y and $(n-1) \times (n-1)$ matrix B . Now use induction. ■.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.
► Then $M_k = \text{span}\{e_1, e_2, \dots, e_k\}$.

A Lemma

- **Lemma 33.3:** Let T be an upper triangular matrix with diagonal entries $d_1, d_2, \dots, d_n$. For $1 \leq k \leq n$, take

$$M_k = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, x_2, \dots, x_k \in \mathbb{C} \right\}.$$

Take $M_0 = \{0\}$. Then for every $1 \leq k \leq n$,

$$(T - d_k I)(M_k) \subseteq M_{k-1}.$$

- **Proof:** Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{C}^n$.
► Then $M_k = \text{span}\{e_1, e_2, \dots, e_k\}$.
► Since T is upper triangular $(T - d_k I)$ is also upper triangular with k -th diagonal entry equal to 0.

- ▶ In particular, the j -th column of $(T - d_k I)$ is in the span of $\{e_1, e_2, \dots, e_{k-1}\}$ for $1 \leq j \leq k$.

- ▶ In particular, the j -th column of $(T - d_k I)$ is in the span of $\{e_1, e_2, \dots, e_{k-1}\}$ for $1 \leq j \leq k$.
- ▶ In other words $(T - d_k I)e_j \in M_{k-1}$ for $1 \leq k \leq n$. ■

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

- ▶ Note that for any polynomial f ,

$$f(A) = Sf(T)S^{-1}.$$

Cayley-Hamilton Theorem

- ▶ Now we present a proof of this famous theorem.
- ▶ **Theorem 32.9 (Cayley Hamilton theorem):** Let A be a complex $n \times n$ matrix and let p be the characteristic polynomial of A . Then

$$p(A) = 0.$$

In other words, the characteristic polynomial is an annihilating polynomial for A .

- ▶ **Proof:** By Theorem 33.2, there exists a non-singular matrix S and an upper triangular matrix T such that

$$A = STS^{-1}.$$

- ▶ Note that for any polynomial f ,

$$f(A) = Sf(T)S^{-1}.$$

- ▶ Let p be the characteristic polynomial of A and let $d_1, \dots, d_n$ be the diagonal entries of T .

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.
- ▶ By the lemma

$$(T - d_n I)x \in M_{n-1}.$$

Continuation

- ▶ Then p is also the characteristic polynomial of T and

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ As $p(A) = Sp(T)S^{-1}$, it suffices to show that $p(T) = 0$.
- ▶ We use the notation of previous lemma. Consider any $x \in \mathbb{C}^n = M_n$.
- ▶ By the lemma

$$(T - d_n I)x \in M_{n-1}.$$

- ▶ As $(T - d_{n-1} I)M_{n-1} \subseteq M_{n-2}$ we get

$$(T - d_{n-1} I)(T - d_n I)x \in M_{n-2}.$$

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ In other words, $p(T)x = 0$ for every $x \in \mathbb{C}^n$.

- ▶ Continuing this way (i.e., by induction) :

$$(T - d_1 I)(T - d_2 I) \cdots (T - d_n I)x \in M_0 = \{0\}.$$

- ▶ In other words, $p(T)x = 0$ for every $x \in \mathbb{C}^n$.
- ▶ This proves the claim. ■

Example

- **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .
- ▶ **Proof:** This is now clear, as the Cayley Hamilton theorem tells us that A^n is a linear combination of $\{I, A, \dots, A^{n-1}\}$.

Example

- ▶ **Example 33.4** Suppose D is a diagonal matrix with diagonal entries $d_1, d_2, \dots, d_n$. Then the characteristic polynomial of D is given by

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ It is clear that $p(D) = 0$.
- ▶ **Corollary 33.5:** Suppose A is an $n \times n$ matrix. Then the dimension of

$$\mathcal{A} = \{f(A) : f \text{ is a polynomial}\}$$

- ▶ is at most n .
- ▶ **Proof:** This is now clear, as the Cayley Hamilton theorem tells us that A^n is a linear combination of $\{I, A, \dots, A^{n-1}\}$.
- ▶ It is then easy to see that A^m for $m \geq n$ are also in the span of $\{I, A, \dots, A^{n-1}\}$.
- ▶ **END OF REVIEW**

Computation of polynomials

- ▶ We have already indicated how the Cayley Hamilton theorem can help us to compute higher powers or general polynomials of matrices.

Computation of polynomials

- ▶ We have already indicated how the Cayley Hamilton theorem can help us to compute higher powers or general polynomials of matrices.
- ▶ Here we present some other simple applications.

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.
- ▶ **Theorem 34.2:** A matrix A is nilpotent if and only if $\sigma(A) = \{0\}$.

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.
- ▶ **Theorem 34.2:** A matrix A is nilpotent if and only if $\sigma(A) = \{0\}$.
- ▶ **Proof:** Suppose $\sigma(A) = \{0\}$. Then the characteristic polynomial of A is $p(x) = x^n$.

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.
- ▶ **Theorem 34.2:** A matrix A is nilpotent if and only if $\sigma(A) = \{0\}$.
- ▶ **Proof:** Suppose $\sigma(A) = \{0\}$. Then the characteristic polynomial of A is $p(x) = x^n$.
- ▶ Hence by Cayley Hamilton theorem, $A^n = 0$.

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.
- ▶ **Theorem 34.2:** A matrix A is nilpotent if and only if $\sigma(A) = \{0\}$.
- ▶ **Proof:** Suppose $\sigma(A) = \{0\}$. Then the characteristic polynomial of A is $p(x) = x^n$.
- ▶ Hence by Cayley Hamilton theorem, $A^n = 0$.
- ▶ Conversely, suppose $A^k = 0$. Now if a is an eigenvalue with eigenvector v , we get $A^k v = a^k v = 0$

Nilpotent matrices

- ▶ **Definition 34.1:** A matrix A is said to be **nilpotent** if $A^k = 0$ for some $k \geq 1$.
- ▶ **Theorem 34.2:** A matrix A is nilpotent if and only if $\sigma(A) = \{0\}$.
- ▶ **Proof:** Suppose $\sigma(A) = \{0\}$. Then the characteristic polynomial of A is $p(x) = x^n$.
- ▶ Hence by Cayley Hamilton theorem, $A^n = 0$.
- ▶ Conversely, suppose $A^k = 0$. Now if a is an eigenvalue with eigenvector v , we get $A^k v = a^k v = 0$
- ▶ As $v \neq 0$, this implies $a^k = 0$. Hence $a = 0$. Therefore $\sigma(A) = \{0\}$. ■

Comparison of coefficients

- ▶ Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$ be the characteristic polynomial of a matrix A .

Comparison of coefficients

- ▶ Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$ be the characteristic polynomial of a matrix A .
- ▶ If $d_1, \dots, d_n$ are the eigenvalues of A , then

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

Comparison of coefficients

- ▶ Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$ be the characteristic polynomial of a matrix A .
- ▶ If $d_1, \dots, d_n$ are the eigenvalues of A , then

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ Comparing the coefficients, we see that $c_{n-1} = (-1)(d_1 + d_2 + \cdots + d_n) = -(\text{trace}(A))$ and $c_0 = (-1)^n d_1 d_2 \cdots d_n = (-1)^n (\det(A))$.

Comparison of coefficients

- ▶ Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0$ be the characteristic polynomial of a matrix A .
- ▶ If $d_1, \dots, d_n$ are the eigenvalues of A , then

$$p(x) = (x - d_1)(x - d_2) \cdots (x - d_n).$$

- ▶ Comparing the coefficients, we see that $c_{n-1} = (-1)(d_1 + d_2 + \cdots + d_n) = -(\text{trace}(A))$ and $c_0 = (-1)^n d_1 d_2 \cdots d_n = (-1)^n (\det(A))$.
- ▶ In particular, A is invertible if and only if $c_0 \neq 0$.

Computation of inverse

- Now by Cayley Hamilton theorem,

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I = 0.$$

Computation of inverse

- ▶ Now by Cayley Hamilton theorem,

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I = 0.$$

- ▶ This implies,

$$A(A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I) = -c_0I.$$

Computation of inverse

- ▶ Now by Cayley Hamilton theorem,

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I = 0.$$

- ▶ This implies,

$$A(A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I) = -c_0I.$$

- ▶ Assuming that A is invertible,

$$-\frac{1}{c_0}(A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I)$$

is the inverse of A .

Special Case

- ▶ Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Special Case

- ▶ Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

- ▶ We have the characteristic polynomial:

$$p(x) = (x - a)(x - d) - bc = x^2 - (a + d)x + (ad - bc).$$

Special Case

- ▶ Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

- ▶ We have the characteristic polynomial:

$$p(x) = (x - a)(x - d) - bc = x^2 - (a + d)x + (ad - bc).$$

- ▶ Note that we are getting the trace and the determinant as the coefficients.

Special Case

- ▶ Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

- ▶ We have the characteristic polynomial:

$$p(x) = (x - a)(x - d) - bc = x^2 - (a + d)x + (ad - bc).$$

- ▶ Note that we are getting the trace and the determinant as the coefficients.
- ▶ We have from the Cayley Hamilton theorem

$$A^2 - (a + d)A + (ad - bc) = 0.$$

Special Case

- ▶ Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

- ▶ We have the characteristic polynomial:

$$p(x) = (x - a)(x - d) - bc = x^2 - (a + d)x + (ad - bc).$$

- ▶ Note that we are getting the trace and the determinant as the coefficients.
- ▶ We have from the Cayley Hamilton theorem

$$A^2 - (a + d)A + (ad - bc) = 0.$$

- ▶ In particular, if the determinant of A is non-zero, then

$$A^{-1} = \frac{-1}{ad - bc}(A - (a + d)I) = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.
- ▶ By the Cayley Hamilton theorem,

$$(A - d_1 I)(A - d_2 I) \cdots (A - d_n I) = 0.$$

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.
- ▶ By the Cayley Hamilton theorem,

$$(A - d_1 I)(A - d_2 I) \cdots (A - d_n I) = 0.$$

- ▶ So if

$$(A - d_2 I) \cdots (A - d_n I) \neq 0$$

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.
- ▶ By the Cayley Hamilton theorem,

$$(A - d_1 I)(A - d_2 I) \cdots (A - d_n I) = 0.$$

- ▶ So if

$$(A - d_2 I) \cdots (A - d_n I) \neq 0$$

- ▶ Then any non-zero column of $(A - d_2 I) \cdots (A - d_n I)$ is an eigenvector of A with eigenvalue d_1 .

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.
- ▶ By the Cayley Hamilton theorem,

$$(A - d_1 I)(A - d_2 I) \cdots (A - d_n I) = 0.$$

- ▶ So if

$$(A - d_2 I) \cdots (A - d_n I) \neq 0$$

- ▶ Then any non-zero column of $(A - d_2 I) \cdots (A - d_n I)$ is an eigenvector of A with eigenvalue d_1 .
- ▶ If the product above is zero, this means that $d_j = d_1$ for some $j > 1$. (Recall that every eigenvalue must be a root of any annihilating polynomial).

Computing eigenvectors

- ▶ Suppose A is a matrix with eigenvalues $d_1, d_2, \dots, d_n$.
- ▶ By the Cayley Hamilton theorem,

$$(A - d_1 I)(A - d_2 I) \cdots (A - d_n I) = 0.$$

- ▶ So if

$$(A - d_2 I) \cdots (A - d_n I) \neq 0$$

- ▶ Then any non-zero column of $(A - d_2 I) \cdots (A - d_n I)$ is an eigenvector of A with eigenvalue d_1 .
- ▶ If the product above is zero, this means that $d_j = d_1$ for some $j > 1$. (Recall that every eigenvalue must be a root of any annihilating polynomial).
- ▶ In such cases, you may drop some of the factors $(A - d_j I)$ with $d_j = d_1$ to get eigenvectors.

Example

- Example 34.3: Consider

$$T = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix}.$$

Example

- ▶ Example 34.3: Consider

$$T = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ It is clear that $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 1.

Example

- ▶ Example 34.3: Consider

$$T = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ It is clear that $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 1.
- ▶ What about eigenvalues 2, 3?

Example

- ▶ Example 34.3: Consider

$$T = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ It is clear that $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 1.
- ▶ What about eigenvalues 2, 3?
- ▶ By direct computation

$$(T - I)(T - 2I) = \begin{bmatrix} 0 & 0 & 29 \\ 0 & 0 & 12 \\ 0 & 0 & 2 \end{bmatrix}.$$

Example

- ▶ Example 34.3: Consider

$$T = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 3 \end{bmatrix}.$$

- ▶ It is clear that $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 1.
- ▶ What about eigenvalues 2, 3?
- ▶ By direct computation

$$(T - I)(T - 2I) = \begin{bmatrix} 0 & 0 & 29 \\ 0 & 0 & 12 \\ 0 & 0 & 2 \end{bmatrix}.$$

- ▶ This shows that $\begin{pmatrix} 29 \\ 12 \\ 2 \end{pmatrix}$ is an eigenvector with eigenvalue 3.

► Similarly

$$(T - I)(T - 3I) = \begin{bmatrix} 0 & -4 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

► Similarly

$$(T - I)(T - 3I) = \begin{bmatrix} 0 & -4 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

► Hence we see that $\begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 2.

- ▶ Similarly

$$(T - I)(T - 3I) = \begin{bmatrix} 0 & -4 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- ▶ Hence we see that $\begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 2.
- ▶ END OF LECTURE 34

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.
- ▶ However, we have the following theorem.

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.
- ▶ However, we have the following theorem.
- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.
- ▶ However, we have the following theorem.
- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ Here the diagonal entries of T are the eigenvalues of A and they can appear in any order we like.

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.
- ▶ However, we have the following theorem.
- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ Here the diagonal entries of T are the eigenvalues of A and they can appear in any order we like.
- ▶ What about off-diagonal entries. Can we bring some order in them?

Lecture 35: Jordan canonical form

- ▶ Recall that we can't diagonalize some matrices.
- ▶ However, we have the following theorem.
- ▶ **Theorem 33.2:** Let A be an $n \times n$ complex matrix. Then there exists an upper triangular matrix T and an invertible matrix S such that

$$A = STS^{-1}.$$

- ▶ Here the diagonal entries of T are the eigenvalues of A and they can appear in any order we like.
- ▶ What about off-diagonal entries. Can we bring some order in them?
- ▶ This is answered by Jordan canonical form theorem.

Jordan blocks

- ▶ We need some notation.

Jordan blocks

- ▶ We need some notation.
- ▶ **Notation:** For $b \in \mathbb{C}$ and $n \in \mathbb{N}$, let $J_b(n)$ denote the $n \times n$ matrix whose diagonal entries are equal to b and the super diagonal entries are equal to 1 and all the other entries are equal to zero:

$$J_b(1) = [b], \quad J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

Jordan blocks

- ▶ We need some notation.
- ▶ **Notation:** For $b \in \mathbb{C}$ and $n \in \mathbb{N}$, let $J_b(n)$ denote the $n \times n$ matrix whose diagonal entries are equal to b and the super diagonal entries are equal to 1 and all the other entries are equal to zero:

$$J_b(1) = [b], \quad J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

- ▶ Alternatively,

$$(J_b(n))_{ij} = \begin{cases} b & \text{if } j = i; \\ 1 & \text{if } j = i + 1; \\ 0 & \text{otherwise.} \end{cases}$$

Eigenvalues and eigenvectors

- ▶ We observe that for any Jordan block matrix $J_b(n)$, the only eigenvalue is b and the only eigenvectors are vectors of the form

$$\begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Eigenvalues and eigenvectors

- ▶ We observe that for any Jordan block matrix $J_b(n)$, the only eigenvalue is b and the only eigenvectors are vectors of the form

$$\begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- ▶ Therefore the eigenvalue b has algebraic multiplicity n and geometric multiplicity 1 in $J_b(n)$.

Eigenvalues and eigenvectors

- ▶ We observe that for any Jordan block matrix $J_b(n)$, the only eigenvalue is b and the only eigenvectors are vectors of the form

$$\begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- ▶ Therefore the eigenvalue b has algebraic multiplicity n and geometric multiplicity 1 in $J_b(n)$.
- ▶ The characteristic polynomial of $J_b(n)$ is $(x - b)^n$.

Jordan block theorem

- ▶ **Theorem 35.1:** Every square matrix is similar to a direct sum of Jordan blocks, which are unique up to permutation.

Jordan block theorem

- ▶ **Theorem 35.1:** Every square matrix is similar to a direct sum of Jordan blocks, which are unique up to permutation.
- ▶ In other words, if A is an $n \times n$ matrix and $a_1, a_2, \dots, a_k$ are the distinct eigenvalues of A , with geometric multiplicities $g_1, g_2, \dots, g_k$, then there exist natural numbers n_{ij} , $1 \leq i \leq k, 1 \leq j \leq g_i$ such that A is similar to

$$\bigoplus_{i=1}^k \bigoplus_{j=1}^{g_i} J_{a_i}(n_{ij}).$$

This decomposition is unique up to permutation.

Jordan block theorem

- ▶ **Theorem 35.1:** Every square matrix is similar to a direct sum of Jordan blocks, which are unique up to permutation.
- ▶ In other words, if A is an $n \times n$ matrix and $a_1, a_2, \dots, a_k$ are the distinct eigenvalues of A , with geometric multiplicities $g_1, g_2, \dots, g_k$, then there exist natural numbers n_{ij} , $1 \leq i \leq k, 1 \leq j \leq g_i$ such that A is similar to

$$\bigoplus_{i=1}^k \bigoplus_{j=1}^{g_i} J_{a_i}(n_{ij}).$$

This decomposition is unique up to permutation.

- ▶ Note that we must have:

$$\sum_{i=1}^k \sum_{j=1}^{g_i} n_{ij} = n.$$

Jordan block theorem

- ▶ **Theorem 35.1:** Every square matrix is similar to a direct sum of Jordan blocks, which are unique up to permutation.
- ▶ In other words, if A is an $n \times n$ matrix and $a_1, a_2, \dots, a_k$ are the distinct eigenvalues of A , with geometric multiplicities $g_1, g_2, \dots, g_k$, then there exist natural numbers n_{ij} , $1 \leq i \leq k, 1 \leq j \leq g_i$ such that A is similar to

$$\bigoplus_{i=1}^k \bigoplus_{j=1}^{g_i} J_{a_i}(n_{ij}).$$

This decomposition is unique up to permutation.

- ▶ Note that we must have:

$$\sum_{i=1}^k \sum_{j=1}^{g_i} n_{ij} = n.$$

- ▶ In other words, the sum of block sizes is equal to n and for each distinct eigenvalue there are as many Jordan blocks (possibly of different sizes) as the geometric multiplicity of the eigenvalue.

Examples

- **Example 35.1:** Suppose A is similar to

$$J_0(1) \oplus J_0(2) \oplus J_0(4) \oplus J_5(3).$$

Then the eigenvalues are 0 and 5. The order of A is $1 + 2 + 4 + 3 = 10$. The algebraic multiplicity of 0 is $(1 + 2 + 4) = 7$ and the geometric multiplicity of 0 is 3. The algebraic multiplicity of 5 is 3 and its geometric multiplicity is 1.

Examples

- **Example 35.1:** Suppose A is similar to

$$J_0(1) \oplus J_0(2) \oplus J_0(4) \oplus J_5(3).$$

Then the eigenvalues are 0 and 5. The order of A is $1 + 2 + 4 + 3 = 10$. The algebraic multiplicity of 0 is $(1 + 2 + 4) = 7$ and the geometric multiplicity of 0 is 3. The algebraic multiplicity of 5 is 3 and its geometric multiplicity is 1.

- Note that if A is diagonalizable (in particular, if it is normal), all Jordan blocks would have size 1 and A is similar to

$$\bigoplus_{i=1}^k \bigoplus_{j=1}^{n_i} J_{a_i}(1).$$

In other words geometric and algebraic multiplicities of a_i are equal to n_i and $n_1 + n_2 + \cdots + n_k = n$.

Minimal polynomial and characteristic polynomial

- Consider a Jordan block

$$J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

Minimal polynomial and characteristic polynomial

- ▶ Consider a Jordan block

$$J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

- ▶ Then the characteristic polynomial of $J_b(n)$ is $p(x) = (x - b)^n$ and is also the minimal polynomial of $J_b(n)$.

Minimal polynomial and characteristic polynomial

- ▶ Consider a Jordan block

$$J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

- ▶ Then the characteristic polynomial of $J_b(n)$ is $p(x) = (x - b)^n$ and is also the minimal polynomial of $J_b(n)$.
- ▶ If we consider $B = J_b(n_1) \oplus J_b(n_2) \oplus \dots \oplus J_b(n_r)$ then the characteristic polynomial of B is

$$p(x) = (x - b)^{n_1 + n_2 + \dots + n_r}$$

and the minimal polynomial is $q(x) = (x - b)^m$ where $m = \max\{n_1, n_2, \dots, n_r\}$.

Minimal polynomial and characteristic polynomial

- ▶ Consider a Jordan block

$$J_b(n) = \begin{bmatrix} b & 1 & 0 & \dots & 0 \\ 0 & b & 1 & \dots & 0 \\ 0 & 0 & b & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b \end{bmatrix}.$$

- ▶ Then the characteristic polynomial of $J_b(n)$ is $p(x) = (x - b)^n$ and is also the minimal polynomial of $J_b(n)$.
- ▶ If we consider $B = J_b(n_1) \oplus J_b(n_2) \oplus \dots \oplus J_b(n_r)$ then the characteristic polynomial of B is

$$p(x) = (x - b)^{n_1 + n_2 + \dots + n_r}$$

and the minimal polynomial is $q(x) = (x - b)^m$ where $m = \max\{n_1, n_2, \dots, n_r\}$.

- ▶ Thus we get the following result.

Minimal polynomial and characteristic polynomial

- **Theorem 35.1** Let A be a matrix similar to Jordan block

$$\bigoplus_{i=1}^k \bigoplus_{j=1}^{g_i} J_{a_i}(n_{ij}).$$

as in Theorem 35.1. Then the characteristic polynomial of A is given by

$$p(x) = \prod_{i=1}^k (x - a_i)^{\sum_{j=1}^{g_i} n_{ij}}$$

The minimal polynomial is given by

$$q(x) = \prod_{i=1}^k (x - a_i)^{m_i},$$

where $m_i = \max\{n_{ij} : 1 \leq j \leq g_i\}$.

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.
- ▶ Take $N = [n_{ij}]$ be the $n \times n$ matrix where

$$n_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases}.$$

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.
- ▶ Take $N = [n_{ij}]$ be the $n \times n$ matrix where

$$n_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Observe that N^2 is given by

$$(N^2)_{ij} = \begin{cases} 1 & \text{if } j = i + 2 \\ 0 & \text{otherwise.} \end{cases}.$$

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.
- ▶ Take $N = [n_{ij}]$ be the $n \times n$ matrix where

$$n_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Observe that N^2 is given by

$$(N^2)_{ij} = \begin{cases} 1 & \text{if } j = i + 2 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Similarly,

$$(N^3)_{ij} = \begin{cases} 1 & \text{if } j = i + 3 \\ 0 & \text{otherwise.} \end{cases}.$$

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.
- ▶ Take $N = [n_{ij}]$ be the $n \times n$ matrix where

$$n_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Observe that N^2 is given by

$$(N^2)_{ij} = \begin{cases} 1 & \text{if } j = i + 2 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Similarly,

$$(N^3)_{ij} = \begin{cases} 1 & \text{if } j = i + 3 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Finally,

$$(N^{n-1})_{ij} = \begin{cases} 1 & \text{if } i = 1, j = n \\ 0 & \text{otherwise.} \end{cases}.$$

Powers of Jordan blocks

- ▶ Consider $B = J_b(n)$ with $n \geq q$.
- ▶ Take $N = [n_{ij}]$ be the $n \times n$ matrix where

$$n_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Observe that N^2 is given by

$$(N^2)_{ij} = \begin{cases} 1 & \text{if } j = i + 2 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Similarly,

$$(N^3)_{ij} = \begin{cases} 1 & \text{if } j = i + 3 \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ Finally,

$$(N^{n-1})_{ij} = \begin{cases} 1 & \text{if } i = 1, j = n \\ 0 & \text{otherwise.} \end{cases}.$$

- ▶ and $N^n = 0$.

Continuation

- ▶ Now $J_b(n) = bl + N$.

Continuation

- ▶ Now $J_b(n) = bI + N$.
- ▶ Since bI and N commute, the binomial expansion is applicable and for $m \geq 1$ we get

$$(bI + N)^m = \sum_{k=0}^m \binom{m}{k} b^{m-k} N^k.$$

Continuation

- ▶ Now $J_b(n) = bI + N$.
- ▶ Since bI and N commute, the binomial expansion is applicable and for $m \geq 1$ we get

$$(bI + N)^m = \sum_{k=0}^m \binom{m}{k} b^{m-k} N^k.$$

- ▶ In other words,

$$(J_b(n)^m)_{ij} = \begin{cases} \binom{m}{j-i} b^{m-(j-i)} & i \leq j \leq n; \\ 0 & \text{otherwise.} \end{cases}$$

Continuation

- More explicitly: $(J_b(n))^m$ equals

$$\begin{bmatrix} b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \binom{m}{3} b^{m-3} & \dots \\ 0 & b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \dots \\ 0 & 0 & b^m & mb^{m-1} & \dots \\ 0 & 0 & 0 & b^m & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Continuation

- More explicitly: $(J_b(n))^m$ equals

$$\begin{bmatrix} b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \binom{m}{3} b^{m-3} & \dots \\ 0 & b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \dots \\ 0 & 0 & b^m & mb^{m-1} & \dots \\ 0 & 0 & 0 & b^m & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

- This way we can write down powers of Jordan blocks explicitly.

Continuation

- ▶ More explicitly: $(J_b(n))^m$ equals

$$\begin{bmatrix} b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \binom{m}{3} b^{m-3} & \dots \\ 0 & b^m & mb^{m-1} & \binom{m}{2} b^{m-2} & \dots \\ 0 & 0 & b^m & mb^{m-1} & \dots \\ 0 & 0 & 0 & b^m & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

- ▶ This way we can write down powers of Jordan blocks explicitly.
- ▶ The proof of Jordan block theorem has been omitted.

Generalized eigenspaces

- ▶ Let A be a complex $n \times n$ matrix and let a be an eigenvalue of A .

Generalized eigenspaces

- ▶ Let A be a complex $n \times n$ matrix and let a be an eigenvalue of A .
- ▶ Then

$$E(a) = \{x \in \mathbb{C}^n : (A - aI)x = 0\}$$

is known as the eigenspace of A with respect to eigenvalue a .
By definition, the dimension of $E(a)$ is the geometric multiplicity of eigenvalue a .

Generalized eigenspaces

- ▶ Let A be a complex $n \times n$ matrix and let a be an eigenvalue of A .
- ▶ Then

$$E(a) = \{x \in \mathbb{C}^n : (A - aI)x = 0\}$$

is known as the eigenspace of A with respect to eigenvalue a .
By definition, the dimension of $E(a)$ is the geometric multiplicity of eigenvalue a .

- ▶ The space

$$F(a) = \{x \in \mathbb{C}^n : (A - aI)^m x = 0, \text{ for some } m \in \mathbb{N}\}$$

is known as the **generalized eigenspace** of eigenvalue a .

Generalized eigenspaces

- ▶ Let A be a complex $n \times n$ matrix and let a be an eigenvalue of A .
- ▶ Then

$$E(a) = \{x \in \mathbb{C}^n : (A - aI)x = 0\}$$

is known as the eigenspace of A with respect to eigenvalue a .
By definition, the dimension of $E(a)$ is the geometric multiplicity of eigenvalue a .

- ▶ The space

$$F(a) = \{x \in \mathbb{C}^n : (A - aI)^m x = 0, \text{ for some } m \in \mathbb{N}\}$$

is known as the **generalized eigenspace** of eigenvalue a .

- ▶ The Jordan Canonical form tells us that the dimension of $F(a)$ is the algebraic multiplicity of eigenvalue a .

Generalized eigenspaces

- ▶ Let A be a complex $n \times n$ matrix and let a be an eigenvalue of A .
- ▶ Then

$$E(a) = \{x \in \mathbb{C}^n : (A - aI)x = 0\}$$

is known as the eigenspace of A with respect to eigenvalue a .
By definition, the dimension of $E(a)$ is the geometric multiplicity of eigenvalue a .

- ▶ The space

$$F(a) = \{x \in \mathbb{C}^n : (A - aI)^m x = 0, \text{ for some } m \in \mathbb{N}\}$$

is known as the **generalized eigenspace** of eigenvalue a .

- ▶ The Jordan Canonical form tells us that the dimension of $F(a)$ is the algebraic multiplicity of eigenvalue a .
- ▶ **END OF LECTURE 35.**

Lecture 36: Simultaneous diagonalization

- ▶ To begin with we have simultaneous upper triangularization.

Lecture 36: Simultaneous diagonalization

- ▶ To begin with we have simultaneous upper triangularization.
- ▶ **Theorem 36.1:** Let A_1, A_2 be two **commuting** $n \times n$ complex matrices. Then there exists a unitary U and two upper triangular matrices T_1, T_2 such that

$$A_j = UT_jU^*, \quad j = 1, 2.$$

Lecture 36: Simultaneous diagonalization

- ▶ To begin with we have simultaneous upper triangularization.
- ▶ **Theorem 36.1:** Let A_1, A_2 be two **commuting** $n \times n$ complex matrices. Then there exists a unitary U and two upper triangular matrices T_1, T_2 such that

$$A_j = UT_jU^*, \quad j = 1, 2.$$

- ▶ Here we are considering the standard inner product on $\mathbb{C}^n$.

Lecture 36: Simultaneous diagonalization

- ▶ To begin with we have simultaneous upper triangularization.
- ▶ **Theorem 36.1:** Let A_1, A_2 be two commuting $n \times n$ complex matrices. Then there exists a unitary U and two upper triangular matrices T_1, T_2 such that

$$A_j = UT_jU^*, \quad j = 1, 2.$$

- ▶ Here we are considering the standard inner product on $\mathbb{C}^n$.
- ▶ The point is that when A_1, A_2 commute, we can find a single unitary U such that both U^*A_1U and U^*A_2U are upper triangular.

- ▶ **Proof:** The proof is by induction. For $n = 1$, there is nothing to show.

Continuation

- ▶ **Proof:** The proof is by induction. For $n = 1$, there is nothing to show.
- ▶ Consider $n \geq 2$. Fix some eigenvalue a_1 of A_1 .
- ▶ Suppose $A_1 v = a_1 v$.

Continuation

- ▶ **Proof:** The proof is by induction. For $n = 1$, there is nothing to show.
- ▶ Consider $n \geq 2$. Fix some eigenvalue a_1 of A_1 .
- ▶ Suppose $A_1 v = a_1 v$.
- ▶ By commutativity, $A_1 A_2 v = A_2 A_1 v = A_2(a_1 v) = a_1(A_2 v)$.

Continuation

- ▶ **Proof:** The proof is by induction. For $n = 1$, there is nothing to show.
- ▶ Consider $n \geq 2$. Fix some eigenvalue a_1 of A_1 .
- ▶ Suppose $A_1 v = a_1 v$.
- ▶ By commutativity, $A_1 A_2 v = A_2 A_1 v = A_2(a_1 v) = a_1(A_2 v)$.
- ▶ In other words, the eigenspace $E_1 = \{v \in \mathbb{C}^n : A v = a_1 v\}$ is left invariant by A_2 .

Continuation

- ▶ Note that the dimension of E_1 is the geometric multiplicity of a_1 for the matrix A_1 , and it is at least 1.

Continuation

- ▶ Note that the dimension of E_1 is the geometric multiplicity of a_1 for the matrix A_1 , and it is at least 1.
- ▶ Let $\{v_1, v_2, \dots, v_k\}$ be an orthonormal basis for E_1 .

Continuation

- ▶ Note that the dimension of E_1 is the geometric multiplicity of a_1 for the matrix A_1 , and it is at least 1.
- ▶ Let $\{v_1, v_2, \dots, v_k\}$ be an orthonormal basis for E_1 .
- ▶ Extend it to an orthonormal basis $\mathcal{B} := \{v_1, \dots, v_n\}$ of $\mathbb{C}^n$.

Continuation

- ▶ Note that the dimension of E_1 is the geometric multiplicity of a_1 for the matrix A_1 , and it is at least 1.
- ▶ Let $\{v_1, v_2, \dots, v_k\}$ be an orthonormal basis for E_1 .
- ▶ Extend it to an orthonormal basis $\mathcal{B} := \{v_1, \dots, v_n\}$ of $\mathbb{C}^n$.
- ▶ Let U_0 be the unitary whose columns are $\{v_1, \dots, v_n\}$.

Continuation

- ▶ We have $A_1 v_j = a_1 v_j$ for $1 \leq j \leq k$.

Continuation

- ▶ We have $A_1 v_j = a_1 v_j$ for $1 \leq j \leq k$.
- ▶ This means that the linear map $x \mapsto A_1 x$, on the basis $\mathcal{B}$ has a block matrix form:

$$R = \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

Continuation

- ▶ We have $A_1 v_j = a_1 v_j$ for $1 \leq j \leq k$.
- ▶ This means that the linear map $x \mapsto A_1 x$, on the basis $\mathcal{B}$ has a block matrix form:

$$R = \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

- ▶ for some $k \times (n - k)$ matrix R_{12} and $(n - k) \times (n - k)$ matrix R_{22} or equivalently,

$$A_1 U_0 = U_0 R \quad (1)$$

Continuation

- ▶ We also know that A_2 leaves $E_1 = \text{span}\{v_1, \dots, v_k\}$ invariant.

Continuation

- ▶ We also know that A_2 leaves $E_1 = \text{span}\{v_1, \dots, v_k\}$ invariant.
- ▶ This means that the linear map $x \mapsto A_2 x$ has in the basis $\mathcal{B}$ has a block matrix form:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix},$$

Continuation

- ▶ We also know that A_2 leaves $E_1 = \text{span}\{v_1, \dots, v_k\}$ invariant.
- ▶ This means that the linear map $x \mapsto A_2 x$ has in the basis $\mathcal{B}$ has a block matrix form:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix},$$

- ▶ or equivalently,

$$A_2 U_0 = U_0 S \quad (2).$$

Continuation

- ▶ We also know that A_2 leaves $E_1 = \text{span}\{v_1, \dots, v_k\}$ invariant.
- ▶ This means that the linear map $x \mapsto A_2x$ has in the basis $\mathcal{B}$ has a block matrix form:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix},$$

- ▶ or equivalently,

$$A_2 U_0 = U_0 S \quad (2).$$

- ▶ From equations (1) and (2), we have

$$A_1 = U_0 R U_0^*, \quad A_2 = U_0 S U_0^*.$$

Continuation

- ▶ Now as A_1, A_2 , commute, R, S also commute and we get:

Continuation

- ▶ Now as A_1, A_2 , commute, R, S also commute and we get:



$$\begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix} \cdot \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

Continuation

- ▶ Now as A_1, A_2 , commute, R, S also commute and we get:



$$\begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix} \cdot \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

- ▶ By block matrix computations,

$$\begin{bmatrix} a_1 S_{11} & a_1 S_{12} + R_{12} S_{22} \\ 0 & R_{22} S_{22} \end{bmatrix} = \begin{bmatrix} a_1 S_{11} & S_{11} R_{12} + S_{12} R_{22} \\ 0 & S_{22} R_{22} \end{bmatrix}.$$

Continuation

- ▶ Now as A_1, A_2 , commute, R, S also commute and we get:



$$\begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix} \cdot \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

- ▶ By block matrix computations,

$$\begin{bmatrix} a_1 S_{11} & a_1 S_{12} + R_{12} S_{22} \\ 0 & R_{22} S_{22} \end{bmatrix} = \begin{bmatrix} a_1 S_{11} & S_{11} R_{12} + S_{12} R_{22} \\ 0 & S_{22} R_{22} \end{bmatrix}.$$

- ▶ In particular, R_{22} and S_{22} commute. Note that they have order $(n - k) \times (n - k)$ with $k \geq 1$. Hence the induction hypothesis is applicable.

Continuation

- ▶ Therefore, there exist a unitary W , two upper triangular matrices M_1, M_2 (all of order $(n - k) \times (n - k)$), such that

$$R_{22} = WM_1W^*, \quad S_{22} = WM_2W^*.$$

Continuation

- ▶ Therefore, there exist a unitary W , two upper triangular matrices M_1, M_2 (all of order $(n - k) \times (n - k)$), such that

$$R_{22} = WM_1W^*, \quad S_{22} = WM_2W^*.$$

- ▶ Further, by Schur's upper triangularization result, there exists a unitary Z and an upper triangular matrix X (all of order $(k \times k)$) such that $S_{11} = ZXZ^*$.

Continuation

- ▶ Therefore, there exist a unitary W , two upper triangular matrices M_1, M_2 (all of order $(n - k) \times (n - k)$), such that

$$R_{22} = WM_1W^*, \quad S_{22} = WM_2W^*.$$

- ▶ Further, by Schur's upper triangularization result, there exists a unitary Z and an upper triangular matrix X (all of order $(k \times k)$) such that $S_{11} = ZXZ^*$.
- ▶ We observe that,

$$\begin{aligned} A_1 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} a_1 I_k & Z^* R_{12} W \\ 0 & M_1 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} a_1 I_k & Z^* R_{12} W \\ 0 & M_1 \end{bmatrix} U^*, \end{aligned}$$

Continuation

- ▶ Therefore, there exist a unitary W , two upper triangular matrices M_1, M_2 (all of order $(n - k) \times (n - k)$), such that

$$R_{22} = WM_1W^*, \quad S_{22} = WM_2W^*.$$

- ▶ Further, by Schur's upper triangularization result, there exists a unitary Z and an upper triangular matrix X (all of order $(k \times k)$) such that $S_{11} = ZXZ^*$.
- ▶ We observe that,

$$\begin{aligned} A_1 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} a_1 I_k & Z^* R_{12} W \\ 0 & M_1 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} a_1 I_k & Z^* R_{12} W \\ 0 & M_1 \end{bmatrix} U^*, \end{aligned}$$

- ▶ where

$$U = U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix}$$

being a product of two unitaries is a unitary.

Continuation

► Similarly,

$$\begin{aligned} A_2 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} U^*, \end{aligned}$$

Continuation

► Similarly,

$$\begin{aligned} A_2 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} U^*, \end{aligned}$$

Continuation

► Similarly,

$$\begin{aligned} A_2 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} U^*, \end{aligned}$$

► where

$$U = U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix}$$

is a unitary.

Continuation

► Similarly,

$$\begin{aligned} A_2 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} U^*, \end{aligned}$$

► where

$$U = U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix}$$

is a unitary.

►

$$T_1 := \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & M_1 \end{bmatrix}$$

is upper triangular as M_1 is upper triangular.

Continuation

- ▶ Similarly,

$$\begin{aligned} A_2 &= U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix} \cdot \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} Z^* & 0 \\ 0 & W^* \end{bmatrix} U_0^* \\ &= U \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix} U^*, \end{aligned}$$

- ▶ where

$$U = U_0 \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix}$$

is a unitary.

- ▶

$$T_1 := \begin{bmatrix} a_1 I_k & R_{12} \\ 0 & M_1 \end{bmatrix}$$

is upper triangular as M_1 is upper triangular.

- ▶

$$T_2 := \begin{bmatrix} X & Z^* S_{12} W \\ 0 & M_2 \end{bmatrix}$$

is upper triangular as X M_2 are upper triangular

Simultaneous diagonalization

- **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.

Simultaneous diagonalization

- ▶ **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.
- ▶ **Proof:** This is clear from the previous theorem, as T_1, T_2 are upper triangular normal matrices they must be diagonal.

Simultaneous diagonalization

- ▶ **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.
- ▶ **Proof:** This is clear from the previous theorem, as T_1, T_2 are upper triangular normal matrices they must be diagonal.
- ▶ Hence take $D_1 = T_1$ and $D_2 = T_2$.

Simultaneous diagonalization

- ▶ **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.
- ▶ **Proof:** This is clear from the previous theorem, as T_1, T_2 are upper triangular normal matrices they must be diagonal.
- ▶ Hence take $D_1 = T_1$ and $D_2 = T_2$.
- ▶ **Corollary 36.3:** Suppose A_1, A_2 are normal matrices. Then A_1, A_2 are commuting if and only if there exists a normal matrix A with polynomials p_1, p_2 such that $A_1 = p_1(A), A_2 = p_2(A)$.

Simultaneous diagonalization

- ▶ **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.
- ▶ **Proof:** This is clear from the previous theorem, as T_1, T_2 are upper triangular normal matrices they must be diagonal.
- ▶ Hence take $D_1 = T_1$ and $D_2 = T_2$.
- ▶ **Corollary 36.3:** Suppose A_1, A_2 are normal matrices. Then A_1, A_2 are commuting if and only if there exists a normal matrix A with polynomials p_1, p_2 such that $A_1 = p_1(A), A_2 = p_2(A)$.
- ▶ **Proof:** If A_1, A_2 are commuting, by the previous theorem, we may assume that both A_1, A_2 are diagonal. Now take A as the diagonal matrix with j -th diagonal entry as j . It is easy to get polynomials p_1, p_2 so that $p_1(j) = (A_1)_{jj}, p_2(j) = (A_2)_{jj}$. Hence $p_1(A) = A_1, p_2(A) = A_2$.

Simultaneous diagonalization

- ▶ **Theorem 36.2:** Suppose A_1, A_2 are commuting normal matrices. Then there exists a unitary U with two diagonal matrices D_1, D_2 such that $A_1 = UD_1U^*$, $A_2 = UD_2U^*$.
- ▶ **Proof:** This is clear from the previous theorem, as T_1, T_2 are upper triangular normal matrices they must be diagonal.
- ▶ Hence take $D_1 = T_1$ and $D_2 = T_2$.
- ▶ **Corollary 36.3:** Suppose A_1, A_2 are normal matrices. Then A_1, A_2 are commuting if and only if there exists a normal matrix A with polynomials p_1, p_2 such that $A_1 = p_1(A), A_2 = p_2(A)$.
- ▶ **Proof:** If A_1, A_2 are commuting, by the previous theorem, we may assume that both A_1, A_2 are diagonal. Now take A as the diagonal matrix with j -th diagonal entry as j . It is easy to get polynomials p_1, p_2 so that $p_1(j) = (A_1)_{jj}, p_2(j) = (A_2)_{jj}$. Hence $p_1(A) = A_1, p_2(A) = A_2$.
- ▶ The converse is to show that for any normal matrix A , $p_1(A), p_2(A)$ commute for any two polynomials and is easy.

Families of commuting matrices

- **Theorem 36.4:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting matrices. Then there exists a unitary U with upper triangular matrices $T_1, \dots, T_k$ such that

$$A_j = UT_jU^*, \quad 1 \leq j \leq k.$$

Families of commuting matrices

- ▶ **Theorem 36.4:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting matrices. Then there exists a unitary U with upper triangular matrices $T_1, \dots, T_k$ such that

$$A_j = UT_jU^*, \quad 1 \leq j \leq k.$$

- ▶ **Proof:** Exercise.

Families of commuting matrices

- ▶ **Theorem 36.4:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting matrices. Then there exists a unitary U with upper triangular matrices $T_1, \dots, T_k$ such that

$$A_j = UT_jU^*, \quad 1 \leq j \leq k.$$

- ▶ **Proof:** Exercise.
- ▶ **Theorem 36.5:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting normal matrices. Then there exists a unitary U with diagonal matrices $D_1, \dots, D_k$ such that

$$A_j = UD_jU^*, \quad 1 \leq j \leq k.$$

Families of commuting matrices

- ▶ **Theorem 36.4:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting matrices. Then there exists a unitary U with upper triangular matrices $T_1, \dots, T_k$ such that

$$A_j = UT_jU^*, \quad 1 \leq j \leq k.$$

- ▶ **Proof:** Exercise.
- ▶ **Theorem 36.5:** Fix $k \geq 1$. Suppose $A_1, A_2, \dots, A_k$ are commuting normal matrices. Then there exists a unitary U with diagonal matrices $D_1, \dots, D_k$ such that

$$A_j = UD_jU^*, \quad 1 \leq j \leq k.$$

- ▶ **Proof:** Clear from the previous theorem.

Simultaneous Jordan form?

- Example 36.6: For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.
- ▶ A_1 is already in Jordan form. By the uniqueness of Jordan form, if $S^{-1}A_1S$ is in Jordan form, then $S^{-1}A_1S = A_1$.

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.
- ▶ A_1 is already in Jordan form. By the uniqueness of Jordan form, if $S^{-1}A_1S$ is in Jordan form, then $S^{-1}A_1S = A_1$.
- ▶ That is, $A_1S = SA_1$.

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.
- ▶ A_1 is already in Jordan form. By the uniqueness of Jordan form, if $S^{-1}A_1S$ is in Jordan form, then $S^{-1}A_1S = A_1$.
- ▶ That is, $A_1S = SA_1$.
- ▶ This implies $A_1^2S = SA_1^2$. Hence $S^{-1}A_2S = A_2$. But A_2 is not in Jordan form.

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.
- ▶ A_1 is already in Jordan form. By the uniqueness of Jordan form, if $S^{-1}A_1S$ is in Jordan form, then $S^{-1}A_1S = A_1$.
- ▶ That is, $A_1S = SA_1$.
- ▶ This implies $A_1^2S = SA_1^2$. Hence $S^{-1}A_2S = A_2$. But A_2 is not in Jordan form.
- ▶ Therefore simultaneous Jordan form is not possible for A_1 and A_2 .

Simultaneous Jordan form?

- ▶ **Example 36.6:** For $n \geq 3$, take $A_1 = J_0(n)$ and $A_2 = J_0(n)^2$.
- ▶ Then A_1 and A_2 are commuting.
- ▶ A_1 is already in Jordan form. By the uniqueness of Jordan form, if $S^{-1}A_1S$ is in Jordan form, then $S^{-1}A_1S = A_1$.
- ▶ That is, $A_1S = SA_1$.
- ▶ This implies $A_1^2S = SA_1^2$. Hence $S^{-1}A_2S = A_2$. But A_2 is not in Jordan form.
- ▶ Therefore simultaneous Jordan form is not possible for A_1 and A_2 .
- ▶ **END OF LECTURE 36.**

Lecture 37: Real symmetric matrices and quadratic forms

- ▶ We know that even if a matrix is real, in general the eigenvalues, i.e, roots of the characteristic polynomial, can be complex.

Lecture 37: Real symmetric matrices and quadratic forms

- ▶ We know that even if a matrix is real, in general the eigenvalues, i.e, roots of the characteristic polynomial, can be complex.
- ▶ **Example 37.1:** Take

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Lecture 37: Real symmetric matrices and quadratic forms

- ▶ We know that even if a matrix is real, in general the eigenvalues, i.e, roots of the characteristic polynomial, can be complex.
- ▶ **Example 37.1:** Take

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

- ▶ Then $A^2 = -I$ and the eigenvalues of A are $\pm i$.

Lecture 37: Real symmetric matrices and quadratic forms

- ▶ We know that even if a matrix is real, in general the eigenvalues, i.e, roots of the characteristic polynomial, can be complex.
- ▶ **Example 37.1:** Take

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

- ▶ Then $A^2 = -I$ and the eigenvalues of A are $\pm i$.
- ▶ Here A is a normal matrix and we can diagonalize it. So we can get a unitary U , such that

$$A = U \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} U^*.$$

Lecture 37: Real symmetric matrices and quadratic forms

- ▶ We know that even if a matrix is real, in general the eigenvalues, i.e, roots of the characteristic polynomial, can be complex.
- ▶ **Example 37.1:** Take

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

- ▶ Then $A^2 = -I$ and the eigenvalues of A are $\pm i$.
- ▶ Here A is a normal matrix and we can diagonalize it. So we can get a unitary U , such that

$$A = U \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} U^*.$$

- ▶ We can't diagonalize A or make it upper triangular in real field.

Real symmetric matrices

- ▶ What if we have a real symmetric matrix? Such matrices appear often in applications.

Real symmetric matrices

- ▶ What if we have a real symmetric matrix? Such matrices appear often in applications.
- ▶ If A is real symmetric, considering it as a complex self-adjoint matrix we know that all eigenvalues are real. So the obstruction for diagonalization mentioned above is not there.

Real symmetric matrices

- ▶ What if we have a real symmetric matrix? Such matrices appear often in applications.
- ▶ If A is real symmetric, considering it as a complex self-adjoint matrix we know that all eigenvalues are real. So the obstruction for diagonalization mentioned above is not there.
- ▶ Recall that a real matrix M is said to be orthogonal if $MM^t = M^tM = I$.

Real symmetric matrices

- ▶ What if we have a real symmetric matrix? Such matrices appear often in applications.
- ▶ If A is real symmetric, considering it as a complex self-adjoint matrix we know that all eigenvalues are real. So the obstruction for diagonalization mentioned above is not there.
- ▶ Recall that a real matrix M is said to be orthogonal if $MM^t = M^tM = I$.
- ▶ **Question:** Can we diagonalize a real symmetric matrix using orthogonal matrices?

Real symmetric matrices

- ▶ What if we have a real symmetric matrix? Such matrices appear often in applications.
- ▶ If A is real symmetric, considering it as a complex self-adjoint matrix we know that all eigenvalues are real. So the obstruction for diagonalization mentioned above is not there.
- ▶ Recall that a real matrix M is said to be orthogonal if $MM^t = M^tM = I$.
- ▶ **Question:** Can we diagonalize a real symmetric matrix using orthogonal matrices?
- ▶ This is answered by the following theorems.

Continuation

- **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

Continuation

- ▶ **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

- ▶ We use the following simple lemma.

Continuation

- ▶ **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

- ▶ We use the following simple lemma.
- ▶ **Lemma 37.2:** Let A be a real matrix and let d be a real eigenvalue of A . Then there exists a real non-zero vector w such that $Aw = dw$ (In other words, there is a real eigenvector.)

Continuation

- ▶ **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

- ▶ We use the following simple lemma.
- ▶ **Lemma 37.2:** Let A be a real matrix and let d be a real eigenvalue of A . Then there exists a real non-zero vector w such that $Aw = dw$ (In other words, there is a real eigenvector.)
- ▶ **Proof:** Suppose $v \in \mathbb{C}^n$ is non-zero and $Av = dv$.

Continuation

- ▶ **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

- ▶ We use the following simple lemma.
- ▶ **Lemma 37.2:** Let A be a real matrix and let d be a real eigenvalue of A . Then there exists a real non-zero vector w such that $Aw = dw$ (In other words, there is a real eigenvector.)
- ▶ **Proof:** Suppose $v \in \mathbb{C}^n$ is non-zero and $Av = dv$.
- ▶ Let $v = x + iy$, where x, y are real vectors.

Continuation

- ▶ **Theorem 37.1:** Let A be a real matrix with **only real eigenvalues**. Then there exists an orthogonal matrix M and an upper triangular matrix T such that

$$A = MTM^t.$$

- ▶ We use the following simple lemma.
- ▶ **Lemma 37.2:** Let A be a real matrix and let d be a real eigenvalue of A . Then there exists a real non-zero vector w such that $Aw = dw$ (In other words, there is a real eigenvector.)
- ▶ **Proof:** Suppose $v \in \mathbb{C}^n$ is non-zero and $Av = dv$.
- ▶ Let $v = x + iy$, where x, y are real vectors.
- ▶ From $Av = dv$, we get $A(x + iy) = dx + idy$. Since A has real entries and d is real, by comparing the real and imaginary parts we get $Ax = dx$ and $Ay = idy$. As $v \neq 0$, at least one of x or y is non-zero and take that as w . ■.

Upper triangular form

- ▶ **Proof of Theorem 37.1:** Using the previous Lemma, the result can be proved through induction exactly like the proof the Schur's upper triangularization theorem.

Upper triangular form

- ▶ **Proof of Theorem 37.1:** Using the previous Lemma, the result can be proved through induction exactly like the proof the Schur's upper triangularization theorem.
- ▶ Just observe that the matrices of lower order appearing in the induction hypothesis also have real entries and real eigenvalues.

Diagonalization of real symmetric matrices

- **Theorem 37.3:** Let A be a real symmetric matrix. Then there exists an orthogonal matrix M and a real diagonal matrix D such that,

$$A = MDM^t.$$

Diagonalization of real symmetric matrices

- ▶ **Theorem 37.3:** Let A be a real symmetric matrix. Then there exists an orthogonal matrix M and a real diagonal matrix D such that,

$$A = MDM^t.$$

- ▶ **Proof:** By considering real symmetric matrices as complex self-adjoint matrices we know that they have only real eigenvalues.

Diagonalization of real symmetric matrices

- ▶ **Theorem 37.3:** Let A be a real symmetric matrix. Then there exists an orthogonal matrix M and a real diagonal matrix D such that,

$$A = MDM^t.$$

- ▶ **Proof:** By considering real symmetric matrices as complex self-adjoint matrices we know that they have only real eigenvalues.
- ▶ So the previous theorem is applicable.

Diagonalization of real symmetric matrices

- ▶ **Theorem 37.3:** Let A be a real symmetric matrix. Then there exists an orthogonal matrix M and a real diagonal matrix D such that,

$$A = MDM^t.$$

- ▶ **Proof:** By considering real symmetric matrices as complex self-adjoint matrices we know that they have only real eigenvalues.
- ▶ So the previous theorem is applicable.
- ▶ Now the result is immediate as symmetric upper triangular matrices are diagonal. ■

Jordan Canonical form

- **Theorem 37.4:** Let A be a real matrix with only real eigenvalues. Then there exists a real invertible matrix M such that $M^{-1}AM$ is in Jordan form, that is, it is a direct sum of Jordan blocks.

Jordan Canonical form

- ▶ **Theorem 37.4:** Let A be a real matrix with only real eigenvalues. Then there exists a real invertible matrix M such that $M^{-1}AM$ is in Jordan form, that is, it is a direct sum of Jordan blocks.
- ▶ The proof of this is omitted.

Real quadratic forms

- **Definition 37.5** Fix $n \in \mathbb{N}$. An n -variable **quadratic form** Q is a function $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$Q(x_1, \dots, x_n) = \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j.$$

Real quadratic forms

- **Definition 37.5** Fix $n \in \mathbb{N}$. An n -variable **quadratic form** Q is a function $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$Q(x_1, \dots, x_n) = \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j.$$

- In other words, it is a quadratic polynomial in n -variables with no constant term.

Real quadratic forms

- **Definition 37.5** Fix $n \in \mathbb{N}$. An n -variable **quadratic form** Q is a function $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$Q(x_1, \dots, x_n) = \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j.$$

- In other words, it is a quadratic polynomial in n -variables with no constant term.
- **Theorem 37.6:** Suppose Q is a quadratic form as above. Then there exists unique real symmetric $n \times n$ matrix A such that

$$Q(x) = \langle x, Ax \rangle, \quad x \in \mathbb{R}^n$$

Real quadratic forms

- **Definition 37.5** Fix $n \in \mathbb{N}$. An n -variable **quadratic form** Q is a function $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$Q(x_1, \dots, x_n) = \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j.$$

- In other words, it is a quadratic polynomial in n -variables with no constant term.
- **Theorem 37.6:** Suppose Q is a quadratic form as above. Then there exists unique real symmetric $n \times n$ matrix A such that

$$Q(x) = \langle x, Ax \rangle, \quad x \in \mathbb{R}^n$$

- Note that here we are considering standard inner product on $\mathbb{R}^n$.

Continuation

► **Proof:** Take $a_{ii} = b_i$, $1 \leq i \leq n$ and $a_{ij} = \frac{1}{2}c_{ij}$, $i \neq j$.

Continuation

- ▶ **Proof:** Take $a_{ii} = b_i$, $1 \leq i \leq n$ and $a_{ij} = \frac{1}{2}c_{ij}$, $i \neq j$.
- ▶ Then $A = [a_{ij}]_{1 \leq i, j \leq n}$ is a real symmetric matrix and

Continuation

- ▶ **Proof:** Take $a_{ii} = b_i$, $1 \leq i \leq n$ and $a_{ij} = \frac{1}{2}c_{ij}$, $i \neq j$.
- ▶ Then $A = [a_{ij}]_{1 \leq i, j \leq n}$ is a real symmetric matrix and
- ▶

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i,j=1}^n x_i a_{ij} x_j \\&= \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i \neq j} a_{ij} x_i x_j \\&= \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j \\&= Q(x).\end{aligned}$$

Continuation

- ▶ **Proof:** Take $a_{ii} = b_i$, $1 \leq i \leq n$ and $a_{ij} = \frac{1}{2}c_{ij}$, $i \neq j$.
- ▶ Then $A = [a_{ij}]_{1 \leq i, j \leq n}$ is a real symmetric matrix and
- ▶

$$\begin{aligned}\langle x, Ax \rangle &= \sum_{i,j=1}^n x_i a_{ij} x_j \\&= \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i \neq j} a_{ij} x_i x_j \\&= \sum_{i=1}^n b_i x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j \\&= Q(x).\end{aligned}$$

- ▶ The uniqueness is clear from comparison of coefficients. ■.

Diagonalization

- ▶ Consider a real quadratic form Q . With out loss of generality, we may take

$$Q(x) = \langle x, Ax \rangle = \sum_{i,j=1}^n a_{ij}x_i x_j$$

for some real symmetric matrix A .

Diagonalization

- ▶ Consider a real quadratic form Q . With out loss of generality, we may take

$$Q(x) = \langle x, Ax \rangle = \sum_{i,j=1}^n a_{ij}x_i x_j$$

for some real symmetric matrix A .

- ▶ We know that we can diagonalize A . Hence there exists a diagonal matrix D and an orthogonal matrix M such that

$$A = MDM^{-1}.$$

Diagonalization

- ▶ Consider a real quadratic form Q . With out loss of generality, we may take

$$Q(x) = \langle x, Ax \rangle = \sum_{i,j=1}^n a_{ij}x_i x_j$$

for some real symmetric matrix A .

- ▶ We know that we can diagonalize A . Hence there exists a diagonal matrix D and an orthogonal matrix M such that

$$A = MDM^{-1}.$$

- ▶ Define new variables, $y_1, \dots, y_n$, by taking

$$y = M^{-1}x.$$

Diagonalization

- ▶ Consider a real quadratic form Q . With out loss of generality, we may take

$$Q(x) = \langle x, Ax \rangle = \sum_{i,j=1}^n a_{ij}x_i x_j$$

for some real symmetric matrix A .

- ▶ We know that we can diagonalize A . Hence there exists a diagonal matrix D and an orthogonal matrix M such that

$$A = MDM^{-1}.$$

- ▶ Define new variables, $y_1, \dots, y_n$, by taking

$$y = M^{-1}x.$$

- ▶ Then,

$$Q(x) = \langle x, MDM^{-1}x \rangle = \langle y, Dy \rangle.$$

- ▶ Let $d_1, \dots, d_n$ be the diagonal entries of D .

- ▶ Let $d_1, \dots, d_n$ be the diagonal entries of D .
- ▶ We get,

$$Q(x) = \sum_{j=1}^n d_j y_j^2.$$

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.
- ▶ Then we get

$$Q(x) = \sum_{j=1}^n d_j y_j^2 = \sum_{j=1}^n z_j^2,$$

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.
- ▶ Then we get

$$Q(x) = \sum_{j=1}^n d_j y_j^2 = \sum_{j=1}^n z_j^2,$$

- ▶ where $z = (D)^{\frac{1}{2}} M^{-1} y$.

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.
- ▶ Then we get

$$Q(x) = \sum_{j=1}^n d_j y_j^2 = \sum_{j=1}^n z_j^2,$$

- ▶ where $z = (D)^{\frac{1}{2}} M^{-1} y$.
- ▶ In other words, every positive quadratic form can be written as a 'sum of squares'.
- ▶ You may see this being applied in Statistics and in other areas.

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.
- ▶ Then we get

$$Q(x) = \sum_{j=1}^n d_j y_j^2 = \sum_{j=1}^n z_j^2,$$

- ▶ where $z = (D)^{\frac{1}{2}} M^{-1} y$.
- ▶ In other words, every positive quadratic form can be written as a 'sum of squares'.
- ▶ You may see this being applied in Statistics and in other areas.
- ▶ **END OF LECTURE 37.**

Positive quadratic forms

- ▶ Assume that the matrix A is positive (that is, the quadratic form $Q(x) = \langle x, Ax \rangle \geq 0, \forall x$).
- ▶ Then we know that the eigenvalues $d_j \geq 0$ for every j .
- ▶ Take $z_j = \sqrt{d_j} y_j$.
- ▶ Then we get

$$Q(x) = \sum_{j=1}^n d_j y_j^2 = \sum_{j=1}^n z_j^2,$$

- ▶ where $z = (D)^{\frac{1}{2}} M^{-1} y$.
- ▶ In other words, every positive quadratic form can be written as a 'sum of squares'.
- ▶ You may see this being applied in Statistics and in other areas.
- ▶ **END OF LECTURE 37.**
- ▶ **END OF THE COURSE**