Chapter 0. Introduction to the Mathematical Method

C.O.S. Sorzano

Biomedical Engineering

September 7, 2013

Outline

0. Mathematical language
   - Axioms, postulates, definitions and propositions (a)
   - Logical operators (a)
   - Qualifiers (b)
   - Mathematical proofs
     - Modus ponens (b)
     - Modus tollens (c)
     - Reductio ad absurdum (c)
     - Induction (c)
     - Case distinction (c)
     - Counterexample (c)
   - Common math mistakes (c)

References

M. de Guzmán Ozámiz. Cómo hablar, demostrar y resolver en Matemáticas. Anaya (2003)

A little bit of history

Modern logic is based on precise calculus rules and was born in the middle of the XIXth century with Gottfried Leibniz (1847), George Boole (1847), Augustus de Morgan (1847) and Bertrand Russell (1910).

To know more about the history of logic visit

- http://individual.utoronto.ca/pking/miscellaneous/history-of-logic.pdf
- http://en.wikipedia.org/wiki/History_of_logic

Outline

0. Mathematical language
   - Axioms, postulates, definitions and propositions (a)
   - Logical operators (a)
   - Qualifiers (b)
   - Mathematical proofs
     - Modus ponens (b)
     - Modus tollens (c)
     - Reductio ad absurdum (c)
     - Induction (c)
     - Case distinction (c)
     - Counterexample (c)
   - Common math mistakes (c)

Axioms, postulates and propositions

Axioms, postulates and propositions

Mathematical language has to be **uniform** (everybody must use it in the same way) and **univocal** (i.e., without any kind of ambiguity). We start from some initial statements called **axioms, postulates and definitions**. These elements are not questioned, they are not true or false, they simply are, and they serve to build a logical reasoning.

Example

- Axiom: If A and B are equal to C, then A is equal to B.
- Postulate: For any two points, there is a unique straight line that joins them.
- Definition: A prime number is a natural number that can only be divided by 1 and itself.

Propositions

Propositions

Based on axioms, postulates and definitions, we can construct **propositions** that are statements that refer to already introduced objects. Propositions can be **true or false**. They are named with capital letters A, B, C, ...

Example

2+3 (is not a proposition)

A: 2+3=5 (is a true proposition)

B: 2+3=7 (is a false proposition)

Construction of new propositions

We can construct new propositions using already existing ones and logical operators

Example

A: 2+2=4 (true)

B: 2+3=5 (true)

C: 2+3=7 (false)

D: A y B (true)

E: A o C (true)

and quantifiers

Example

A: **Some** numbers are prime (true)

B: **All** even numbers can be divided by 2 (true)

C: **None** of odd numbers can be divided by 2 (true)

Outline

0 Mathematical language
- Axioms, postulates, definitions and propositions (a)
- Logical operators (a)
- Qualifiers (b)
- Mathematical proofs
  - Modus ponens (b)
  - Modus tollens (c)
  - Reductio ad absurdum (c)
  - Induction (c)
  - Case distinction (c)
  - Counterexample (c)
- Common math mistakes (c)

$\overline{A}$ (not A)

Definition

$\overline{A}$ is true if A is false, and $\overline{A}$ is false if A is true.

Example

A: 3+2=5 (true)

B: $\neg A \equiv 3 + 2 \neq 5$ (false)

C: 3+2=6 (false)

D: $\neg C \equiv 3 + 2 \neq 6$ (true)

Truth table

A	$\overline{A}$
F	T
T	F

Properties

$$\overline{\overline{A}} = A$$

$\overline{A}$ (not A)

A double negation is a positive statement.

Example

A: 3+2=5 (true)

B: $\overline{A} \equiv 3 + 2 \neq 5$ (false)

C: $\overline{B} \equiv 3 + 2 = 5$ (true)

Example

It is not true that John is not at home.

A: John is at home

B: $\overline{A} \equiv$ Not (John is at home) $\equiv$ John is not at home

C: $\overline{B} \equiv$ Not (John is not at home) $\equiv$ John is at home $\equiv$ A

If C is true, then A is true. Therefore, John is at home.

A ∩ B (A and B)

Truth table

A	B	$A \cap B$
F	F	F
F	T	F
T	F	F
T	T	T

Properties

$A \cap B = B \cap A$

Example

A: 3+2=5 (true)

B: 2+2=4 (true)

C: $A \cap B \equiv$ 3+2=5 and 2+2=4 (true)

D: 3+2=6 (false)

E: $D \cap B \equiv$ 3+2=6 and 2+2=4 (false)

$A \cap B$ (A and B)

The common language AND is sometimes equivalent to the mathematical AND

Example

Triangle ABC and triangle A'B'C' are equilateral $\Rightarrow$

A: ABC is equilateral

B: A'B'C' is equilateral

C: $A \cap B \equiv$ Triangle ABC is equilateral AND Triangle A'B'C' is equilateral

and sometimes not

Example

Triangle ABC and triangle A'B'C' are similar $\not\Rightarrow$

A: ABC is similar

B: A'B'C' is similar

C: $A \cap B \equiv$ Triangle ABC is similar AND Triangle A'B'C' is similar

$A \cup B$ (A or B; A and/or B)

15% discounts for customers having a student card or university card. Of course, people with both cards have a 15% discount. Inclusive OR.

Truth table

A	B	$A \cup B$
F	F	F
F	T	T
T	F	T
T	T	T

Properties

$A \cup B = B \cup A$

Example

A: 3+2=5 (true)

B: 2+2=4 (true)

C: $A \cup B \equiv$ 3+2=5 or 2+2=4 (true)

D: 3+2=6 (false)

E: $D \cup B \equiv$ 3+2=6 or 2+2=4 (true)

$A \oplus B$ (either A or B; A xor B (eXclusive or))

We'll go to Paris or Berlin. Either Paris or Berlin, we cannot go to both places at the same time. Exclusive OR.

Truth table

A	B	$A \oplus B$
F	F	F
F	T	T
T	F	T
T	T	F

Properties

$A \oplus B = B \oplus A$

Example

A: a<5

B: a=5

C: $A \oplus B \equiv a \leq 5$

If $a = 3$, then C is true. If $a = 6$, then C is false.

Negation of and

$\overline{A \cap B} = \overline{A} \cup \overline{B}$
This is one of Morgan's laws.

A	B	$A \cap B$	$\overline{A \cap B}$	$\overline{A}$	$\overline{B}$	$\overline{A} \cup \overline{B}$
F	F	F	**T**	T	T	**T**
F	T	F	**T**	T	F	**T**
T	F	F	**T**	F	T	**T**
T	T	T	**F**	F	F	**F**

Example

A: It rained on Monday

B: It rained on Tuesday

C: $\overline{A \cap B} \equiv$ It is not true that it rained on both days $\equiv$ Either it did not rain on Monday or it did not rain on Tuesday.

Negation of or

Inclusive OR: $\overline{A \cup B} = \overline{A} \cap \overline{B}$
This is another Morgan's law.

A	B	$A \cup B$	$\overline{A \cup B}$	$\overline{A}$	$\overline{B}$	$\overline{A} \cap \overline{B}$
F	F	F	**T**	T	T	**T**
F	T	T	**F**	T	F	**F**
T	F	T	**F**	F	T	**F**
T	T	T	**F**	F	F	**F**

Exclusive OR: $\overline{A \oplus B} = (\overline{A} \cap B) \cup (A \cap \overline{B})$

A	B	$A \oplus B$	$\overline{A \oplus B}$	$\overline{A}$	$\overline{B}$	$\overline{A} \cap B$	$A \cap \overline{B}$	$(\overline{A} \cap B) \cup (A \cap \overline{B})$
F	F	F	**T**	T	T	F	F	**T**
F	T	T	**F**	T	F	T	F	**F**
T	F	T	**F**	F	T	F	T	**F**
T	T	F	**T**	F	F	F	F	**T**

$A \Rightarrow B$ (A implies B)

Natural language

- A implies B
- A is sufficient for B
- A guarantees B
- B is necessary for A
- If A, then B
- If not B, then not A

Truth table

A	B	$A \Rightarrow B$
F	F	T
F	T	T
T	F	F
T	T	T

$A \Rightarrow B$ (A implies B)

In natural language "If ..., then ..." is not used in the mathematical sense.

Example

If it rains, I'll stay at home.

If he is at home, is it raining?
We don't know, he didn't say what he would do if it was not raining.

Example

I'm going to the bank. If it is open, I'll bring 1000 euros.

If he is back with 1000 euros, is the bank open?
We don't know, maybe a very good friend of his gave him 1000 euros.

$A \Rightarrow B$ (If not B, then not A)

Example

I'm going to the bank. If it is open, I'll bring 1000 euros.

If I'm back without 1000 euros, is the bank open?
No, let's see why

A: Bank is open

B: I bring 1000 euros

A	B	$A \Rightarrow B$	Why
F	F	T	The bank was closed
F	T	T	A friend gave me
T	F	F	I lied
T	T	T	I withdrew 1000 euros from bank

There is only one situation in which my statement is true (I did not lie) and in which I do not bring 1000 euros (B is false) that is when the bank is closed (A is also false).

$A \Rightarrow B$ (If not B, then not A)

We can generally formulate this analysis as

Properties

$$A \Rightarrow B = \overline{B} \Rightarrow \overline{A}$$

A	B	$A \Rightarrow B$	$\overline{B}$	$\overline{A}$	$\overline{B} \Rightarrow \overline{A}$
F	F	**T**	T	T	**T**
F	T	**T**	F	T	**T**
T	F	**F**	T	F	**F**
T	T	**T**	F	F	**T**

$A \Rightarrow B$ (Not (A and not B))

Another interesting property

Properties

$$A \Rightarrow B = \overline{A \cap \overline{B}}$$
$$\overline{A \Rightarrow B} = A \cap \overline{B}$$

The proof of these properties is left to the reader.

Example

I'm going to the bank. If it is open, I'll bring 1000 euros.

It is equivalent to:

It will not be the case that (the bank is open (A) and I don't bring 1000 euros (not B)).

$A \Leftrightarrow B$ (A if and only if B)

Truth table

A	B	$A \Leftrightarrow B$
F	F	T
F	T	F
T	F	F
T	T	T

Properties

$$A \Leftrightarrow B = (A \Rightarrow B) \cap (B \Rightarrow A)$$

In plain language, we say:
A is necessary and sufficient for B
B is necessary and sufficient for A

Outline

0 Mathematical language

- Axioms, postulates, definitions and propositions (a)
- Logical operators (a)
- Qualifiers (b)
- Mathematical proofs
  - Modus ponens (b)
  - Modus tollens (c)
  - Reductio ad absurdum (c)
  - Induction (c)
  - Case distinction (c)
  - Counterexample (c)
- Common math mistakes (c)

Qualifiers

Example

There might be **a** person that reads **all** newspapers **every** day.
Every day, there might be **a** person that reads **all** newspapers.
Every one reads a newspaper **every** day.
Every day, there is **a** newspaper that **everybody** reads.

Example

We say that the limit of the function $f(x)$ when x goes to x_0 is y if and only if **for all** positive numbers (ϵ), **there exists** another positive number (δ) **such that** if the distance between x and x_0 is smaller than δ, then the distance between $f(x)$ and y is smaller than ϵ.

$$\lim_{x \to x_0} f(x) = y \Leftrightarrow \forall \epsilon > 0 \, \exists \delta > 0 \mid |x - x_0| < \delta \Rightarrow |f(x) - y| < \epsilon$$

$\forall$ (for all), $\exists$ (exists) and $\exists!$ (exists only one)

For all x in P	$\forall x, x \in P$; $\forall x \in P$
For any x in P	$\forall x, x \in P$; $\forall x \in P$
For each x with the property P	$\forall x, P(x)$
There **exists at least one** x in P	$\exists x, x \in P$; $\exists x \in P$
For at least one x in P	$\exists x, x \in P$; $\exists x \in P$
There **exists at least one** x with the property P	$\exists x, P(x)$
There **exists exactly one** x in P	$\exists! x, x \in P$; $\exists! x \in P$

Example

For all real numbers	$\forall x \in \mathbb{R}$
For all real numbers smaller than 4	$\forall x \in \mathbb{R}, x < 4$
There **exists at least one** real number	$\exists x \in \mathbb{R}$
There **exists at least one** real number greater than 2	$\exists x \in \mathbb{R}, x > 2$
There **exists a single** real number such that ...	$\exists! x \in \mathbb{R} \mid$ ...

$|$ (such that, it is verified, verifying)

Example

There must be people that read all newspapers everyday. Let P be the set of all persons, let N be the set of all newspapers, and let D be the set of all days. Then, the previous sentence is formalized as

$\exists p \in P | \forall d \in D | \forall n \in N|$ p reads n on d.

Literal reading: There exist at least one person **such that** for all days and for all newspapers **it is verified that** p reads n on d.

Example

Every day, there must be someone that reads all newspapers.

$\forall d \in D | \exists p \in P | \forall n \in N|$ p reads n on d.

Literal reading: For all days **it is verified that** there exists at least one person **verifying that** for all newspapers **it is verified that** p reads n on d.

$|$ (such that, it is verified, verifying)

Example

$$\lim_{x \to x_0} f(x) = y \Leftrightarrow \forall \epsilon > 0 \, \exists \delta > 0 \mid |x - x_0| < \delta \Rightarrow |f(x) - y| < \epsilon$$

Literal reading: the limit of $f(x)$ when x goes to x_0 is y if and only if for any ϵ greater than 0, there exists δ greater than 0 **such that** if $|x - x_0| < \delta$ is true, then $|f(x) - y| < \epsilon$ is also true.

Example

Fermat-Wiles Theorem:
$\forall n \in \mathbb{Z}, n > 2 | \forall (x, y, z) \in \mathbb{R}^3, x^n + y^n = z^n | xyz = 0$

Literal reading: For all integer numbers **it is verified that** for any real numbers x, y z with the property $x^n + y^n = z^n$ **it is verified that** at least one of the three numbers is 0.

Negation of qualifiers

Let's say we state that all elements in a given set S has a certain property ($\forall x \in S | P(x)$). The negation of this statement is that there exists at least one element of S that does not have that property ($\exists x \in S | \overline{P(x)}$).

Similarly, if we state that there exists at least one element in a given set S that has a certain property ($\exists x \in S | P(x)$). The negation of this statement is that none of the elements of S have that property ($\forall x \in S | \overline{P(x)}$).

Negation of qualifiers

Example

In a previous example we had: **There must be people that read all newspapers everyday**. Its negation is

$$\overline{\exists p \in P | \forall d \in D | \forall n \in N | \text{p reads n on d.}} =$$
$$\forall p \in P | \overline{\forall d \in D | \forall n \in N | \text{p reads n on d.}} =$$
$$\forall p \in P | \exists d \in D | \overline{\forall n \in N | \text{p reads n on d.}} =$$
$$\forall p \in P | \exists d \in D | \exists n \in N | \overline{\text{p reads n on d.}} =$$
$$\forall p \in P | \exists d \in D | \exists n \in N | \text{p does not read n on d.}$$

That is, **For everybody, there is at least one day and one paper, such that p did not read n on d.**

Outline

Modus ponens

The following proofs follow a reasoning model called *Modus ponens* which is formally written as

$$(A \cap (A \Rightarrow B)) \Rightarrow B.$$

The intuitive meaning is that if A is true and $A \Rightarrow B$, then B is also true. Most proofs follow this way of reasoning. They can be performed in a forward way

$$A \Rightarrow B_1 \Rightarrow B_2 \Rightarrow ... \Rightarrow B$$

or in a backward way

$$B \Leftarrow B_n \Leftarrow B_{n-1} \Leftarrow ... \Leftarrow A.$$

Forward proofs ($A \Rightarrow B$; B is necessary for A)

Example

Prove that the third power of an odd number is odd.

Proof

Let there be the following propositions:

A: x is odd.

B: x^3 is odd.

We need to prove that $A \Rightarrow B$ (B is necessary for A).

Proof $A \Rightarrow B$

Since x is an odd number we can write $x = 2k + 1$ for some integer number k. Then,
$x^3 = (2k+1)^3 = 8k^3 + 12k^2 + 6k + 1 = 2(4k^3 + 6k^2 + 3k) + 1 = 2k' + 1$.
For $k' = 4k^3 + 6k^2 + 3k$, which is another integer number. Therefore, x^3 is odd.

Forward proofs ($A \Rightarrow B$; B is necessary for A)

Example

A necessary condition for a natural number to be a multiple of 360 is that it is a multiple of 3 and 120.

Proof

Let there be the following propositions:

A: To be multiple of 360

B: To be multiple of 3 and 120

We need to prove that $A \Rightarrow B$ (B is necessary for A).

Proof $A \Rightarrow B$

Let x be a multiple of 360 (A) $\Rightarrow$ There exists a natural number k such that $x = 360 \cdot k \Rightarrow x = 120 \cdot 3 \cdot k$. From this factorization, it is obvious that x is a multiple of 120 and a multiple of 3 (B).

Forward proofs ($A \Leftarrow B$; B is sufficient for A)

Example

A sufficient condition for a natural number to be a multiple of 360 is that it is a multiple of 3 and 120. *Proof*

Let there be the following propositions:

A: To be multiple of 360

B1: To be multiple of 3

B2: To be multiple of 120

B: B1 $\cap$ B2

We need to prove that $B \Rightarrow A$ (B is sufficient for A).

Proof $B \Rightarrow A$

We can easily prove that $B \not\Rightarrow A$ with a counterexample. Let us consider $x = 240$. It is a multiple of 3 (B1). It is a multiple of 120 (B2). Therefore, B is true. However, 240 is not a multiple of 360 (A is false). Therefore, we have proved that $B \not\Rightarrow A$.

Forward proofs ($A \Rightarrow B$)

Example

Show that $x = \frac{-b+\sqrt{b^2-4ac}}{2a}$ is solution of the equation $ax^2 + bx + c = 0$

Proof

Let there be the following propositions:

A: $x = \frac{-b+\sqrt{b^2-4ac}}{2a}$

B: $ax^2 + bx + c = 0$

We need to prove that $A \Rightarrow B$.

If $A \Rightarrow B$ is true, then it must also be true that $A \Rightarrow B_1$

$$B_1\text{: } a\left(\frac{-b+\sqrt{b^2-4ac}}{2a}\right)^2 + b\frac{-b+\sqrt{b^2-4ac}}{2a} + c = 0$$

that we can rewrite as

$$B_1\text{: } a\left(\frac{b^2}{4a^2} + \frac{b^2-4ac}{4a^2} - \frac{2b\sqrt{b^2-4ac}}{4a^2}\right) + b\frac{-b+\sqrt{b^2-4ac}}{2a} + c = 0$$

Forward proofs ($A \Rightarrow B$)

Example (continued)

that we can simplify to

$$B_1\colon\ \frac{b^2}{4a} + \frac{b^2}{4a} - c - \frac{b\sqrt{b^2-4ac}}{2a} + \frac{-b^2}{2a} + \frac{b\sqrt{b^2-4ac}}{2a} + c = 0$$

$$B_1\colon\ \cancel{\frac{b^2}{4a}} + \cancel{\frac{b^2}{4a}} - \cancel{c} - \cancel{\frac{b\sqrt{b^2-4ac}}{2a}} + \cancel{\frac{-b^2}{2a}} + \cancel{\frac{b\sqrt{b^2-4ac}}{2a}} + \cancel{c} = 0$$

$$B_1\colon\ 0 = 0$$

Since B_1 is always true (a statement that is always true is called a tautology), then $A \Rightarrow B_1$ is true, as we wanted.

Forward proofs ($A \Leftrightarrow B$)

In this case we have to prove both directions: $A \Rightarrow B$ and $B \Rightarrow A$.

Example

A necessary and sufficient condition for a natural number to be a multiple of 360 is that it is a multiple of 5 and 72.

Proof

Let there be the following propositions:

A: To be multiple of 360

B1: To be multiple of 5

B2: To be multiple of 72

B: B1 $\cap$ B2

We need to prove that $A \Leftrightarrow B$, that is, $A \Rightarrow B$ and $B \Rightarrow A$

Proof $A \Rightarrow B$

Let x be a multiple of 360 (A) $\Rightarrow$ There exists a natural number k such that $x = 360 \cdot k \Rightarrow x = 72 \cdot 5 \cdot k$. From this factorization, it is obvious that x is a multiple of 72 and a multiple of 5 (B).

Forward proofs ($A \Leftrightarrow B$)

Example (continued)

Proof $B \Rightarrow A$

$B1 \Rightarrow$ There is a natural number k_1 such that $x = 5 \cdot k_1$
$B2 \Rightarrow$ There is a natural number k_2 such that $x = 72 \cdot k_2$
Therefore, $5k_1 = 72k_2 \Rightarrow k_1 = \frac{72}{5}k_2$. But k_1 is a natural number not a rational number, therefore, k_2 needs to be a multiple of 5, i.e., there exists a natural number k_3 such that $k_2 = 5 \cdot k_3$. Consequently, considering B2, we have $x = 72 \cdot 5 \cdot k_3 = 360 \cdot k_3$. That is x is a multiple of 360. Therefore, we have proved that $A \Rightarrow B$.

More forward proofs

If I want to prove that A does not imply B ($A \Rightarrow B$ is false), I have to prove that B is false, but A is true.

Example

In our example, I have to prove that you did not bring 1000 euros (B is false), but the bank is open (A is true). I don't have to prove that

- B is false (you did not bring 1000 euros)
- A is false (the bank is closed)
- B is true but A is false (you brought 1000 euros, but the bank is closed)
- A and B are false (you did not bring 1000 euros, and the bank is closed)

More forward proofs

If I know that B is false, and I want to proof that A implies B ($A \Rightarrow B$ is true), then I have to prove that A is also false.

A	B	$A \Rightarrow B$
F	**F**	**T**
F	T	T
T	F	F
T	T	T

Example

If I know that you did not bring 1000 euros (B is false), all I have to prove to show that $A \Rightarrow B$ is true, is that the bank is closed (A is false).

More forward proofs

If I want to proof that A implies B or C ($A \Rightarrow B \cup C$ is true), and I prove that it is false that $A \Rightarrow B \cap C$, have I finished? No,let's see why

A	B	C	$B \cup C$	$A \Rightarrow B \cup C$	$B \cap C$	$A \Rightarrow B \cap C$
F	F	F	F	T	F	T
F	F	T	T	T	F	T
F	T	F	T	T	F	T
F	T	T	T	T	T	T
T	F	F	F	F	F	F
T	F	T	T	T	F	F
T	T	F	T	T	F	F
T	T	T	T	T	T	T

More forward proofs

If I prove that $A \Rightarrow B \cap C$ is false, that amounts to selecting the following rows from the table

A	B	C	$B \cup C$	$A \Rightarrow B \cup C$	$B \cap C$	$A \Rightarrow B \cap C$
T	F	F	F	F	F	F
T	F	T	T	T	F	F
T	T	F	T	T	F	F

In those lines, $A \Rightarrow B \cup C$ is true for two of the A, B, C combinations (that's good), but false for the other (that's bad). Therefore, we have not finished yet and we have to prove that either B or C is true, so that we can finally reduce the table to

A	B	C	$B \cup C$	$A \Rightarrow B \cup C$	$B \cap C$	$A \Rightarrow B \cap C$
T	F	T	T	T	F	F
T	T	F	T	T	F	F

in which $A \Rightarrow B \cup C$ is true, and consequently, we have proved that $A \Rightarrow B \cup C$.

Backward proofs ($A \Rightarrow B$)

Example

Show that if $x > 0$, then $x + \frac{1}{x} \geq 2$.

Proof

Let there be the following propositions:

A: $x > 0$

B: $x + \frac{1}{x} \geq 2$

It is obvious that $C_1 \Rightarrow B$, $C_2 \Rightarrow C_1$, $C_3 \Rightarrow C_2$ being

C1: $x + \frac{1}{x} - 2 \geq 0$

C2: $\frac{x^2+1-2x}{x} \geq 0$

C3: $\frac{(x-1)^2}{x} \geq 0$

It is also obvious that $A \Rightarrow C_3$ and, in this way, we have proved that $A \Rightarrow B$. We can simplify the writing of this proof as:

$x + \frac{1}{x} \geq 2 \Leftarrow x + \frac{1}{x} - 2 \geq 0 \Leftarrow \frac{x^2+1-2x}{x} \geq 0 \Leftarrow \frac{(x-1)^2}{x} \geq 0 \Leftarrow x > 0$

Backward proofs ($A \Rightarrow B$)

Example

If $x, y \in \mathbb{R}$, $x, y > 0$, then $\sqrt{xy} \leq \frac{x+y}{2}$

<u>*Proof*</u>

$\sqrt{xy} \leq \frac{x+y}{2} \Leftarrow \sqrt{xy} - \frac{x+y}{2} \leq 0 \Leftarrow \frac{x+y}{2} - \sqrt{xy} \geq 0$

Since x and y are positive numbers, we can write them as $x = a^2$ and $y = b^2$.

Then, $\frac{x+y}{2} - \sqrt{xy} \geq 0 \Leftarrow \frac{a^2+b^2}{2} - ab \geq 0 \Leftarrow a^2 + b^2 - 2ab \geq 0 \Leftarrow (a-b)^2 \geq 0$

This last proposition is always true, therefore $\sqrt{xy} \leq \frac{x+y}{2}$ is also true.

Outline

Modus tollens

The following proofs follow a reasoning model called *Modus tollens* which is formally written as

$$\left(\overline{B} \cap (A \Rightarrow B)\right) \Rightarrow \overline{A}.$$

The intuitive meaning is that if $A \Rightarrow B$ is true and B is false, then A must also be false. Another way of writing this reasoning is

$$(A \Rightarrow B) \Leftrightarrow (\overline{B} \Rightarrow \overline{A}).$$

That is if we want to prove $A \Rightarrow B$, it is enough to prove $\overline{B} \Rightarrow \overline{A}$.

Modus tollens

Example

Show that if x^3 is even, then x is even.

Proof

Let there be the following propositions:

A: x^3 is even

B: x is even

We want to prove that $A \Rightarrow B$. Instead, we'll prove that $\overline{B} \Rightarrow \overline{A}$, with

$\overline{B}$: x is odd

$\overline{A}$: x^3 is odd

But we already proved this in a previous example. Therefore, $A \Rightarrow B$ is true.

Modus tollens

Example

Show that if c is odd, then the equation $n^2 + n - c = 0$ has no integer solution.

Proof

Let there be the following propositions:

A: c is odd

B: $n^2 + n - c = 0$ has no integer solution

We want to prove that $A \Rightarrow B$. Instead, we'll prove that $\overline{B} \Rightarrow \overline{A}$, with

$\overline{B}$: $n^2 + n - c = 0$ has an integer solution

$\overline{A}$: c is even

Proof $\overline{B} \Rightarrow \overline{A}$

Let's assume that $n \in \mathbb{Z}$ is solution of $n^2 + n - c = 0$.
If n is even, then c is even because $c = n^2 + n = (2k)^2 + 2k = 2(2k^2 + k)$.
If n is odd, then c is also even because $c = n^2 + n = (2k+1)^2 + (2k+1) = 2(2k^2 + 3k + 1)$.

Outline

0. Mathematical language
   - Axioms, postulates, definitions and propositions (a)
   - Logical operators (a)
   - Qualifiers (b)
   - Mathematical proofs
     - Modus ponens (b)
     - Modus tollens (c)
     - Reductio ad absurdum (c)
     - Induction (c)
     - Case distinction (c)
     - Counterexample (c)
   - Common math mistakes (c)

Reductio ad absurdum

The following proofs follow a reasoning model called *Reductio ad absurdum* which is formally written as

$$A \Rightarrow B \Leftrightarrow (A \cap \overline{B} \Rightarrow \mathrm{absurdum}).$$

Absurdum is a statement that is always false, like $P \cap \overline{P}$. Let's analyze the truth table for this proposition

Truth table

A	B	$A \Rightarrow B$	$A \cap \overline{B}$	$P \cap \overline{P}$	$A \cap \overline{B} \Rightarrow (P \cap \overline{P})$
F	F	T	F	F	T
F	T	T	F	F	T
T	F	F	T	F	F
T	T	T	F	F	T

We see that the third and sixth columns are identical.

Reductio ad absurdum

Example

Show that $\sqrt{2}$ is irrational.

Proof

It does not appear in the form $A \Rightarrow B$ but it can be put with

A: All facts we know about numbers

B: $\sqrt{2}$ is irrational

Let's assume that $\sqrt{2}$ is rational ($\overline{B}$), that is $\exists p, q \in \mathbb{Z} | \sqrt{2} = \frac{p}{q}$ and p, q are irreducible (they don't have any common factor). If this is true, then $2q^2 = p^2$, i.e., 2 must be a factor of p and consequently p must be $p = 2r$. Substituting this knowledge into $2q^2 = p^2$ we obtain $2q^2 = (2r)^2 \Rightarrow q^2 = 2r^2$. Consequently, 2 is another factor of q. But we presumed that

P: p and q were irreducible

So, if $\sqrt{2}$ is rational, then we have P and $\overline{P}$ at the same time, which is a contradiction, and therefore $\sqrt{2}$ cannot be rational.

Reductio ad absurdum

Example

Show that there are infinite prime numbers.

Proof

Let's presume they is a finite list of prime numbers (in ascending order):

$$2, 3, 5, 7, ..., P$$

Now we construct the number $M = 2 \cdot 3 \cdot 5 \cdot 7 \cdot ... \cdot P + 1$.
If M is prime, then we have a contradiction is M is prime and is larger than P.
If M is not prime, then it has as a factor at least one of the prime numbers in the list. Let's assume it is 3, that is

$$M = 3H = 2 \cdot 3 \cdot 5 \cdot 7 \cdot ... \cdot P + 1 \Rightarrow 1 = 3(H - 2 \cdot 5 \cdot 7 \cdot ... \cdot P)$$

that means that 3 is a factor of 1, which is an absurdum.

Outline

0 Mathematical language

- Axioms, postulates, definitions and propositions (a)
- Logical operators (a)
- Qualifiers (b)
- Mathematical proofs
  - Modus ponens (b)
  - Modus tollens (c)
  - Reductio ad absurdum (c)
  - Induction (c)
  - Case distinction (c)
  - Counterexample (c)
- Common math mistakes (c)

Weak induction

This is a strategy to prove a property of a natural number, $P(n)$. We follow the strategy below:

1. Prove that $P(k)$ is true.
2. Prove that if $P(n-1)$ is true, then $P(n)$ is also true

Example

Show that $S_n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

Proof

1. $S_1 = \sum_{i=1}^{1} i = \frac{1(1+1)}{2} = 1$, which is obviusly true.
2. Let's assume that $S_{n-1} = \sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$. Then, we need to prove that $S_n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. But
   $S_n = S_{n-1} + n = \frac{(n-1)n}{2} + n = n\left(\frac{n-1}{2} + 1\right) = \frac{n(n+1)}{2}$. q.e.d.

Strong induction

The goal is similar to the previous method, but now in the second step we assume that the property is true for all previous integers

1. Prove that $P(k)$ is true.
2. Prove that if $P(k)$ is true and $P(k+1)$ is true and ... $P(n-1)$ is true, then $P(n)$ is also true

Example: Fundamental theorem of arithmetics

Show that for all natural numbers larger than 1 either it is prime or it is the product of prime numbers

Proof

1. The property is true for 2.
2. Let's assume that it is true for $2, 3, 4, ..., n-1$.
   If n is prime, then the property is also true for n.
   If n is not prime, then it can be written as the product of several numbers between 2 and $n-1$. But the property is true for all these numbers, and therefore, the property is also true for n.

Outline

0. Mathematical language
   - Axioms, postulates, definitions and propositions (a)
   - Logical operators (a)
   - Qualifiers (b)
   - Mathematical proofs
     - Modus ponens (b)
     - Modus tollens (c)
     - Reductio ad absurdum (c)
     - Induction (c)
     - Case distinction (c)
     - Counterexample (c)
   - Common math mistakes (c)

Case distinction

For each case we follow a different strategy.

Example: Triangular inequality

Show that $\forall a, b \in \mathbb{R} | |a+b| \leq |a| + |b|$

Proof

We remind that the absolute value is a function defined by parts:

$$|x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$$

Case $a+b \geq 0$: $a+b \leq |a|+|b|$

For all real numbers it is obvious that $x \leq |x|$. Therefore, we have $a \leq |a|$ and $b \leq |b|$. Consequently, $a+b \leq |a|+|b|$.

Case $a+b < 0$: $-(a+b) \leq |a|+|b|$

For all real numbers it is also true that $-x \leq |x|$. Therefore, we have $-a \leq |a|$ and $-b \leq |b|$. Consequently, $-(a+b) = -a-b \leq |a|+|b|$.

Outline

Counterexample

To prove that something is not true, it is enough to show that it is not true for one example. This example is called a counterexample.

Example

Show that $\forall x, y, z \in \mathbb{R}^+$ and $\forall n \in \mathbb{Z}, n \geq 2$ it is verified that $x^n + y^n \neq z^n$

Proof

The proposition is false because, for instance, for $x = 3$, $y = 4$, $z = 5$ and $n = 2$ we have

$$3^2 + 4^2 = 5^2$$

Outline

0 Mathematical language

Common math mistakes

Avoid some common mathematical mistakes (many of them, algebraic):

- Common math mistakes Video 1:
  http://www.youtube.com/watch?v=VHo_sfVdieM
- Common math mistakes PDF:
  http://tutorial.math.lamar.edu/pdf/Common_Math_Errors.pdf
- Common math mistakes Video 2:
  http://www.youtube.com/watch?v=qHSUU_q_2wA
- Common math mistakes Video 3:
  http://www.youtube.com/watch?v=cTiuocJfyCs
- Common math mistakes Video 4:
  http://www.youtube.com/watch?v=r5Yro2GdJ6w

Outline

0. Mathematical language
   - Axioms, postulates, definitions and propositions (a)
   - Logical operators (a)
   - Qualifiers (b)
   - Mathematical proofs
     - Modus ponens (b)
     - Modus tollens (c)
     - Reductio ad absurdum (c)
     - Induction (c)
     - Case distinction (c)
     - Counterexample (c)
   - Common math mistakes (c)

Chapter 1. Vectors

C.O.S. Sorzano

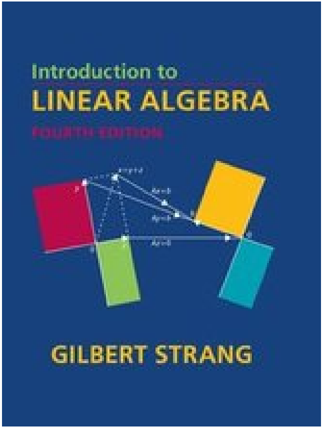
Biomedical Engineering

December 3, 2013

Outline

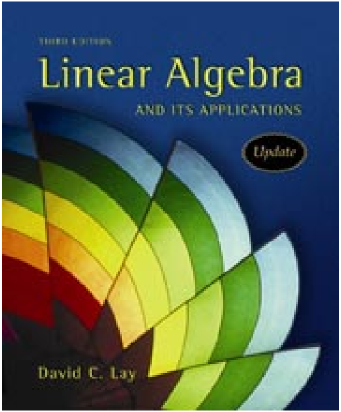
1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

References



G. Strang. Introduction to linear algebra (4th ed). Wellesley Cambridge Press (2009). Chapter 1.

References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 1.

A little bit of history

Vectors were developed during the XIX$^{\text{th}}$ century by mathematicians and physicists like Carl Friedrich Gauss (1799), William Rowan Hamilton (1837), and James Clerk Maxwell (1873), mostly as a tool to represent complex numbers, and later as a tool to perform geometrical reasoning. Their modern algebra was formalized by Josiah Willard Gibbs (1901), a university professor at Yale.

To know more about the history of vectors visit

- http://www.math.mcgill.ca/labute/courses/133f03/VectorHistory.html
- https://www.math.ucdavis.edu/~temple/MAT21D/SUPPLEMENTARY-ARTICLES/Crowe_History-of-Vectors.pdf

Outline

What is a vector?

Definition 1.1

*Informally, a **vector** is a collection of n numbers of the same type. We say it has n components (1,2,...,n)*

We'll see that this definition is terribly simplistic since many other things (like functions, infinite sequences, etc.) can be vectors. But, for the time being, let's stick to this simple definition.

Example

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{Z}^3 \quad \text{is a collection of 3 integer numbers}$$

$$\begin{pmatrix} -1.1 \\ 1.1 \end{pmatrix} \in \mathbb{Q}^2 \quad \text{is a collection of 2 rational numbers}$$

$$\begin{pmatrix} -1.1 \\ \sqrt{2} \end{pmatrix} \in \mathbb{R}^2 \quad \text{is a collection of 2 real numbers}$$

Matlab:

```
[-1.1; sqrt(2)]
```

Transpose

We distinguish between column vectors (for instance **v** below) and row vectors (**w**). In the first case, we say **v** is a $n \times 1$ vector, while in the second, we say **w** is a $1 \times n$ vector.

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix} \text{ and } \mathbf{w} = (w_1 w_2 \dots w_n).$$

Definition 1.2

*The **transpose** is the operation that transforms a column vector into a row vector and viceversa.*

Example

$$(-1\ 1)^T = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Matlab:

```
[-1 1]'
```

Addition of vectors

Definition 1.3

Given two vectors $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix}$ *and* $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix}$ *the **sum** of these two vectors is another vector defined as* $\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \dots \\ v_n + w_n \end{pmatrix}$. *Note that you can only add two column vectors or two row vectors, but not a column and a row vector.*

Example

$$\begin{pmatrix} -1.1 \\ 1.1 \end{pmatrix} + \begin{pmatrix} -1.1 \\ \sqrt{2} \end{pmatrix} = \begin{pmatrix} -2.2 \\ 1.1 + \sqrt{2} \end{pmatrix}$$

Matlab:

```
[-1.1; 1.1]+[-1.1; sqrt(2)]
```

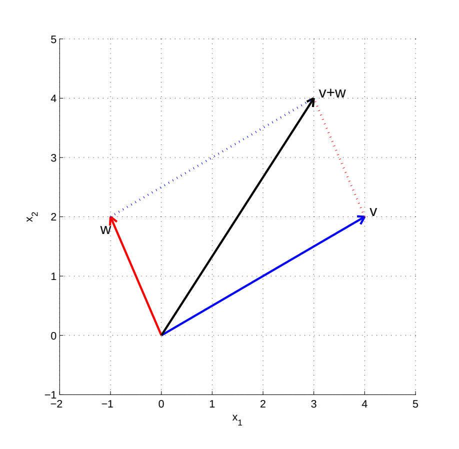
Properties 1.1

Commutativity:
$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$

Addition of vectors

Example

$$\begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$



Product by scalar

Definition 1.4

Given a vector $\mathbf{v}$ *and a scalar* c*, the* ***multiplication*** *of* c *and* $\mathbf{v}$ *is defined as*

$$c\mathbf{v} = \begin{pmatrix} cv_1 \\ cv_2 \\ \dots \\ cv_n \end{pmatrix}$$

Example

$$2\begin{pmatrix} -1.1 \\ 1.1 \end{pmatrix} = \begin{pmatrix} -2.2 \\ 2.2 \end{pmatrix}$$

$$-\begin{pmatrix} -1.1 \\ 1.1 \end{pmatrix} = \begin{pmatrix} 1.1 \\ -1.1 \end{pmatrix}$$

Matlab:

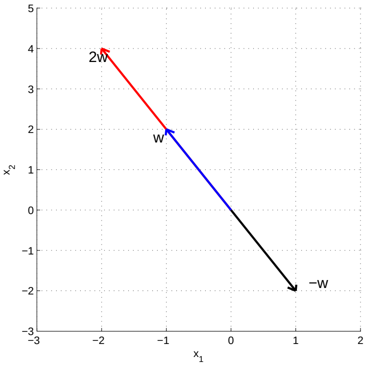
```
2*[-1.1; 1.1] -[1.1; 1.1]
```

Product by scalar

Example

$$\mathbf{w} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

What is the shape of all scaled vectors of the form $c\mathbf{w}$?
If $\mathbf{w} = \mathbf{0}$, then it is a single point ($\mathbf{0}$). If $\mathbf{w} \neq \mathbf{0}$, then it is the straight line that passes through $\mathbf{0}$ and $\mathbf{w}$.



Properties

For simplification we will present them as properties for $\mathbb{R}^n$, but they apply to all vector spaces. Given any three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and any two scalars $c, d \in \mathbb{R}$, we have

Vector operation properties

Regarding the sum of vectors:

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ Commutativity
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ Associativity
3. $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ Existence of neutral element
4. $\mathbf{u} + -\mathbf{u} = -\mathbf{u} + \mathbf{u} = \mathbf{0}$ Existence of symmetric element

Regarding the sum of vectors and scalar product:

5. $\mathrm{c}(\mathbf{u} + \mathbf{v}) = c\mathbf{v} + c\mathbf{u}$ Distributivity with respect to the sum of vectors
6. $(\mathrm{c+d})\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ Distributivity with respect to the sum of scalars

Regarding the scalar product:

7. $\mathrm{c}(\mathrm{d}\mathbf{u}) = (cd)\mathbf{u}$ Associativity
8. $1\mathbf{u} = \mathbf{u}$ Existence of neutral element

Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Linear combination

Definition 2.1

*Given a collection of p scalars (x_i, $i = 1, 2, ..., p$) and p vectors ($\mathbf{v}_i$), the **linear combination** of the p vectors using the **weights** given by the p scalars is defined as*

$$\sum_{i=1}^{p} x_i \mathbf{v}_i = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + ... + x_p \mathbf{v}_p$$

Example

$$\frac{1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -\frac{5}{6} \\ -\frac{11}{6} \end{pmatrix}$$

Matlab:

```
    format rational
-1/2*[-1; 1]-2/3*[2; 2]
```

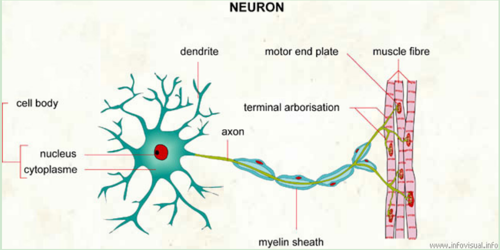
Linear combination

Example

A very basic model of the activity of neurons is

$$output = f(\sum_i weight_i input_i)$$

where $f(x)$ is a non-linear function. In fact, this is the model used in artificial neuron networks.

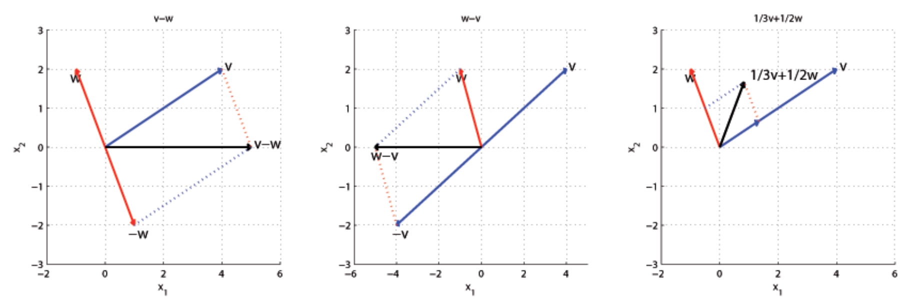


The human brain has in the order of 10^{11} neurons and about 10^{18} connections. See https://www.youtube.com/watch?v=zLp-edwiGUU.

Linear combination

Example

$$\mathbf{v} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \mathbf{w} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$



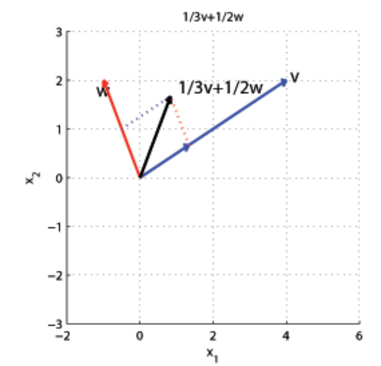
We may think of the weight coefficients as the "travelling" instructions. For instance, for the figure in the right, the instructions say: *"Travel $\frac{1}{3}$ of* $\mathbf{v}$ *along* $\mathbf{v}$*, then travel $\frac{1}{2}$ of* $\mathbf{w}$ *along* $\mathbf{w}$*"*.

What is the shape of all linear combinations of the form $c\mathbf{v} + d\mathbf{w}$

If the two vectors are not collinear (i.e., $\mathbf{w} \neq k\mathbf{v}$), then it is the whole plane passing by $\mathbf{0}$, $\mathbf{v}$ and $\mathbf{w}$. We can think of it as the sum of all vectors belonging to the line $\overline{\mathbf{0v}}$ and $\overline{\mathbf{0w}}$.

The plane generated by $\mathbf{v}$ and $\mathbf{w}$ is the set of all vectors that can be generated as a linear combination of both vectors.

$\Pi = \{\mathbf{r} | \mathbf{r} = c\mathbf{v} + d\mathbf{w} \; \forall c, d \in \mathbf{R}\}$



Linear combination

The previous example prompts the following definition:

Definition 2.2 (Spanned subspace)

The subspace spanned by the vectors $\mathbf{v}_i$, $i = 1, 2, ..., p$, is the set of all vectors that can be expressed as the linear combination of them. Formally,

$$\langle \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p \rangle = \mathrm{Span}\left\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\right\} \triangleq \left\{\mathbf{v} \in \mathbb{R}^n | \mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + ... + x_p\mathbf{v}_p\right\}$$

Example

Assuming all vectors below are linearly independent:
$\mathrm{Span}\left\{\mathbf{v}_1\right\}$ is a straight line.
$\mathrm{Span}\left\{\mathbf{v}_1, \mathbf{v}_2\right\}$ is a plane.
$\mathrm{Span}\left\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{n-1}\right\}$ is a hyperplane.

Properties

$$\mathbf{0} \in \mathrm{Span}\left\{\cdot\right\}$$

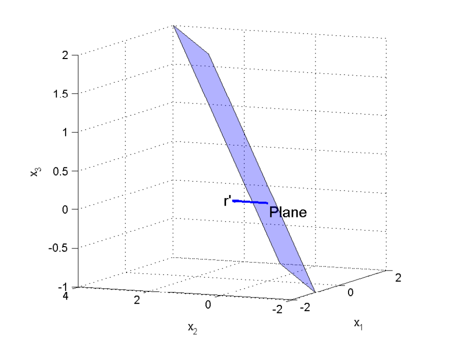
Linear combination

Outside the plane

Let $\mathbf{v} = (1, 1, 0)$ and $\mathbf{w} = (0, 1, 1)$. The linear combinations of $\mathbf{v}$ and $\mathbf{w}$ fill a plane in 3D. All points belonging to this plane are of the form

$$\Pi = \{\mathbf{r} | \mathbf{r} = c(1, 1, 0) + d(0, 1, 1) \, \forall c, d \in \mathbf{R}\} = \{\mathbf{r} = (c, c + d, d) \, \forall c, d \in \mathbf{R}\}$$

It is clear that the vector $\mathbf{r}' = (0, 1, 0) \notin \Pi$, therefore, it is outside the plane.

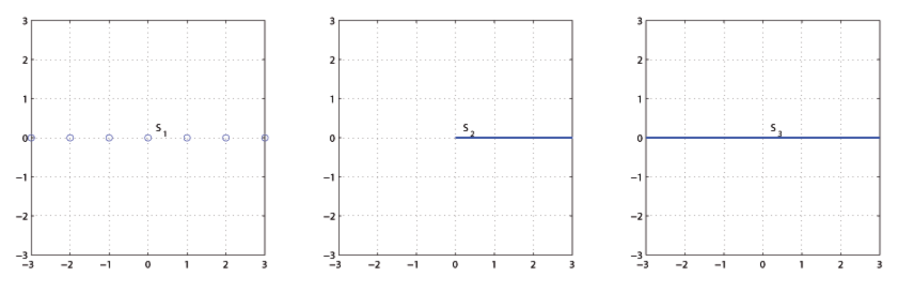


Linear combination

Sets of points

Let $\mathbf{v} = (1, 0)$.

1. $S_1 = \{\mathbf{r} = c\mathbf{v} \ \forall c \in \mathbb{Z}\}$ is a set of points
2. $S_2 = \{\mathbf{r} = c\mathbf{v} \ \forall c \in \mathbb{R}^+\}$ is a semiline
3. $S_3 = \{\mathbf{r} = c\mathbf{v} \ \forall c \in \mathbb{R}\}$ is a line

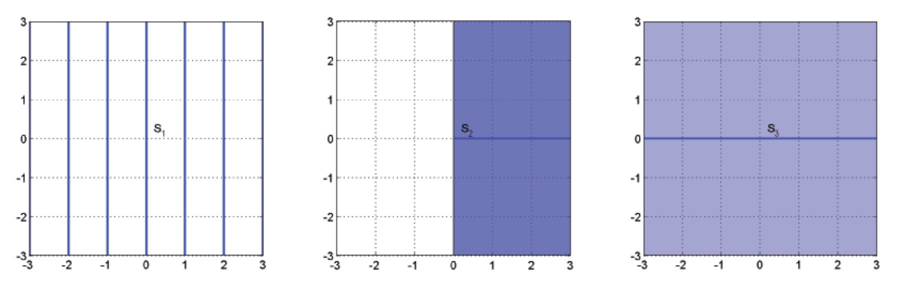


Linear combination

Sets of points

Let $\mathbf{v} = (1, 0)$ and $\mathbf{w} = (0, 1)$.

1. $S_1 = \{\mathbf{r} = c\mathbf{v} + d\mathbf{w}\ \forall c \in \mathbb{Z}, \forall d \in \mathbb{R}\}$ is a set of lines
2. $S_2 = \{\mathbf{r} = c\mathbf{v} + d\mathbf{w}\ \forall c \in \mathbb{R}^+, \forall d \in \mathbb{R}\}$ is a semiplane
3. $S_3 = \{\mathbf{r} = c\mathbf{v} + d\mathbf{w}\ \forall c, d \in \mathbb{R}\}$ is a plane



Combination coefficients

Let $\mathbf{v} = (2, -1)$, $\mathbf{w} = (-1, 2)$ and $\mathbf{b} = (1, 0)$. Find c and d such that $\mathbf{b} = c\mathbf{v} + d\mathbf{w}$.

Solution

We need to find c and d such that

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = c \begin{pmatrix} 2 \\ -1 \end{pmatrix} + d \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2c - d \\ 2d - c \end{pmatrix}$$

This gives a simple equation system

$$\begin{aligned} 2c - d &= 1 \\ 2d - c &= 0 \end{aligned}$$

whose solution is $c = \frac{2}{3}$ and $d = \frac{1}{3}$. We can easily check it with Matlab:

```
2/3*[2 -1]'+1/3*[-1 2]'
```

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 3:

- 1.3.1
- 1.3.3
- 1.3.6
- 1.3.7
- 1.3.25
- 1.3.27
- 1.3.29
- 1.3.31

Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Inner product

Definition 3.1

Given two vectors $\mathbf{v}$ *and* $\mathbf{w}$ *the* ***inner or dot product*** *between* $\mathbf{v}$ *and* $\mathbf{w}$ *is defined as*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} \triangleq \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$$

Mathematically, the concept of inner product is much more general, and this operational definition is just a particularization for vectors in $\mathbb{R}^n$. Although, the introduced inner product is the most common, it is not the only one that can be defined in $\mathbb{R}^n$. But, let's leave these generalization for the moment.

Example

$$\begin{pmatrix} 4 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \end{pmatrix} = 4 \cdot (-1) + 2 \cdot 2 = 0$$

Matlab:

```
dot([4; 2],[-1; 2])
```

Properties 3.1

Commutativity:

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$$

Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Vector norm and vector length

Definition 4.1

Given a vector $\mathbf{v}$, *its* ***length or norm*** *is defined as*

$$\|\mathbf{v}\| \triangleq \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

In the particular case of working with the previously introduced inner product, this definition boils down to

$$\|\mathbf{v}\| \triangleq \sqrt{\mathbf{v}^T \mathbf{v}} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

that is known as the ***Euclidean norm*** *of vector* $\mathbf{v}$.

Properties 4.1

$$\begin{aligned} \|-\mathbf{v}\| &= \|\mathbf{v}\| \\ \|c\mathbf{v}\| &= |c|\|\mathbf{v}\| \end{aligned}$$

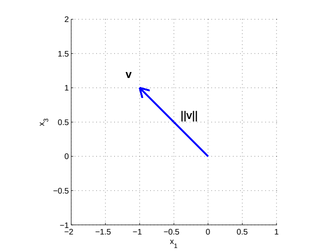
Vector norm and vector length

Example

$$\|(-1,0,1)\| = \sqrt{(-1)^2 + 0^2 + 1^2} = \sqrt{2}$$

Matlab:

```
norm([-1;0;1])
```



Unit vectors

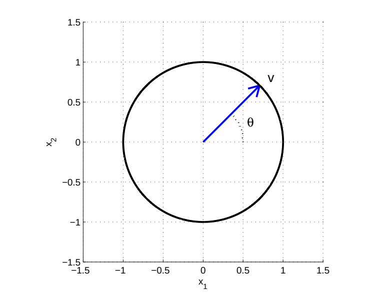
Definition 4.2

$\mathbf{v}$ *is* ***unitary*** *iff* $\|\mathbf{v}\| = 1$.

Example

$\mathbf{e}_1 = (1, 0)$
$\mathbf{e}_2 = (0, 1)$
$\mathbf{e}_\theta = (\cos(\theta), \sin(\theta))$
Matlab:

```
theta=pi/4;
e_theta=[cos(theta);sin(theta)];
norm(e_theta)
```



Unit vectors

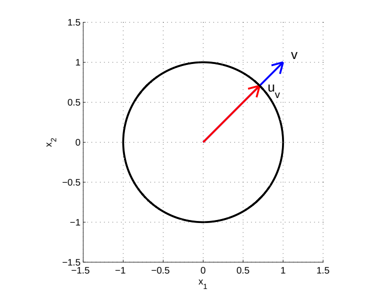
Definition 4.3 (Construction of a unit vector)

Given any vector $\mathbf{v}$ *(whose norm is not null), we can always construct a unitary vector with the same direction of* $\mathbf{v}$ *as* $\mathbf{u_v} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$.

Example

$\mathbf{v} = (1, 1)$

$\mathbf{u_v} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{(1,1)}{\sqrt{2}} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$



Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Distance and angle between two vectors

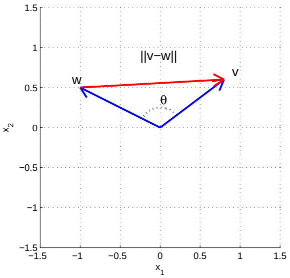
Definition 5.1

Given two vectors $\mathbf{v}$ *and* $\mathbf{w}$*, the* ***distance*** *between both is defined as*

$$d(\mathbf{v},\mathbf{w}) \triangleq \|\mathbf{v}-\mathbf{w}\|$$

and their ***angle*** *is*

$$\angle(\mathbf{v},\mathbf{w}) \triangleq \mathrm{acos}\frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} = \theta$$



Definition 5.2

Two vectors are ***orthogonal*** *(perpendicular) iff their inner product is 0. We then write* $\mathbf{v}\perp\mathbf{w}$*. In this case,* $\angle(\mathbf{v},\mathbf{w}) = \frac{\pi}{2}$*.*

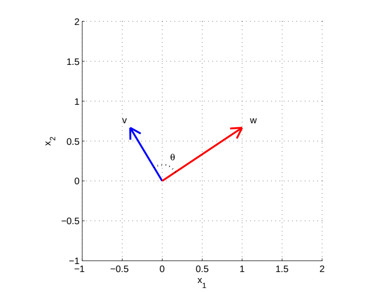
Distance and angle between two vectors

Example

Let $\mathbf{v} = (-\frac{2}{5}, \frac{2}{3})$ and $\mathbf{w} = (1, \frac{2}{3})$. The angle between these two vectors can be calculated as

$$\mathbf{v} \cdot \mathbf{w} = (-\tfrac{2}{5})1 + \tfrac{2}{3}\tfrac{2}{3} = \frac{2}{45}$$
$$\|\mathbf{v}\| = \sqrt{(-\tfrac{2}{5})^2 + (\tfrac{2}{3})^2} = \frac{\sqrt{136}}{15}$$
$$\|\mathbf{w}\| = \sqrt{(1)^2 + (\tfrac{2}{3})^2} = \frac{\sqrt{13}}{3}$$
$$\angle(\mathbf{v}, \mathbf{w}) = \operatorname{acos}\frac{\frac{2}{45}}{\frac{\sqrt{136}}{15}\frac{\sqrt{13}}{3}} = 87.27^\circ$$

$\mathbf{v}$ and $\mathbf{w}$ are almost orthogonal.



Example

Let $\mathbf{v} = (1, 0, 0, 1, 0, 0, 1, 0, 0, 1)$ and $\mathbf{w} = (0, 1, 1, 0, 1, 1, 0, 1, 1, 0)$. These two vectors in a 10-dimensional space are orthogonal because

$$\mathbf{v} \cdot \mathbf{w} = 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 = 0$$

Distance and angle between two vectors

Example

Search for a vector that is orthogonal to $\mathbf{v} = (-\frac{2}{5}, \frac{2}{3})$

Solution

Let the vector $\mathbf{w} = (w_1, w_2)$ be such a vector. Since it is orthogonal to $\mathbf{v}$ it must meet

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0 = (-\tfrac{2}{5})w_1 + \tfrac{2}{3}w_2 \Rightarrow w_2 = \tfrac{3}{5}w_1$$

That is, any vector of the form $\mathbf{w} = (w_1, \frac{3}{5}w_1) = w_1(1, \frac{3}{5})$ is perpendicular to $\mathbf{v}$. This is the line passing by the origin and with direction $(1, \frac{3}{5})$. In particular, for $w_1 = \frac{2}{3}$ we have that $\mathbf{w} = (\frac{2}{3}, \frac{2}{5})$ and for $w_1 = -\frac{2}{3}$ we have $\mathbf{w} = (-\frac{2}{3}, -\frac{2}{5})$.

This is a general rule in 2D. Given a vector $\mathbf{v} = (a, b)$, the vectors $\mathbf{w} = (b, -a)$ and $\mathbf{w} = (-b, a)$ are orthogonal to $\mathbf{v}$.

$$(a, b) \perp (b, -a) \text{ and } (a, b) \perp (-b, a)$$

Distance and angle between two vectors

Theorem 5.1 (Pythagorean theorem)

If $\mathbf{v} \perp \mathbf{w}$*, then* $\|\mathbf{v}-\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$*.*

Proof

$$\|\mathbf{v}-\mathbf{w}\|^2 = (\mathbf{v}-\mathbf{w})^T(\mathbf{v}-\mathbf{w}) = \mathbf{v}^T\mathbf{v} - \mathbf{v}^T\mathbf{w} - \mathbf{w}^T\mathbf{v} + \mathbf{w}^T\mathbf{w} = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\langle\mathbf{v},\mathbf{w}\rangle$$

But, because $\mathbf{v} \perp \mathbf{w}$*, we have* $\langle\mathbf{v},\mathbf{w}\rangle = 0$*, and consequently*

$\|\mathbf{v}-\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$ *(q.e.d.)*

Corollary 5.1

- *If* $\langle\mathbf{v},\mathbf{w}\rangle < 0$*, then* $\frac{\pi}{2} < \theta \leq \pi$*.*
- *If* $\langle\mathbf{v},\mathbf{w}\rangle > 0$*, then* $0 \leq \theta < \frac{\pi}{2}$*.*
- *For two unit vectors,* $\mathbf{u}_1$ *and* $\mathbf{u}_2$*, we have* $\cos\theta = \langle\mathbf{u}_1,\mathbf{u}_2\rangle$*, and as a consequence* $-1 \leq \langle\mathbf{u}_1,\mathbf{u}_2\rangle \leq 1$*.*

Distance and angle between two vectors

Theorem 5.2 (Cosine formula)

For any two vectors, $\mathbf{v}$ *and* $\mathbf{w}$*, such that* $\|\mathbf{v}\| \neq 0$ *and* $\|\mathbf{w}\| \neq 0$*, we have*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$$

<u>Proof</u>

By use of Definition 4.3, we can construct the unit vectors associated to $\mathbf{v}$ *and* $\mathbf{w}$*, that is* $\mathbf{u_v}$ *and* $\mathbf{u_w}$*. Then by Corollary 5.1 we know that*

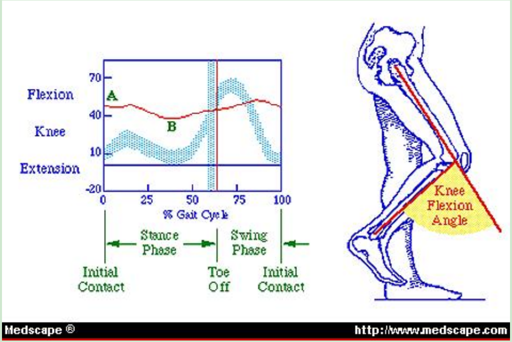
$$\cos\theta = \langle \mathbf{u_v}, \mathbf{u_w} \rangle = \left(\frac{\mathbf{v}}{\|\mathbf{v}\|}\right)^T \left(\frac{\mathbf{w}}{\|\mathbf{w}\|}\right) = \frac{1}{\|\mathbf{u}\|\|\mathbf{w}\|}\mathbf{u}^T\mathbf{w} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{u}\|\|\mathbf{w}\|}$$

From this point it is trivial to deduce that $\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$ *(q.e.d.)*

Distance and angle between two vectors

Example

To compute the knee flexion angle, we need to calculate the dot product between the vectors aligned with the leg before and after the knee.



Distance and angle between two vectors

Theorem 5.3 (Cauchy-Schwarz inequality)

For any two vectors, $\mathbf{v}$ *and* $\mathbf{w}$*, it is verified that*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| < \|\mathbf{v}\| \|\mathbf{w}\|$$

Proof

From the cosine formula (Theorem 5.2), we know that

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta \Rightarrow \\ |\langle \mathbf{v}, \mathbf{w} \rangle| &= |\|\mathbf{v}\| \|\mathbf{w}\| \cos\theta| = \|\mathbf{v}\| \|\mathbf{w}\| \, |\cos\theta| \leq \|\mathbf{v}\| \|\mathbf{w}\| \end{aligned}$$

Example

Let $\mathbf{v} = (-\frac{2}{5}, \frac{2}{3})$ and $\mathbf{w} = (1, \frac{2}{3})$. We already know that $\mathbf{v} \cdot \mathbf{w} = \frac{2}{45}$, $\|\mathbf{v}\| = \frac{\sqrt{136}}{15}$, and $\|\mathbf{w}\| = \frac{\sqrt{13}}{3}$. Let us check Cauchy-Schwarz inequality

$$|\tfrac{2}{45}| < \frac{\sqrt{136}}{15} \frac{\sqrt{13}}{3} \Leftrightarrow 0.0444 < 0.9344$$

Distance and angle between two vectors

Example

Show that for any two positive numbers, x and y, the geometric mean ($\sqrt{xy}$) is always smaller or equal than the arithmetic mean ($\frac{x+y}{2}$). For instance, the statement is verified for $x = 2$ and $y = 3$: $\sqrt{6} \leq \frac{5}{2} \Leftrightarrow 2.4495 \leq 2.5$.

Proof

Let there be vectors $\mathbf{v} = (a, b)$ and $\mathbf{w} = (b, a)$. Then, by Cauchy-Schwarz inequality we know that

$$|\langle \mathbf{v}, \mathbf{w} \rangle| < \|\mathbf{v}\| \|\mathbf{w}\| \Rightarrow |2ab| \leq a^2 + b^2$$

Since x and y are positive numbers, we may consider them to be $x = a^2$ and $y = b^2$. Consequently, we can rewrite the previous expression as

$$2\sqrt{x}\sqrt{y} \leq x + y \Rightarrow \sqrt{xy} \leq \frac{x+y}{2} \text{ (q.e.d.)}$$

In fact, the geometric mean is nothing more than the arithmetic mean in logarithmic units

$$\log(\sqrt{xy}) = \log(xy)^{\frac{1}{2}} = \frac{1}{2}(\log x + \log y) = \frac{\log x + \log y}{2}$$

Theorem 5.4 (Triangular inequality)

For any two vectors, $\mathbf{v}$ *and* $\mathbf{w}$*, it is verified that*

$$\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$

Proof

By definition we know that

$$\|\mathbf{v}+\mathbf{w}\|^2 = (\mathbf{v}+\mathbf{w})^T(\mathbf{v}+\mathbf{w}) = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\langle\mathbf{v},\mathbf{w}\rangle$$

Applying the Cauchy-Schwarz inequality (Theorem 5.3), we have

$$\|\mathbf{v}+\mathbf{w}\|^2 \leq \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\|\mathbf{v}\|\|\mathbf{w}\| = (\|\mathbf{v}\| + \|\mathbf{w}\|)^2$$

Taking the square root we have

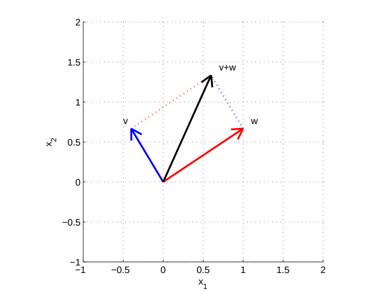
$$\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$

Distance and angle between two vectors

Example

Let $\mathbf{v} = (-\frac{2}{5}, \frac{2}{3})$ and $\mathbf{w} = (1, \frac{2}{3})$. We already know that $\|\mathbf{v}\| = \frac{\sqrt{136}}{15}$ and $\|\mathbf{w}\| = \frac{\sqrt{13}}{3}$. Let us check the triangular inequality

$$\mathbf{v}+\mathbf{w} = (\tfrac{3}{5}, \tfrac{4}{3}) \Rightarrow \|\mathbf{v}+\mathbf{w}\| = \frac{\sqrt{481}}{15}$$
$$\frac{\sqrt{481}}{15} \leq \frac{\sqrt{136}}{15} + \frac{\sqrt{13}}{3} \Leftrightarrow 1.4621 \leq 1.9793$$



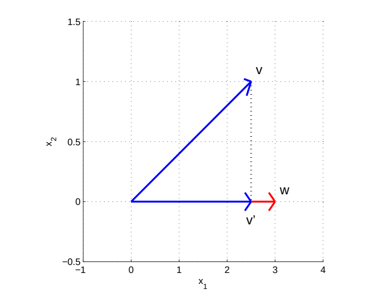
Distance and angle between two vectors

Orthogonal projections

Let us consider the orthogonal projection of **v** onto **w**.

$$\mathbf{v}' = \langle \mathbf{v}, \mathbf{w} \rangle \frac{\mathbf{w}}{\|\mathbf{w}\|^2} = \frac{\langle \mathbf{v},\mathbf{w} \rangle}{\|\mathbf{w}\|} \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

The length of this vector is $\frac{\langle \mathbf{v},\mathbf{w} \rangle}{\|\mathbf{w}\|}$



Example

Let $\mathbf{v} = (\frac{5}{2}, 1)$ and $\mathbf{w} = (3, 0)$. Then, $\mathbf{v}' = \frac{\frac{5}{2}3+1\cdot 0}{3}(1,0) = (\frac{5}{2}, 0)$. See the figure above.

Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Multiplication by matrices

Example

Let's consider three vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Let's consider the linear combination

$$\mathbf{y} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = x_1 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{pmatrix}$$

I can obtain the same result by constructing a matrix

$$A = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

And making the multiplication

$$\mathbf{y} = A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{pmatrix}$$

Multiplication by matrices

Example

We can also achieve the same result by calculating **y** as the inner product of the rows of the matrix A and the weight vector.

$$\mathbf{y} = \begin{pmatrix} \left\langle (1,0,0), (x_1, x_2, x_3) \right\rangle \\ \left\langle (-1,1,0), (x_1, x_2, x_3) \right\rangle \\ \left\langle (0,-1,1), (x_1, x_2, x_3) \right\rangle \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{pmatrix}$$

Matlab:

```
 syms x1 x2 x3
x=[x1; x2; x3]
A=[1 0 0; -1 1 0; 0 -1 1];
y=A*x
```

Multiplication by matrices

Matrix multiplication as a linear combination

This is a general rule: a matrix multiplication can be seen as the linear combination of the columns of the matrix.

$$A = (\mathbf{c}_1 \ \mathbf{c}_2 \ ...\mathbf{c}_p) \Rightarrow \mathbf{y} = A\mathbf{x} = \sum_{i=1}^{p} x_i \mathbf{c}_i$$

Matrix multiplication as inner products

Also, a matrix multiplication can be seen as the dot product of the weight vector with the rows of the matrix.

$$A = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ ... \\ \mathbf{r}_n^T \end{pmatrix} \Rightarrow \mathbf{y} = A\mathbf{x} = \begin{pmatrix} \langle \mathbf{r}_1, \mathbf{x} \rangle \\ \langle \mathbf{r}_2, \mathbf{x} \rangle \\ ... \\ \langle \mathbf{r}_n, \mathbf{x} \rangle \end{pmatrix}$$

Multiplication by matrices

Properties of multiplication by matrices

$$A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$$
$$A(c\mathbf{u}) = c(A\mathbf{u})$$

Outline

1. Vectors
   - Vectors and basic operations (a)
   - Linear combination (a)
   - Inner product or dot product (b)
   - Norm, vector length and unit vectors (b)
   - Distances and angles (b)
   - Multiplication by matrices (b)

Chapter 2. Linear equation systems

C.O.S. Sorzano

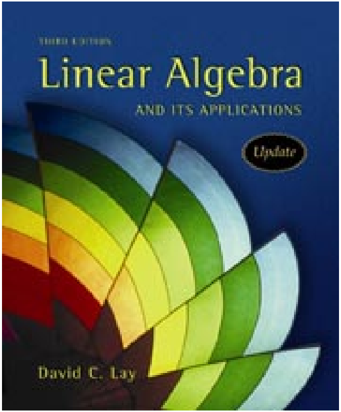
Biomedical Engineering

September 24, 2013

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

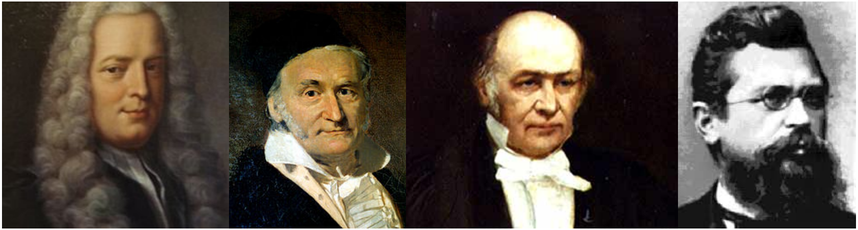
References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 1.

A little bit of history

Linear equations in their modern form are known since the middle of the XVIIIth century and they were strongly developed during the XIXth century with important contributions of people like Gabriel Cramer (1750), Carl Friedrich Gauss (1801), Sir William Rowan Hamilton (1843) and Wilhelm Jordan (1873). They were mostly developed to explain the mechanics of celestial objects.

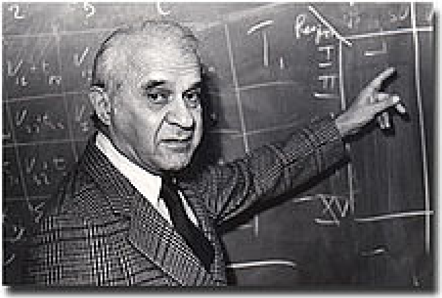


To know more about the history of linear equations visit

- http://hom.wikidot.com/cramer-s-method-and-cramer-s-paradox

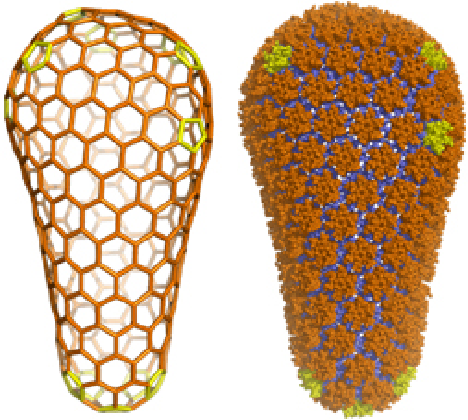
A little bit of history

Wassily Leontief was a Russian-American economist that worked in Harvard. In 1949 he performed an analysis with the early computers at Harvard using data from the U.S. Bureau of Labor Statistics to classify the U.S. economy into 500 sectors, that were later simplified to 42. He used linear equation systems to do so. It took 56 hours in Mark II (one of the first computers) to solve it. He was awarded the Nobel prize in 1970 for his work on input-output tables that analyze how outputs from some industries are inputs to some other industries.



A little bit of history

Currently, we need about two weeks in a supercomputer (128 cores) to solve the structure of a macromolecular assembly (in the figure, the HIV virus capsid). We have 1,000 million equations with about 3 million unknowns.



Outline

2. Linear equation system
    - Introduction (a)
    - Gauss-Jordan algorithm (b)
    - Interpretation as a subspace (b)
    - Existence and uniqueness of solutions (c)
    - Applications (c)
    - Linear independence (c)
    - Linear transformations (d)
    - Geometrical transformations (e)
    - Classification of functions (e)
    - More applications (e)

What is a linear equation system?

Definition 1.1 (Linear equation system)

*A **linear equation** is one that can be expressed in the form*

$$\sum_{i=1}^{n} a_i x_i = b$$
$$a_1 x_1 + a_2 x_2 + ... + a_n x_n = b$$
$$\langle \mathbf{a}, \mathbf{x} \rangle = b$$

*The unknowns are x_i ($i = 1, 2, ..., n$) while a_i's and b are coefficients. When we have several of these equations, we have a **linear equation system**.*

Example

Examples of linear equations

$$7x_1 - 2x_2 = 4$$
$$7(x_1 - \sqrt{3}x_2) = \frac{1}{\sqrt{2}}x_1 \Rightarrow$$
$$(7 - \frac{1}{\sqrt{2}})x_1 - 7\sqrt{3}x_2 = 0$$

Examples of non-linear equations

$$x_1 + x_2 + x_1 x_2 = 1$$
$$\sqrt{x_1} + x_2 = 1$$

Set of solutions of a linear equation

Definition 1.2 (Set of solutions of a linear equation system)

The set of solutions of a linear equation system $S \subseteq \mathbb{R}^n$ is the set of all those values that we can assign to x_1, x_2, ..., x_n such that the equation system is fulfilled.

Example

Consider the following equation system

$$\begin{aligned} 2x_1 - x_2 &= 7 \\ x_1 + 2x_2 &= 11 \end{aligned}$$

$\mathbf{x} = (5, 3)$ is a solution to this equation system because

$$\begin{aligned} 2 \cdot 5 - 3 &= 7 \\ 5 + 2 \cdot 3 &= 11 \end{aligned}$$

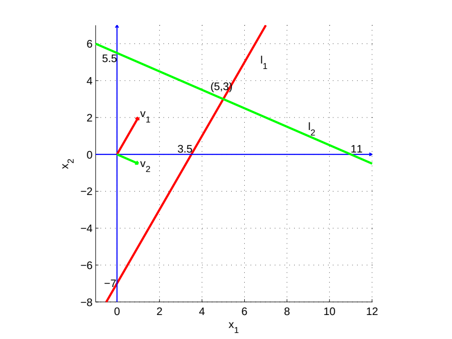
In fact it is its unique solution and, therefore, $S = \{(5, 3)\} \subset \mathbb{R}^2$.

Geometric interpretation

Example

$$l_1\colon\ 2x_1 - x_2 = 7 \Rightarrow x_2 = 2x_1 - 7 \Rightarrow \mathbf{v}_1 = (1,2)$$
$$l_2\colon\ x_1 + 2x_2 = 11 \Rightarrow x_2 = 11 - \tfrac{1}{2}x_1 \Rightarrow \mathbf{v}_2 = (1,-\tfrac{1}{2})$$

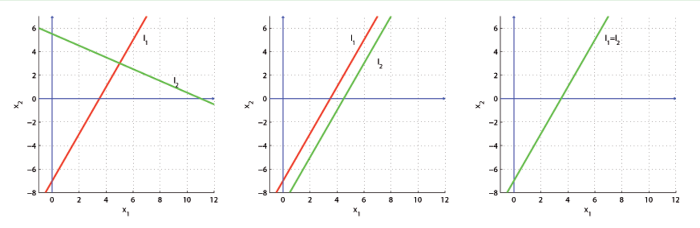
Each one of the equations is actually representing a line, and both lines, in this case intersect at the point $(5,3)$, the unique solution of this equation system.



Geometric interpretation

Example

There can be a single solution (left), no solution (middle), or infinite ($l_1 = l_2$; right)



In general

With linear equations we can represent:

a **line** in 2D: $a_1x_1 + a_2x_2 = b$

a **plane** in 3D: $a_1x_1 + a_2x_2 + a_3x_3 = b$

a **hyperplane** in nD: $a_1x_1 + a_2x_2 + ... + a_nx_n = b$

Matrix notation

Example

The equation system

$$\begin{array}{rrrcr} x_1 & -2x_2 & +x_3 & = & 0 \\ & 2x_2 & -8x_3 & = & 8 \\ -4x_1 & +5x_2 & +9x_3 & = & -9 \end{array}$$

can be represented as

$$\left(\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right) \quad [\tilde{A}]$$

or

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 8 \\ -9 \end{pmatrix} \quad [A\mathbf{x} = \mathbf{b}]$$

Matrix notation

In general

$A \in \mathcal{M}_{m\times n}$ is called the **system matrix** of an equation system with m equations and n unknowns.
$\tilde{A} \in \mathcal{M}_{m\times(n+1)}$ is called the **augmented system matrix** of an equation system with m equations and n unknowns.

Basic row iterations

To solve the equation system with the augmented system matrix, we used the so-called basic row operations:

Substitution: $\mathbf{r}_i \leftarrow k_i\mathbf{r}_i + k_j\mathbf{r}_j$: Row i is substituted by a linear combination of rows i and j

Swapping: $\mathbf{r}_i \leftrightarrow \mathbf{r}_j$: Row i swapped with row j

Scaling: $\mathbf{r}_i \leftarrow k_i\mathbf{r}_i$: Row i is multiplied by a scale factor

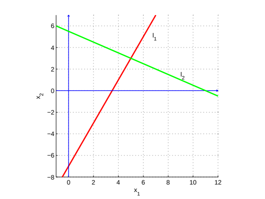
All these operations transform the equation system into an **equivalent** system (with the same set of solutions). The two matrices (original and transformed) are said to be **row equivalent**.

Solving the equation system

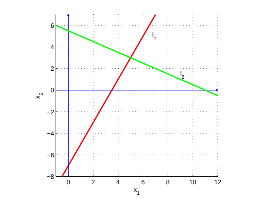
Example

In the following example we will see how linear combinations are actually changing the equation system to a different one, while scaling is not.

$$\begin{array}{rrcr} 2x_1 & -x_2 & = & 7 \\ x_1 & +2x_2 & = & 11 \end{array} \qquad \left(\begin{array}{rr|r} 2 & -1 & 7 \\ 1 & 2 & 11 \end{array}\right)$$



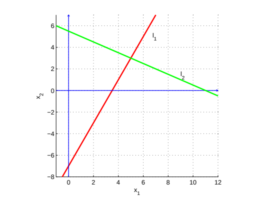
$$\mathbf{r}_1 \leftarrow \tfrac{1}{2}\mathbf{r}_1 \qquad \left(\begin{array}{rr|r} 1 & -\frac{1}{2} & \frac{7}{2} \\ 1 & 2 & 11 \end{array}\right)$$



Solving the equation system

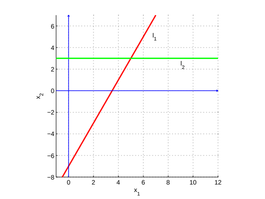
Example

$$\left(\begin{array}{cc|c} 1 & -\frac{1}{2} & \frac{7}{2} \\ 1 & 2 & 11 \end{array}\right)$$



$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - \mathbf{r}_1$

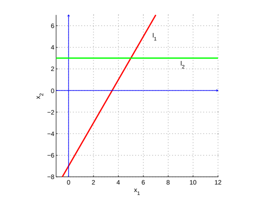
$$\left(\begin{array}{cc|c} 1 & -\frac{1}{2} & \frac{7}{2} \\ 0 & \frac{5}{2} & \frac{15}{2} \end{array}\right)$$



Solving the equation system

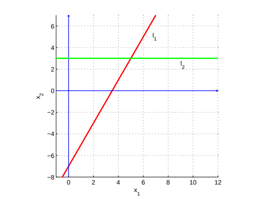
Example

$$\left(\begin{array}{cc|c} 1 & -\frac{1}{2} & \frac{7}{2} \\ 0 & \frac{5}{2} & \frac{15}{2} \end{array}\right)$$



$\mathbf{r}_2 \leftarrow \frac{2}{5}\mathbf{r}_2$

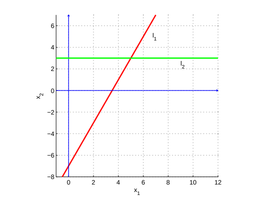
$$\left(\begin{array}{cc|c} 1 & -\frac{1}{2} & \frac{7}{2} \\ 0 & 1 & 3 \end{array}\right)$$



Solving the equation system

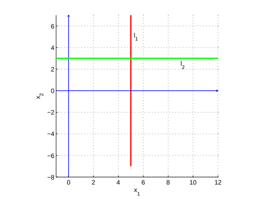
Example

$$\left(\begin{array}{cc|c} 1 & -\frac{1}{2} & \frac{7}{2} \\ 0 & 1 & 3 \end{array}\right)$$



$\mathbf{r}_1 \leftarrow \mathbf{r}_1 + \frac{1}{2}\mathbf{r}_2$

$$\left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 3 \end{array}\right)$$



Existence and uniqueness of solutions

Example

$$\begin{array}{rrrcr} x_1 & -2x_2 & +x_3 & = & 0 \\ & 2x_2 & -8x_3 & = & 8 \\ -4x_1 & +5x_2 & +9x_3 & = & -9 \end{array} \sim \ldots \sim \left(\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

I can solve for x_3 ($x_3 = 3$), then use this value in the second equation to solve for x_2, and finally use these two values in the first equation to solve for x_1. Thus, the equation system **has a solution and it is unique**. We say the equation system is **compatible**. The set of solutions is $S = \{(29, 16, 3)\}$.

Matlab:

```
 A=[1 -2 1; 0 2 -8; -4 5 9];
b=[0; 8; -9];
x=A\b
```

Existence and uniqueness of solutions

Example

$$\begin{array}{rrrcl} & x_2 & -4x_3 & = & 8 \\ 2x_1 & -3x_2 & +2x_3 & = & 1 \\ 5x_1 & -8x_2 & +7x_3 & = & 1 \end{array} \sim \ldots \sim \left(\begin{array}{rrr|c} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{array}\right)$$

Last equation implies $0 = \frac{5}{2}$ which is impossible. Consequently, there is **no solution** and we say that the equation system is **incompatible**. The set of solutions is $S = \varnothing$.

Example

$$\begin{array}{rrcl} x_1 & +x_2 & = & 1 \\ 2x_1 & +2x_2 & = & 2 \end{array} \sim \ldots \sim \left(\begin{array}{rr|c} 1 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right)$$

There are **infinite** solutions. The system is **compatible indeterminate**. The set of solutions is $S = \{(x_1, 1 - x_1)\}$.

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 1:

- 1.1.11
- 1.1.4
- 1.1.15
- 1.1.18
- 1.1.25
- 1.1.26
- 1.1.33

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Echelon matrices

Example

The following matrices are echelon matrices:

$$A_1 = \begin{pmatrix} \Diamond & \heartsuit & \heartsuit & \heartsuit \\ 0 & \Diamond & \heartsuit & \heartsuit \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 0 & \Diamond & \heartsuit & \heartsuit \\ 0 & 0 & 0 & \Diamond \end{pmatrix}$$

In the previous matrices we have marked with $\Diamond$ the **leading elements** (the first ones different from 0 in their row), and with $\heartsuit$ the rest of the elements different from 0.

Echelon matrices

Definition 2.1 (Echelon matrix)

*A rectangular matrix has an **echelon** form iif:*

1. *Within each row, the first element different from zero (called the **leading entry**) is in a column to the right of the leading entry of the previous row.*
2. *Within each column, all values below a leading entry are zero.*
3. *All rows without a leading entry (i.e., they only have zeros) are below all the rows in which at least one element is not zero.*

Definition 2.2 (Reduced echelon matrix)

*A rectangular matrix has a **reduced echelon** form iif:*

1. *It is echelon.*
2. *The leading entry of each row is 1.*
3. *The leading entry is the only 1 in its column.*

Echelon matrices

Theorem 2.1

Each matrix is row equivalent to one and only one reduced echelon matrix.

Example

$$
\begin{array}{l|l}
 & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & -1 & 0 \end{pmatrix} \\
\begin{array}{l} \mathbf{r}_2 \leftarrow \mathbf{r}_2 - 4\mathbf{r}_1 \\ \mathbf{r}_3 \leftarrow \mathbf{r}_3 + \mathbf{r}_1 \end{array} & \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 1 & 3 \end{pmatrix} \\
\mathbf{r}_2 \leftrightarrow \mathbf{r}_3 & \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & -3 & -6 \end{pmatrix} \\
\mathbf{r}_3 \leftrightarrow \mathbf{r}_3 + 3\mathbf{r}_2 & \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix}
\end{array}
\qquad
\begin{array}{l|l}
\begin{array}{l} \mathbf{r}_1 \leftarrow \mathbf{r}_1 - 2\mathbf{r}_2 \\ \mathbf{r}_3 \leftarrow \frac{1}{3}\mathbf{r}_3 \end{array} & \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \\
\begin{array}{l} \mathbf{r}_1 \leftarrow \mathbf{r}_1 + 3\mathbf{r}_3 \\ \mathbf{r}_2 \leftarrow \mathbf{r}_2 - 3\mathbf{r}_3 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{array}
$$

Echelon matrices

Example (continued)

Matlab:

```
 A=[1 2 3; 4 5 6; -1 -1 0]
A(2,:)=A(2,:)-4*A(1,:)
A(3,:)=A(3,:)+A(1,:)
aux=A(2,:); A(2,:)=A(3,:); A(3,:)=aux
A(3,:)=A(3,:)+3*A(2,:)
A(1,:)=A(1,:)-2*A(2,:)
A(3,:)=1/3*A(3,:)
A(1,:)=A(1,:)+3*A(3,:)
A(2,:)=A(2,:)-3*A(3,:)
```

Echelon matrices

Example

Now, we'll repeat the same example using different row operations:

$$
\begin{array}{l|l}
 & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & -1 & 0 \end{pmatrix} \\
\mathbf{r}_1 \leftarrow \mathbf{r}_3 & \begin{pmatrix} -1 & -1 & 0 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix} \\
\mathbf{r}_1 \leftarrow -\mathbf{r}_1 & \begin{pmatrix} 1 & 1 & 0 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix} \\
\begin{array}{l} \mathbf{r}_2 \leftarrow \mathbf{r}_2 - 4\mathbf{r}_1 \\ \mathbf{r}_3 \leftarrow \mathbf{r}_3 - \mathbf{r}_1 \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 6 \\ 0 & 1 & 3 \end{pmatrix}
\end{array}
\qquad
\begin{array}{l|l}
\begin{array}{l} \mathbf{r}_1 \leftarrow \mathbf{r}_1 - \mathbf{r}_2 \\ \mathbf{r}_3 \leftarrow \mathbf{r}_3 - \mathbf{r}_2 \end{array} & \begin{pmatrix} 1 & 0 & -6 \\ 0 & 1 & 6 \\ 0 & 0 & -3 \end{pmatrix} \\
\mathbf{r}_3 \leftarrow -\frac{1}{3}\mathbf{r}_3 & \begin{pmatrix} 1 & 0 & -6 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{pmatrix} \\
\begin{array}{l} \mathbf{r}_1 \leftarrow \mathbf{r}_1 + 6\mathbf{r}_3 \\ \mathbf{r}_2 \leftarrow \mathbf{r}_2 - 6\mathbf{r}_3 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{array}
$$

Gauss-Jordan algorithm

Definition 2.3 (Pivot and pivot column)

A pivot element is the element of a matrix that is used to perform certain calculations. For the Gauss-Jordan algorithm it corresponds to the first element different from zero in a given row. A pivot column is a column that contains a pivot.

Step 1

Choose the left-most pivot column. The pivot element (marked in red) is any value within this column different from 0. Note: Normally, we should take the one with maximum absolute value to avoid numerical errors.

Example

$$\begin{pmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ \color{red}{3} & -9 & 12 & -9 & 6 & 15 \end{pmatrix}$$

Gauss-Jordan algorithm

Step 2

Sort rows if necessary so that the pivot is as high as possible.

Example

$$\mathbf{r}_3 \leftrightarrow \mathbf{r}_1 \quad \left|\quad \begin{pmatrix} \color{red}{3} & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{pmatrix}\right.$$

Step 3

Use row operations to force the elements below the pivot to be 0.

Example

$$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - \mathbf{r}_1 \quad \left|\quad \begin{pmatrix} \color{red}{3} & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{pmatrix}\right.$$

Gauss-Jordan algorithm

Step 4

Repeat Steps 1 to 3 with the rows below the pivot.

Example

$$
\begin{array}{r|l}
 & \begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & \color{red}{3} & -6 & 6 & 4 & -5 \\ 0 & 2 & -4 & 4 & 2 & -6 \end{pmatrix} \\
\mathbf{r}_3 \leftarrow \mathbf{r}_3 - \frac{2}{3}\mathbf{r}_2 & \begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & \color{red}{3} & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & -\frac{2}{3} & -\frac{8}{3} \end{pmatrix} \\
 & \begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & \color{red}{-\frac{2}{3}} & -\frac{8}{3} \end{pmatrix}
\end{array}
$$

Gauss-Jordan algorithm

Step 5

Starting from the lowest and right-most pivot, force the elements above that pivot to be zero. If the pivot is not 1, then rescale the row. Repeat with the next pivot on the left.

Example

$$
\begin{array}{l|l}
\mathbf{r}_3 \leftarrow -\frac{3}{2}\mathbf{r}_3 & \begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & \color{red}{1} & 4 \end{pmatrix} \\
\begin{array}{l} \mathbf{r}_2 \leftarrow \mathbf{r}_2 - 4\mathbf{r}_3 \\ \mathbf{r}_1 \leftarrow \mathbf{r}_1 - 6\mathbf{r}_3 \end{array} & \begin{pmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 3 & -6 & 6 & 0 & -21 \\ 0 & 0 & 0 & 0 & \color{red}{1} & 4 \end{pmatrix} \\
\mathbf{r}_2 \leftarrow \frac{1}{3}\mathbf{r}_2 & \begin{pmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & \color{red}{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \\
\mathbf{r}_1 \leftarrow \mathbf{r}_1 + 9\mathbf{r}_2 & \begin{pmatrix} 3 & 0 & -6 & 9 & 0 & -72 \\ 0 & \color{red}{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \\
\mathbf{r}_1 \leftarrow \frac{1}{3}\mathbf{r}_1 & \begin{pmatrix} \color{red}{1} & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}
\end{array}
$$

Computing the inverse of a $n \times n$ matrix costs in the order of n^3 operations ($O(n^3)$). However, calculating the reduced echelon form is only in the order of n^2 ($O(n^2)$). This difference is more and more important as n grows.

Existence and uniqueness of solutions (revisited)

We can now review the issue of existence and uniqueness under the light of the reduced echelon matrix.

Example

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

The system is compatible and the set of solutions is formed by a single point $S = \{(1, 4, 0)\}$.

Example

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

There are **infinite** solutions. The system is **compatible indeterminate**. The set of solutions is $S = \left\{(1, 4 - x_3, x_3)\ \forall x_3 \in \mathbb{R}^3\right\}$. Because the set of solutions depends on a single variable, the set of solutions is a line.

Existence and uniqueness of solutions (revisited)

Example

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

There are **infinite** solutions. The system is **compatible indeterminate**. The set of solutions is $S = \left\{(1, 4 - x_3 - x_4, x_3, x_4)\ \forall x_3, x_4 \in \mathbb{R}^3\right\}$. Now, the set of solutions depends on 2 variables and, consequently, it is a plane.

Example

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

The system is incompatible since the last equation is $0 = 1$. The set of solutions is the empty set, $S = \varnothing$.

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 2:

- 1.2.2
- 1.2.8
- 1.2.19
- 1.2.33
- 1.2.34

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Interpretation as a subspace

Subspace spanned by columns

Consider the equation system given by the matricial equation $A\mathbf{x} = \mathbf{b}$, where $A \in \mathcal{M}_{n\times p}$. Let us call the p columns of A as $\mathbf{c}_i \in \mathbb{R}^n$. The previous equation can be rewritten as

$$(\mathbf{c}_1\ \mathbf{c}_2\ \dots\ \mathbf{c}_p)\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{pmatrix} = \mathbf{b} \Rightarrow \sum_{i=1}^{p} x_i\mathbf{c}_i = \mathbf{b}$$

That is, $A\mathbf{x}$ is the subspace spanned by the columns of matrix A.

$$\mathrm{Span}\left\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p\right\} = \left\{\mathbf{v} \in \mathbb{R}^n | \mathbf{v} = A\mathbf{x}\ \forall \mathbf{x} \in \mathbb{R}^p\right\}$$

The equation system $A\mathbf{x} = \mathbf{b}$ the poses the question: **Find the weight coefficients x_i such that vector $\mathbf{b}$ belongs to $\mathrm{Span}\left\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p\right\}$.**

Interpretation as a subspace

Example

The equation system

$$\begin{array}{rrrcl} x_1 & +2x_2 & -x_3 & = & 4 \\ & -5x_2 & +3x_3 & = & 1 \end{array}$$

can be represented as

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

That is, which are the weight coefficients x_1, x_2 and x_3 such that the vector $(4, 1)$ belongs to the subspace generated by the vectors $(1, 0)$, $(2, -5)$, and $(-1, 3)$.

Interpretation as a subspace

Theorem 3.1

The matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution as the vector equation $\sum_{i=1}^{p} x_i \mathbf{c}_i = \mathbf{b}$ and as the equation system whose augmented matrix is $\tilde{A} = (A|\mathbf{b})$.

Theorem 3.2

For any $A \in \mathcal{M}_{n\times p}$ and vector $\mathbf{b} \in \mathbb{R}^n$, the following four statements are equivalent, that is, $P_1 \Leftrightarrow P_2 \Leftrightarrow P_3 \Leftrightarrow P_4$

- P_1: *The equation $A\mathbf{x} = \mathbf{b}$ has a solution.*
- P_2: *$\mathbf{b}$ is a linear combination of the columns of A.*
- P_3: *The columns of A span all $\mathbb{R}^n$, i.e., $\mathrm{Span}\{\mathbf{c}_i\} = \mathbb{R}^n$.*
- P_4: *A has a pivot in each row.*

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 4:

- 1.4.13
- 1.4.18
- 1.4.26
- 1.4.27
- 1.4.32
- 1.4.39
- 1.4.41 (bring computer)

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Existence and uniqueness of solutions (revisited once again)

Let us consider the **homogeneous system** $A\mathbf{x} = \mathbf{0}$. It obviously has the trivial solution $\mathbf{x} = \mathbf{0}$. Non-trivial solutions can be found through the echelon matrix

Example

$$\begin{array}{rrrcl} 3x_1 & +5x_2 & -4x_3 & = & 0 \\ -3x_1 & -2x_2 & +4x_3 & = & 0 \\ 6x_1 & +x_2 & -8x_3 & = & 0 \end{array} \Rightarrow \ldots \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This is a compatible indeterminate system whose set of solutions is $S = \left\{(-\frac{4}{3}x_3, 0, x_3)\ \forall x_3 \in \mathbb{R}\right\}$, or what is the same

$$S = \mathrm{Span}\left\{(-\tfrac{4}{3}, 0, 1)\right\}.$$

That is, any of the infinite points in the straight line whose director vector is $(-\frac{4}{3}, 0, 1)$ is a solution of the equation system.

Existence and uniqueness of solutions (revisited once again)

Let us consider the **non-homogeneous system** $A\mathbf{x} = \mathbf{b}$.

Example

$$\begin{array}{rrrcr} 3x_1 & +5x_2 & -4x_3 & = & 7 \\ -3x_1 & -2x_2 & +4x_3 & = & -1 \\ 6x_1 & +x_2 & -8x_3 & = & -4 \end{array} \Rightarrow \ldots \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{4}{3} & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

This is a compatible indeterminate system whose set of solutions is $S = \left\{(-\frac{4}{3}x_3, 2, x_3)\ \forall x_3 \in \mathbb{R}\right\}$, or what is the same

$$S = \left\{(0,2,0) + (-\tfrac{4}{3}x_3, 0, x_3)\ \forall x_3 \in \mathbb{R}\right\} = (0,2,0) + \mathrm{Span}\left\{(-\tfrac{4}{3}, 0, 1)\right\}.$$

That is, any of the infinite points in the straight line whose director vector is $(-\frac{4}{3}, 0, 1)$ and passes through the point $(0,2,0)$ is a solution of the equation system.

Existence and uniqueness of solutions (revisited once again)

Consider the following homogeneous equation system

Example

$$10x_1 \quad -3x_2 \quad -2x_3 \quad = \quad 0 \;\Rightarrow \ldots \sim \left(\begin{array}{ccc|c} 10 & -3 & -2 & 0 \end{array} \right)$$

This is a compatible indeterminate system whose set of solutions is $S = \left\{ \left(\frac{3}{10}x_2 + \frac{1}{5}x_3, x_2, x_3\right) \forall x_2, x_3 \in \mathbb{R} \right\}$, or what is the same

$$S = \mathrm{Span}\left\{ \left(\frac{3}{10}, 1, 0\right), \left(\frac{1}{5}, 0, 1\right) \right\}.$$

That is, any of the infinite points in the plane containing the vectors $(\frac{3}{10}, 1, 0)$ and $(\frac{1}{5}, 0, 1)$ is a solution of the equation system.

Existence and uniqueness of solutions (revisited once again)

Consider now the following non-homogeneous equation system

Example

$$\begin{array}{rrrcl} 10x_1 & -3x_2 & -2x_3 & = & 10 \end{array} \Rightarrow \ldots \sim \left(\begin{array}{rrr|r} 10 & -3 & -2 & 10 \end{array} \right)$$

This is a compatible indeterminate system whose set of solutions is $S = \left\{(1 + \frac{3}{10}x_2 + \frac{1}{5}x_3, x_2, x_3)\ \forall x_2, x_3 \in \mathbb{R}\right\}$, or what is the same

$$S = \left\{(1, 0, 0) + (\tfrac{3}{10}x_2 + \tfrac{1}{5}x_3, x_2, x_3)\ \forall x_2, x_3 \in \mathbb{R}\right\} = (1, 0, 0) + \mathrm{Span}\left\{(\tfrac{3}{10}, 1, 0), (\tfrac{1}{5}, 0, 1)\right\}.$$

Existence and uniqueness of solutions (revisited once again)

Corollary 4.1

Consider the compatible, non-homogeneous equation system given by $A\mathbf{x} = \mathbf{b}$ *and its homogeneous counterpart* $A\mathbf{x} = \mathbf{0}$. *Let* S_h *be the set of solutions of the homogeneous equation system. Then, the set of solutions of the non-homogeneous equation system is of the form*

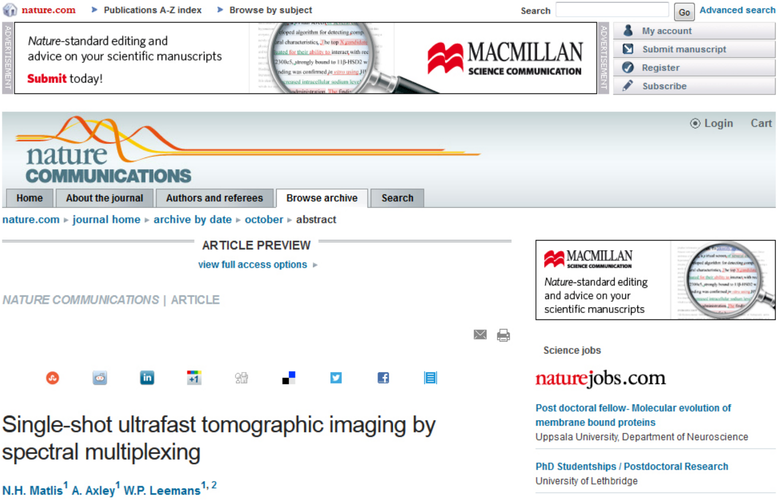
$$S_{nh} = \mathbf{x}_0 + S_h$$

For some $\mathbf{x}_0 \in \mathbb{R}^n$.

Definition 4.1 (Null space of A)

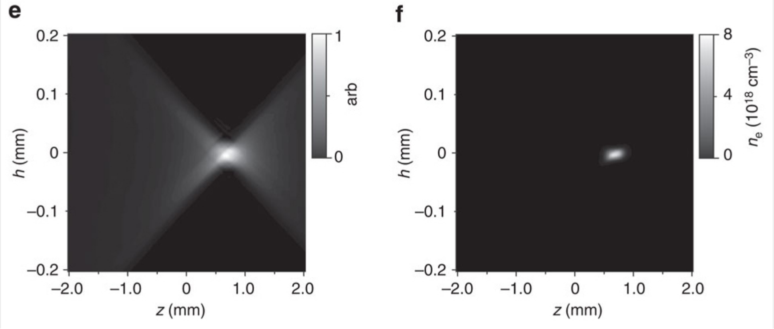
S_h *is called the null space of the matrix* A. *It has the property that given an equation system* $A\mathbf{x} = \mathbf{b}$, *if* $\mathbf{x}_0$ *is a solution of the equation system, then* $\mathbf{x}_0 + \mathbf{x}_h$ *is also a solution, for any* $\mathbf{x}_h \in S_h$.

Existence and uniqueness of solutions (revisited once again)



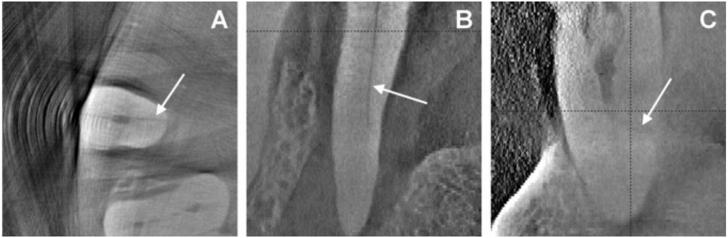
Existence and uniqueness of solutions (revisited once again)

In this example, the authors describe how to solve a problem appearing in the tomographic use of a certain microscope due to the absence of some measurements (resulting in an important null space of the tomographic problem).



Existence and uniqueness of solutions (revisited once again)

In this example, the authors describe how the exact location of a tooth fracture is uncertain (Fig. C) due to the artifacts introduced by the null space of the tomographic problem.



Mora, M. A.; Mol, A.; Tyndall, D. A., Rivera, E. M. *In vitro assessment of local computed tomography for the detection of longitudinal tooth fractures*. Oral Surg Oral Med Oral Pathol Oral Radiol Endod, **2007**, 103, 825-829.

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 5:

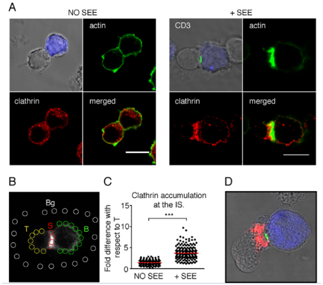
- 1.5.11
- 1.5.13
- 1.5.19
- 1.5.21
- 1.5.25
- 1.5.26
- 1.5.36
- 1.5.39

Outline

2. Linear equation system
    - Introduction (a)
    - Gauss-Jordan algorithm (b)
    - Interpretation as a subspace (b)
    - Existence and uniqueness of solutions (c)
    - **Applications (c)**
    - Linear independence (c)
    - Linear transformations (d)
    - Geometrical transformations (e)
    - Classification of functions (e)
    - More applications (e)

Applications

In fluorescence microscopy, we can quantitatively measure the amount of fluorescence coming from each source with a linear equation system.

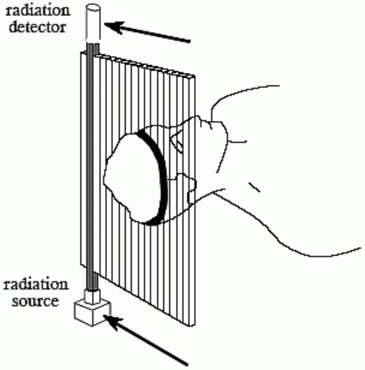


$$\begin{pmatrix} Bg \\ B \\ T \\ S \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \widehat{Bg} \\ \hat{B} \\ \hat{T} \\ \hat{T}_s \end{pmatrix}$$

C. Calabia-Linares, M. Pérez-Martínez, N. Martín-Cofreces, M. Alfonso-Pérez, C. Gutiérrez-Vázquez, M. Mittelbrunn, S. Ibiza, F.R. Urbano-Olmos, C. Aguado-Ballano, C.O.S. Sorzano, F. Sánchez-Madrid, E. Veiga. *Clathrin drives actin accumulation at the immunological synapse*. J. Cell Science, 124: 820-830 (**2011**)

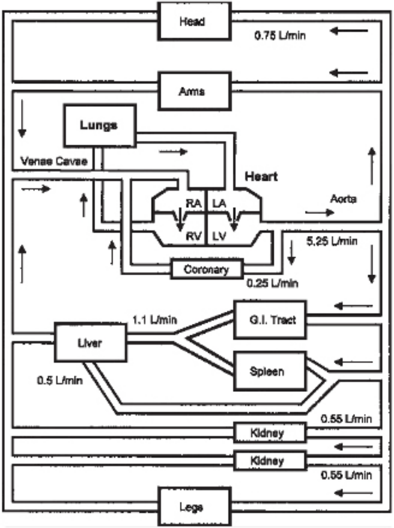
In computed tomography, a simple model (but widely used) for data collection states that the data observed is the sum of the values of the density found along the X-ray path.

FIGURE 25-14
CT data acquisition. A simple CT system passes a narrow beam of x-rays through the body from source to detector. The source and detector are then translated to obtain a complete view. The remaining views are obtained by rotating the source and detector in about 1° increments, and repeating the translation process.



Applications

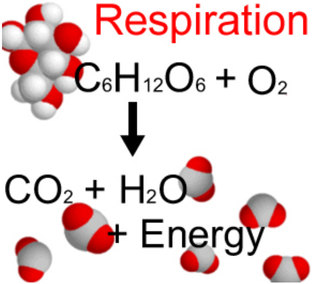
In the blood system, at each node, the sum of output flows must be equal to the sum of input flows.



Applications

In a very simplified model, respiration is the burning of glucose that can be written as

$$x_1 C_6 H_{12} O_6 + x_2 O_2 \to x_3 CO_2 + x_4 H_2 O$$



C: $6x_1 = x_3$

H: $6x_1 = 2x_4$

O: $6x_1 + 2x_2 = 2x_3 + x_4$

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 6:

- 1.6.5
- 1.6.7
- 1.6.12

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - **Linear independence (c)**
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Linear independence

Definition 6.1 (Linear independence)

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ *is* ***linearly independent*** *if*

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + ... + x_p\mathbf{v}_p = \mathbf{0} \Rightarrow x_1 = x_2 = ... = x_p = 0$$

That is the only solution of the previous equation is the trivial solution $\mathbf{x} = \mathbf{0}$*.*
The set is ***linearly dependent*** *if at least two* x_i*'s are different from 0.*

Linear independence

Example

Determine if the vectors $\mathbf{v}_1 = (1, 2, 3)$, $\mathbf{v}_2 = (4, 5, 6)$, and $\mathbf{v}_3 = (2, 1, 0)$ are linearly independent.

Solution

The augmented matrix associated to the equation system in Definition 6.1 is

$$\left(\begin{array}{ccc|c} 1 & 4 & 2 & 0 \\ 2 & 5 & 1 & 0 \\ 3 & 6 & 0 & 0 \end{array}\right) \sim \ldots \sim \left(\begin{array}{rrr|c} 1 & 4 & 2 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Since the system is compatible indeterminate, there exists a solution apart from the trivial solution and, therefore, the vectors are linearly dependent.

Linear independence

Example

If possible, find a linear relationship among the three vectors. *Solution*
We continue transforming the augmented matrix to its reduced echelon form

$$\left(\begin{array}{ccc|c} 1 & 4 & 2 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \sim \ldots \sim \left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

From which $x_1 = 2x_3$ and $x_2 = -x_3$. Simply by choosing $x_3 = 1$, we obtain have that a possible solution to the equation system in Definition 6.1 is $x_1 = 2$, $x_2 = -1$ and $x_3 = 1$, consequently we have that

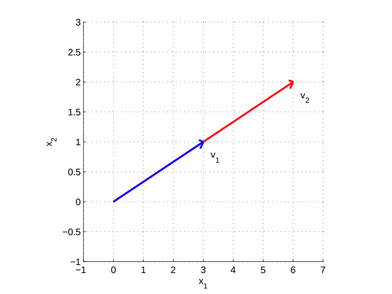
$$2\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$$

Linear independence

Example

$\mathbf{v}_1 = (3, 1)$ and $\mathbf{v}_2 = (6, 2)$ are linearly dependent because

$$\mathbf{v}_2 = 2\mathbf{v}_1 \Rightarrow -2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0} \Rightarrow \mathbf{v}_1 = \frac{1}{2}\mathbf{v}_2$$

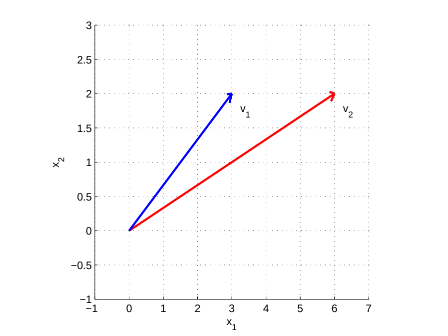


If two vectors are linearly dependent of each other, then any one of them is a multiple of the other.

Linear independence

Example

$\mathbf{v}_1 = (3, 2)$ and $\mathbf{v}_2 = (6, 2)$ are linearly independent

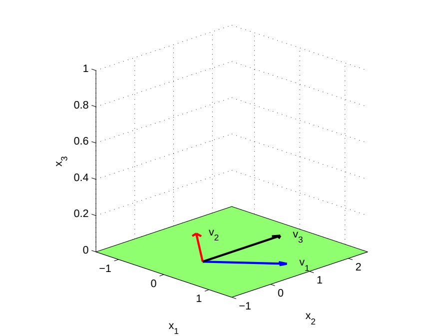


Linear independence

Example

$\mathbf{v}_1 = (1, 1, 0)$, $\mathbf{v}_2 = (-1, 1, 0)$ and $\mathbf{v}_3 = (0, 2, 0)$ are linearly dependent because

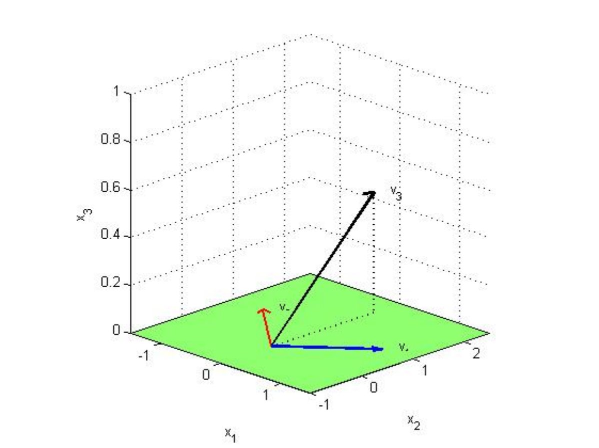
$$\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2$$



Linear independence

Example

$\mathbf{v}_1 = (1, 1, 0)$, $\mathbf{v}_2 = (-1, 1, 0)$ and $\mathbf{v}_3 = (0, 2, 1)$ are linearly independent



Linear independence

Theorem 6.1 (Linear independence of matrix columns)

The columns of the matrix A are linearly independent iff the only solution of $A\mathbf{x} = \mathbf{0}$ is the trivial one.

Proof

Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_p]$ so that the columns of the matrix A are the vectors $\mathbf{a}_i$. According to Definition 6.1 these vectors are linearly independent iff

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_p\mathbf{a}_p = \mathbf{0} \Rightarrow x_1 = x_2 = \ldots = x_p = 0$$

or what is the same

$$A\mathbf{x} = \mathbf{0} \Rightarrow x_1 = x_2 = \ldots = x_p = 0$$

as stated by the theorem (q.e.d.)

Linear independence

Theorem 6.2

Any set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ *with* $\mathbf{v}_i \in \mathbb{R}^n$ *is linearly dependent if* $p > n$.

Proof

Let $A = [\mathbf{v}_1\ \mathbf{v}_2\ ...\ \mathbf{v}_p]$ *and let us consider the equation system* $A\mathbf{x} = \mathbf{0}$. *If* $p > n$ *there are more unknowns than equations, and consequently, there are free variables and the system is compatible indeterminate. Thus, there are more solutions apart from the trivial one and the set of vectors is linearly dependent.*

Theorem 6.3

If any set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ *with* $\mathbf{v}_i \in \mathbb{R}^n$ *contains the vector* $\mathbf{0}$, *then the set of vectors is linearly dependent.*

Proof

We can assume, without loss of generality, that $\mathbf{v}_1 = \mathbf{0}$. *Then, we can set* $x_1 = 1$, $x_2 = x_3 = ... = x_p = 0$ *so that the following equation is met:*

$$1\mathbf{v}_1 + 0\mathbf{v}_2 + ... + 0\mathbf{v}_p = \mathbf{0} \ (q.e.d.)$$

Linear independence

Theorem 6.4

A set of vectors is linearly dependent iff at least 1 of the vectors is linearly dependent on the rest

Proof

Proof $\Leftarrow$

Let us assume that $\mathbf{v}_j$ *is a linear combination of the rest of the vectors, that is,*

$$\mathbf{v}_j = \sum_{k \neq j} x_k \mathbf{v}_k$$

Then, we can write $\mathbf{v}_j - \sum_{k \neq j} x_k \mathbf{v}_k = \mathbf{0} \Rightarrow$

$$-x_1 \mathbf{v}_1 - x_2 \mathbf{v}_2 - ... - x_{j-1} \mathbf{v}_{j-1} + \mathbf{v}_j - x_{j+1} \mathbf{v}_{j+1} - x_p \mathbf{v}_p = \mathbf{0}$$

And consequently there exists a non-trivial solution of the equation of Definition 6.1.

Linear independence

Proof $\Rightarrow$

If $\mathbf{v}_1 = \mathbf{0}$, then we have already a vector that is a trivial combination of the rest ($\mathbf{v}_1 = 0\mathbf{v}_2 + 0\mathbf{v}_3 + \ldots + 0\mathbf{v}_p$).

If $\mathbf{v}_1 \neq \mathbf{0}$, then there exist some coefficients such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_p\mathbf{v}_p = \mathbf{0}$$

Let j be the largest index for which $x_j \neq 0$ (that is, $x_{j+1} = x_{j+2} = \ldots = x_p = 0$).

If $j = 1$, then $x_1\mathbf{v}_1 = \mathbf{0}$, but this is not possible because $\mathbf{v}_1 \neq \mathbf{0}$. Then, $j > 1$ and consequently

$$\begin{gathered} x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_j\mathbf{v}_j + 0\mathbf{v}_{j+1} + \ldots + 0\mathbf{v}_p = \mathbf{0} \Rightarrow \\ x_j\mathbf{v}_j = -x_1\mathbf{v}_1 - x_2\mathbf{v}_2 - \ldots - x_{j-1}\mathbf{v}_{j-1} \Rightarrow \\ \mathbf{v}_j = -\tfrac{x_1}{x_j}\mathbf{v}_1 - \tfrac{x_2}{x_j}\mathbf{v}_2 - \ldots - \tfrac{x_{j-1}}{x_j}\mathbf{v}_{j-1} \text{ (q.e.d.)} \end{gathered}$$

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 7:

- 1.7.9
- 1.7.39
- 1.7.40
- 1.7.41 (bring computer)

Outline

2. Linear equation system
    - Introduction (a)
    - Gauss-Jordan algorithm (b)
    - Interpretation as a subspace (b)
    - Existence and uniqueness of solutions (c)
    - Applications (c)
    - Linear independence (c)
    - Linear transformations (d)
    - Geometrical transformations (e)
    - Classification of functions (e)
    - More applications (e)

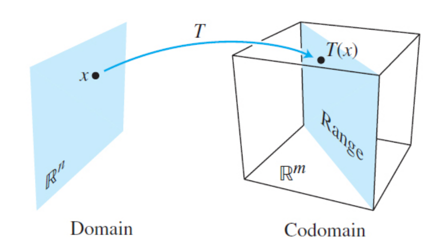
Linear transformations

Definition 7.1 (Transformation)

*A **transformation** (or **function** or **mapping**), T, from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector of $\mathbb{R}^n$ a vector of $\mathbb{R}^m$.*

$$\begin{aligned} T : \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \mathbf{x} &\rightarrow T(\mathbf{x}) \end{aligned}$$

*$\mathbb{R}^n$ is called the **domain** of the transformation, and $\mathbb{R}^m$ its **codomain**. $T(\mathbf{x})$ is the image of vector $\mathbf{x}$ under the action of T. The set of all images is the **range** of T.*



Linear transformations

Definition 7.2 (Matrix transformation)

*T is a **matrix transformation** iff $T(\mathbf{x}) = A\mathbf{x}$ for some matrix $A \in \mathcal{M}_{m\times n}$.*

Example

Let us consider $A = \begin{pmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{pmatrix}$ and the matrix transformation $\mathbf{y} = A\mathbf{x}$. For instance, the image of $\mathbf{x} = (1, 1, 1, 1)$ is

$$\mathbf{y} = \begin{pmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

The equation system $A\mathbf{x} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$ looks for all those $\mathbf{x}$, if any, such that $T(\mathbf{x}) = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$. The domain of this transformation is $\mathbb{R}^4$ and its codomain $\mathbb{R}^2$.

Linear transformations

Example

Let us consider $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ and the matrix transformation $\mathbf{y} = A\mathbf{x}$. The domain of this transformation is $\mathbb{R}^2$ and its codomain $\mathbb{R}^3$. However, not all points in $\mathbb{R}^3$ need to be an image of some point $\mathbf{x} \in \mathbb{R}^2$, only a subset of them may be. In this case,

$$\mathbb{R}^3 \supset \mathrm{Range}(T) = \langle (1,0,0), (0,1,0) \rangle$$

In general, the range of the transformation T is the subspace spanned by the columns of the matrix A.

Linear transformations

Example

Let us consider $A = \begin{pmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{pmatrix}$ and the matrix transformation $\mathbf{y} = A\mathbf{x}$.

1. What is the image of $\mathbf{u} = (2, -1)$ under T?
   $T(\mathbf{u}) = A\mathbf{u} = (5, 1, 9)$
2. Let $\mathbf{b} = (3, 2, -5)$. Which is $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{b}$?
   $$\left(\begin{array}{cc|c} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{array}\right)$$
   From which we deduce $\mathbf{x} = (\frac{3}{2}, -\frac{1}{2})$.
3. Is there any other $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{b}$?
   No, the previous solution is unique because the equation system is definite compatible.

Example

4. Does $\mathbf{c} = (3, 2, 5)$ belong to $\mathrm{Range}(T)$?

$$\left(\begin{array}{rr|r} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{array}\right) \sim \left(\begin{array}{rr|r} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -35 \end{array}\right)$$

Since the system is incompatible, we deduce that there is no vector $\mathbf{x}$ whose image is $\mathbf{c}$ and, consequently, $\mathbf{c} \notin \mathrm{Range}(T)$.

5. Which is the function $\mathbf{y} = T(\mathbf{x})$?

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{pmatrix}$$

6. Which is $\mathrm{Range}(T)$?

$\mathrm{Range}(T) = \langle (1, 3, -1), (-3, 5, 7) \rangle =$

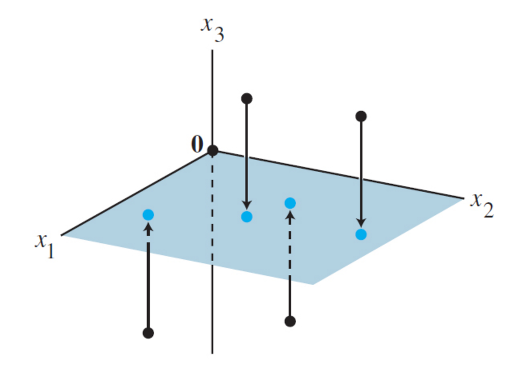
$$\left\{ \mathbf{y} \in \mathbb{R}^3 | \mathbf{y} = x_1 \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} + x_2 \begin{pmatrix} -3 \\ 5 \\ 7 \end{pmatrix} \ \forall x_1, x_2 \in \mathbb{R} \right\}$$

Because $(1, 3, -1)$ and $(-3, 5, 7)$ are linearly independent, $\mathrm{Range}(T)$ is a plane.

Linear transformations

Example

Consider the transformation $T(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$. This is a projection transformation that projects any 3D point onto the XY plane.



Linear transformations

Definition 7.3 (Linear transformation)

*T is a **linear transformation** iff* $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathrm{Dom}(T)$, $\forall c \in \mathbb{R}$

1. $T(\mathbf{x}_1 + \mathbf{x}_2) = T(\mathbf{x}_1) + T(\mathbf{x}_2)$
2. $T(c\mathbf{x}_1) = cT(\mathbf{x}_1)$

Theorem 7.1

If $T(\mathbf{x})$ is a linear transformation, then

1. $T(\mathbf{0}) = \mathbf{0}$
2. $T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1 T(\mathbf{x}_1) + c_2 T(\mathbf{x}_2)\ \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathrm{Dom}(T),\ \forall c_1, c_2 \in \mathbb{R}$

Proof

1. $T(\mathbf{0}) = T(0\mathbf{x}_1) =$*[(2), Def. 7.3]*$= 0T(\mathbf{x}_1) = \mathbf{0}$ *(q.e.d.)*
2. $T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) =$*[(1), Def. 7.3]*$= T(c_1\mathbf{x}_1) + T(c_2\mathbf{x}_2) =$*[(2), Def. 7.3]* $c_1 T(\mathbf{x}_1) + c_2 T(\mathbf{x}_2)$ *(q.e.d.)*

Theorem 7.2

If $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathrm{Dom}(T)$, $\forall c_1, c_2 \in \mathbb{R}$ it is verified that $T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1 T(\mathbf{x}_1) + c_2 T(\mathbf{x}_2)$, then $T(\mathbf{x})$ is a linear transformation.

Proof

1. *Let us consider the case $c_1 = c_2 = 1$, then according to the assumption of the theorem we have $T(\mathbf{x}_1 + \mathbf{x}_2) = T(\mathbf{x}_1) + T(\mathbf{x}_2)$, which implies (1) in Def. 7.3.*
2. *Let us consider the case $c_2 = 0$, then according to the assumption of the theorem we have $T(c_1\mathbf{x}_1) = c_1 T(\mathbf{x}_1)$, which implies (2) in Def. 7.3.*

(q.e.d.)

Linear transformations

Corollary: Principle of superposition

If $\forall \mathbf{x}_i \in \mathrm{Dom}(T)$, $\forall c_i \in \mathbb{R}$ it is verified that $T\left(\sum_i c_i \mathbf{x}_i\right) = \sum_i c_i T(\mathbf{x}_i)$.

Proof

Apply the previous theorem multiple times. (q.e.d.)

Linear transformations

Example

Show that $T(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is a linear transformation.

<u>*Proof*</u>

1. Show that $T(\mathbf{x}_1 + \mathbf{x}_2) = T(\mathbf{x}_1) + T(\mathbf{x}_2)$
   On one side we have $T(\mathbf{x}_1 + \mathbf{x}_2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_{11} + x_{21} \\ x_{12} + x_{22} \end{pmatrix} = \begin{pmatrix} x_{11} + x_{21} \\ -x_{12} - x_{22} \end{pmatrix}$
   On the other side we have $T(\mathbf{x}_1) + T(\mathbf{x}_2) =$
   $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} = \begin{pmatrix} x_{11} \\ -x_{12} \end{pmatrix} + \begin{pmatrix} x_{21} \\ -x_{22} \end{pmatrix} = \begin{pmatrix} x_{11} + x_{21} \\ -x_{12} - x_{22} \end{pmatrix}$
   Obviously, these two calculations give the same result.
2. Show that $T(c_1\mathbf{x}_1) = c_1 T(\mathbf{x}_1)$

$$\begin{aligned} T(c_1\mathbf{x}_1) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} c_1x_{11} \\ c_1x_{12} \end{pmatrix} = \begin{pmatrix} c_1x_{11} \\ -c_1x_{12} \end{pmatrix} = c_1 \begin{pmatrix} x_{11} \\ -x_{12} \end{pmatrix} \\ &= c_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = c_1 T(\mathbf{x}_1) \end{aligned}$$

Linear transformations

Theorem 7.3

Any matrix transformation is a linear transformation.

Proof

1. *Show that* $T(\mathbf{x}_1 + \mathbf{x}_2) = T(\mathbf{x}_1) + T(\mathbf{x}_2)$
   $T(\mathbf{x}_1 + \mathbf{x}_2) = A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = T(\mathbf{x}_1) + T(\mathbf{x}_2)$ *(q.e.d.)*
2. *Show that* $T(c_1\mathbf{x}_1) = c_1 T(\mathbf{x}_1)$
   $T(c_1\mathbf{x}_1) = A(c_1\mathbf{x}_1) = c_1(A\mathbf{x}_1) = c_1 T(\mathbf{x}_1)$ *(q.e.d.)*

Reinterpreting the columns of a matrix

Example

Consider $T(\mathbf{x}) = A\mathbf{x}$ with $A = \begin{pmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{pmatrix}$. Consider the standard canonical basis of $\mathbb{R}^4$ formed by the vectors $\mathbf{e}_1 = (1, 0, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0, 0)$, $\mathbf{e}_3 = (0, 0, 1, 0)$, and $\mathbf{e}_4 = (0, 0, 0, 1)$. Let us consider the transformation of each one of these vectors

$$T(\mathbf{e}_1) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \quad T(\mathbf{e}_2) = \begin{pmatrix} -3 \\ 0 \end{pmatrix} \quad T(\mathbf{e}_3) = \begin{pmatrix} 1 \\ 5 \end{pmatrix} \quad T(\mathbf{e}_4) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

In general, we note that the transformation of $\mathbf{e}_i$ is the i-th column of matrix A.

Corollary

The columns of the matrix $A \in \mathcal{M}_{m\times n}$ can be understood as the transformations of the canonical basis of $\mathbb{R}^n$:

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{pmatrix}$$

Reinterpreting the columns of a matrix

Example (continued)

In the previous example consider transforming the vector $\mathbf{x} = (1, -2, 3, 5)$. This vector is equal to

$$\mathbf{x} = \mathbf{e}_1 - 2\mathbf{e}_2 + 3\mathbf{e}_3 + 5\mathbf{e}_4$$

Then, we have

$$\begin{aligned} T(\mathbf{x}) &= T(\mathbf{e}_1 - 2\mathbf{e}_2 + 3\mathbf{e}_3 + 5\mathbf{e}_4) = T(\mathbf{e}_1) - 2T(\mathbf{e}_2) + 3T(\mathbf{e}_3) + 5T(\mathbf{e}_4) \\ &= \begin{pmatrix} 4 \\ 2 \end{pmatrix} - 2\begin{pmatrix} -3 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 5 \end{pmatrix} + 5\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 28 \\ 22 \end{pmatrix} \end{aligned}$$

Exercises

Exercises

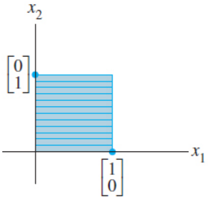
From Lay (4th ed.), Chapter 1, Section 8:

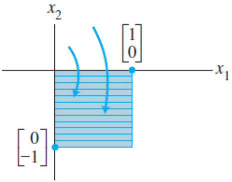
- 1.8.23
- 1.8.25
- 1.8.26
- 1.8.30
- 1.8.34

Outline

Geometrical transformations

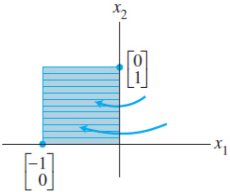
Certain matrix transformations are used to transform the unit square into different shapes. The following table shows some of such transformations.



Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$

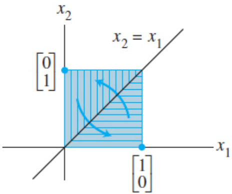
Geometrical transformations

Reflection through the x_2-axis



$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

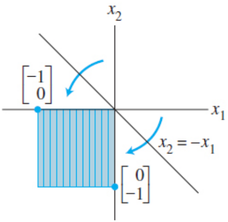
Reflection through the line $x_2 = x_1$



$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

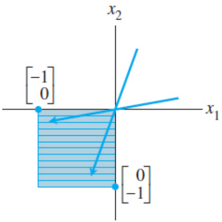
Geometrical transformations

Reflection through the line $x_2 = -x_1$



$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

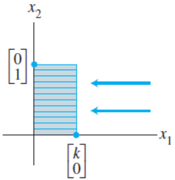
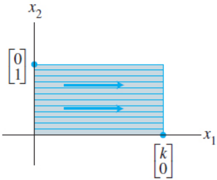
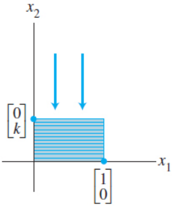
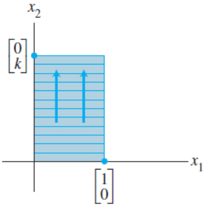
Reflection through the origin



$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

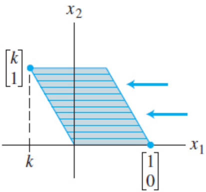
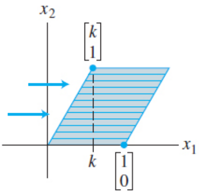
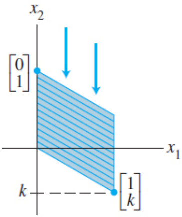
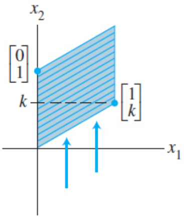
Geometrical transformations

TABLE 2 Contractions and Expansions

Transformation	Image of the Unit Square		Standard Matri
Horizontal contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

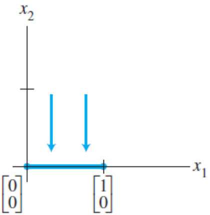
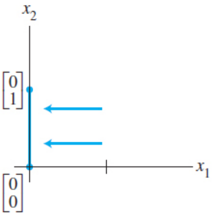
Geometrical transformations

TABLE 3 Shears

Transformation	Image of the Unit Square		Standard Matrix
Horizontal shear	 $k < 0$	 $k > 0$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear	 $k < 0$	 $k > 0$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

Geometrical transformations

TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Classification of functions

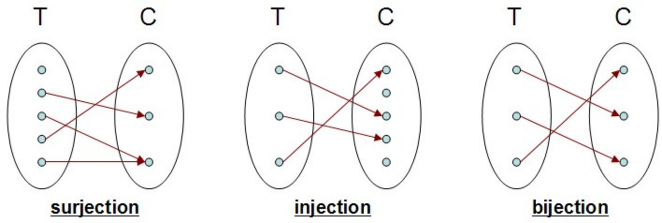
Definition 9.1

Functions can be classified as ***surjective, injective or bijective****:*

Surjective: *A function is surjective if every point of the codomain has* ***at least one*** *point of the domain that maps onto it. They are also called* ***onto*** *functions.*

Injective: *A function is injective if every point of the codomain has* ***at most one*** *point in the domain that maps onto it. They are also called* ***one-to-one*** *functions.*

Bijective: *A function is bijective if it is injective and surjective.*



Classification of functions

Example

Here we have some examples of the classification of functions applied to linear transformations

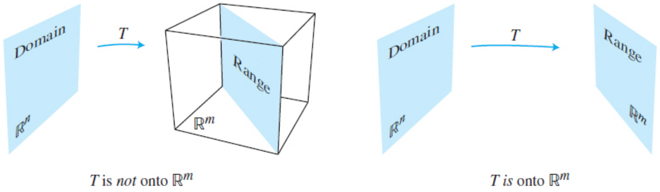


FIGURE 3 Is the range of T all of $\mathbb{R}^m$?

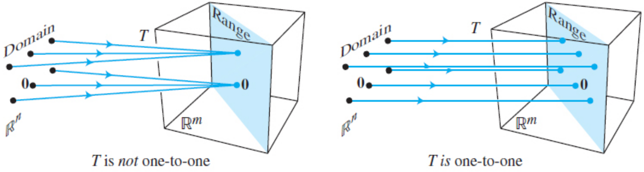


FIGURE 4 Is every $\mathbf{b}$ the image of at most one vector?

Classification of functions

Example

Consider $T(\mathbf{x}) = A\mathbf{x}$ with $A = \begin{pmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{pmatrix}$. This is a transformation from $\mathbb{R}^4$ onto $\mathbb{R}^3$. The columns of A $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_4$ are linearly independent and span $\mathbb{R}^3$ (that is, the function is surjective). Therefore, there must be points in $\mathbb{R}^3$ that come from several points in $\mathbb{R}^4$ (the function is not injective). Let us find some of these points.

$$\left(\begin{array}{cccc|c} 1 & -4 & 8 & 1 & y_1 \\ 0 & 2 & -1 & 3 & y_2 \\ 0 & 0 & 0 & 5 & y_3 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 6 & 0 & y_1 - 2y_2 - \frac{4}{5}y_3 \\ 0 & 1 & -\frac{1}{2} & 0 & \frac{1}{2}y_2 - \frac{3}{10}y_3 \\ 0 & 0 & 0 & 1 & \frac{1}{5}y_3 \end{array}\right) \Rightarrow$$

$$\begin{aligned} x_1 &= y_1 - 2y_2 - \tfrac{4}{5}y_3 - 6x_3 \\ x_2 &= \tfrac{1}{2}y_2 - \tfrac{3}{10}y_3 + \tfrac{1}{2}x_3 \\ x_4 &= \tfrac{1}{5}y_3 \end{aligned}$$

Since x_3 is a free variable, we have that for each point in the codomain, there is a straight line that maps onto it (the equation of the line is the one given above).

Classification of functions

Theorem 9.1

Let $T(\mathbf{x})$ be a linear transformation. $T(\mathbf{x})$ is an injective function iff $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

Proof

Proof $\Rightarrow$

If T is injective, then, by definition, every point of the codomain, in particular $\mathbf{0}$ is the mapping of at most one point in the domain. We already know that for any linear transformation $T(\mathbf{0}) = \mathbf{0}$, therefore, $\mathbf{x} = \mathbf{0}$ must be the unique solution of the equation $T(\mathbf{x}) = \mathbf{0}$.

Classification of functions

Proof $\Leftarrow$

For any linear transformation we know that $T(\mathbf{0}) = \mathbf{0}$. Let us assume that the statement is false, that is $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, but T is not injective. IF T is not injective there exist a point $\mathbf{y}$ in the codomain that is the image of two points in the domain

$$\begin{array}{c} T(\mathbf{x}_1) = \mathbf{y} \\ T(\mathbf{x}_2) = \mathbf{y} \end{array}$$

If we know subtract the two equations we have

$$\begin{array}{cl} T(\mathbf{x}_1) - T(\mathbf{x}_2) = \mathbf{0} & \\ T(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0} & T \text{ is linear} \\ \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0} & \text{There is only one solution of } T(\mathbf{x}) = \mathbf{0} \\ \mathbf{x}_1 = \mathbf{x}_2 & \text{contradiction (q.e.d.)} \end{array}$$

Classification of functions

Theorem 9.2

Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation. Then:

1. $\mathrm{Range}(T) = \mathbb{R}^m$ *iff* $\mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n) = \mathbb{R}^m$.
2. *T is injective iff all columns of A are linearly independent.*

Proof

1. *According to Theorem 3.2, the columns of A span $\mathbb{R}^m$ if for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ is consistent, that is, if there exists at least one solution of $T(\mathbf{x}) = \mathbf{b}$. If this is true, then $\mathrm{Range}(T) = \mathbb{R}^m$.*
2. *According to Theorem 9.1, T is injective iff $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, or what is the same iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This happens only if the columns of A are linearly independent as stated by Theorem 6.1.*

Classification of functions

Example

Let $T(\mathbf{x}) = \begin{pmatrix} 3x_1 + x_2 \\ 5x_1 + 7x_2 \\ x_1 + 3x_2 \end{pmatrix}$:

1. Show that it is a linear transformation
2. Does it map $\mathbb{R}^2$ onto $\mathbb{R}^3$?

Solution

1. The transformation is of the form $T(\mathbf{x}) = A\mathbf{x}$ with $A = \begin{pmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{pmatrix}$ and, therefore, the transformation is linear.
2. The columns of A are linearly independent (because they are not multiples of each other), then, by the previous theorem, the transformation is injective. However, they do not span all $\mathbb{R}^3$ (since they are only two vectors and for spanning all $\mathbb{R}^3$ we need at least 3 vectors). Consequently, the transformation is not surjective, and it does not map $\mathbb{R}^2$ onto $\mathbb{R}^3$.

Exercises

Exercises

From Lay (4th ed.), Chapter 1, Section 9:

- 1.9.1
- 1.9.3
- 1.9.17
- 1.9.33
- 1.9.36
- 1.9.37
- 1.9.39

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

More applications

Construction of a diet

Given the following nutritional information:

Amounts (g) Supplied per 100 g of Ingredient				Amounts (g) Supplied by Cambridge Diet in One Day
Nutrient	Nonfat milk	Soy flour	Whey	
Protein	36	51	13	33
Carbohydrate	52	34	74	45
Fat	0	7	1.1	3

What is the amount of nonfat milk, soy flour and whey needed to provide the protein carbohydrate and fat planned for one day?

Solution

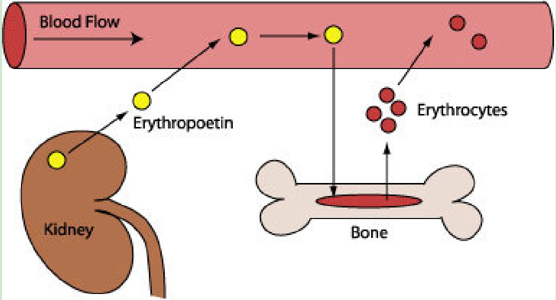
$$\left(\begin{array}{ccc|c} 36 & 51 & 13 & 33 \\ 52 & 34 & 74 & 45 \\ 0 & 7 & 11 & 3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0.277 \\ 0 & 1 & 0 & 0.392 \\ 0 & 0 & 1 & 0.233 \end{array}\right)$$

That is, we need $x_1 = 0.277 \cdot 100\text{g} = 277\text{g}$ of non-fat milk, $x_2 = 392\text{g}$ of soy flour and $x_3 = 233$ g of whey.

More applications

Dynamic systems: difference equations

In a simplistic model red blood cells (erythrocytes) are created in the bone marrow, then some of them pass to the blood. After some time, old red blood cells are destroyed in the spleen (bazo).



More applications

Dynamic systems: difference equations (continued)

Let's say that at every time interval:

- 5% of the erythrocytes in the marrow leave to the blood stream.
- 2% of the erythrocytes in the blood stream are destroyed by the spleen.
- 1M new red blood cells are created at the marrow.

The following equation can be used to determine the amount of erythrocytes at any moment

$$\begin{pmatrix} x_{marrow}^{(k+1)} \\ x_{blood}^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0.95 & 0 \\ 0.05 & 0.98 \end{pmatrix} \begin{pmatrix} x_{marrow}^{(k)} \\ x_{blood}^{(k)} \end{pmatrix} + \begin{pmatrix} 10^6 \\ 0 \end{pmatrix}$$

This kind of models is called difference equations.

Outline

2. Linear equation system
   - Introduction (a)
   - Gauss-Jordan algorithm (b)
   - Interpretation as a subspace (b)
   - Existence and uniqueness of solutions (c)
   - Applications (c)
   - Linear independence (c)
   - Linear transformations (d)
   - Geometrical transformations (e)
   - Classification of functions (e)
   - More applications (e)

Chapter 3. Matrix algebra

C.O.S. Sorzano

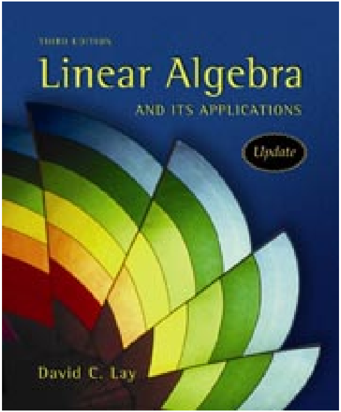
Biomedical Engineering

December 3, 2013

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 2.

A little bit of history

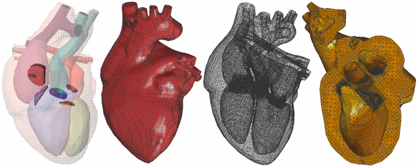
Matrices appeared as a regular arrangement of numbers more than 2,000 years ago. However, it was during the XVIIth, XVIIIth and XIXth centuries that they developed in the way we know them now. Some important names in their modern development are Seki Takakazu (1683), Gottfried Leibniz (1693), Gabriel Cramer (1750), James Sylvester (1850), and Arthur Cayley (1858). They were applied in all kind of mathematical problems as a way to organize calculations.

To know more about the history of matrix algebra visit

- http://www-groups.dcs.st-and.ac.uk/~history/PrintHT/Matrices_and_determinants.html

Applications

Finite elements has been one of the most successful approaches to biomechanical modeling. In the figure we show one of such a model for the heart. Using this model, all kind of local stresses can be calculated.



(a) (b) (c) (d)

$$\begin{array}{ccccc} nx1 & & nxn & & nx1 \\ \begin{pmatrix} f_{bc} \\ f_{int} \\ f_{vis} \\ f_{con} \end{pmatrix} & = & \begin{bmatrix} K_{bc1} & K_{bc2} & K_{bc3} & K_{bc4} \\ K_{int1} & K_{int2} & K_{int3} & K_{int4} \\ K_{vis1} & K_{vis2} & K_{vis3} & K_{vis4} \\ K_{con1} & K_{con2} & K_{con3} & K_{con4} \end{bmatrix} & & \begin{pmatrix} d_{bc} \\ d_{int} \\ d_{vis} \\ d_{con} \end{pmatrix} \end{array}$$

(a)

$$\begin{pmatrix} f'_{bc} \\ f'_{int} \\ f'_{vis} \\ f'_{con} \end{pmatrix} = \begin{bmatrix} K'_{bc1} & K'_{bc2} & K'_{bc3} & K'_{bc4} \\ 0 & K'_{int2} & K'_{int3} & K'_{int4} \\ 0 & 0 & K'_{vis3} & K'_{vis4} \\ 0 & 0 & 0 & K'_{con4} \end{bmatrix} \begin{pmatrix} d_{bc} \\ d_{int} \\ d_{vis} \\ d_{con} \end{pmatrix}$$

(b)

$$\begin{pmatrix} f''_{bc} \\ f''_{int} \\ f''_{vis} \\ f''_{con} \end{pmatrix} = \begin{bmatrix} K''_{bc} & 0 & 0 & 0 \\ 0 & K''_{int} & 0 & 0 \\ 0 & 0 & K''_{vis} & 0 \\ 0 & 0 & 0 & K''_{con} \end{bmatrix} \begin{pmatrix} d_{bc} \\ d_{int} \\ d_{vis} \\ d_{con} \end{pmatrix}$$

(c)

J. Berkley, S. Weghorst, H. Gladstone, G. Raugi, D. Berg, M. Ganter. Banded Matrix Approach to Finite Element Modeling for Soft Tissue Simulation.

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

Definition 1.1 (Matrix)

*Informally, we can define a **matrix** as a regular arrangement of numbers that are laid out in a grid of m rows and n columns. More formally, we say that* $A \in \mathcal{M}_{m\times n}$. *We denote as* $\mathbf{a}_j$ *as its j-th **column**, and* a_{ij} *the element in the i-th row and the j-th column.*

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

*The **main diagonal** is the vector given by* $(a_{11}, a_{22}, \ldots)$. *Two important special matrices are the **identity matrix*** ($I \in \mathcal{M}_{n\times n}$) *that is zero everywhere except the main diagonal that is full of 1s; and the **zero matrix*** ($0 \in \mathcal{M}_{m\times n}$) *that is zero everywhere.*

Example

MATLAB: `A=[1 2 3; 4 5 6]`

Matrix operations

Definition 1.2 (Sum with a scalar)

*We define the **sum operator between a scalar and a matrix** as:*

$$\begin{array}{rcll} + : \ \mathbb{R} \times \mathcal{M}_{m\times n} & \rightarrow & \mathcal{M}_{m\times n} & \\ +(k, A) & \rightarrow & B = k + A & |b_{ij} = k + a_{ij} \end{array}$$

*We overload the notation to define the **sum operator between a matrix and a scalar** as*

$$\begin{array}{rcll} + : \ \mathcal{M}_{m\times n} \times \mathbb{R} & \rightarrow & \mathcal{M}_{m\times n} & \\ +(A, k) & \rightarrow & B = A + k & |b_{ij} = a_{ij} + k \end{array}$$

Example

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{pmatrix} \\ B = 1 + A &= \begin{pmatrix} 2 & 3 & 4 \\ 0 & -1 & -2 \end{pmatrix} \end{aligned}$$

MATLAB: `B=1+A`

Properties

$$\begin{aligned} k + A &= A + k \\ (k_1 + k_2) + A &= k_1 + (k_2 + A) \end{aligned}$$

Matrix operations

Definition 1.3 (Multiplication with a scalar)

We define the ***multiplication operator between a scalar and a matrix*** *as:*

$$\begin{array}{rlll} \cdot : & \mathbb{R} \times \mathcal{M}_{m\times n} & \to & \mathcal{M}_{m\times n} \\ & \cdot(k, A) & \to & B = k + A \quad | b_{ij} = k a_{ij} \end{array}$$

We overload the notation to define the ***multiplication operator between a matrix and a scalar*** *as*

$$\begin{array}{rlll} \cdot : & \mathcal{M}_{m\times n} \times \mathbb{R} & \to & \mathcal{M}_{m\times n} \\ & \cdot(A, k) & \to & B = Ak \quad | b_{ij} = a_{ij} k \end{array}$$

Example

$$\begin{array}{rcl} A & = & \begin{pmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{pmatrix} \\ B = 2A & = & \begin{pmatrix} 2 & 4 & 6 \\ -2 & -4 & -6 \end{pmatrix} \end{array}$$

MATLAB: `B=2*A`

Properties

$$\begin{array}{rcl} kA & = & Ak \\ (k_1 k_2)A & = & k_1(k_2 A) \\ (k_1 + k_2)A & = & k_1 A + k_2 A \end{array}$$

Matrix operations

Definition 1.4 (Sum of two matrices)

We define the ***sum operator between two matrices*** *as:*

$$\begin{array}{rcl} + : \quad \mathcal{M}_{m\times n} \times \mathcal{M}_{m\times n} & \rightarrow & \mathcal{M}_{m\times n} \\ +(A,B) & \rightarrow & C = A + B \quad | c_{ij} = a_{ij} + b_{ij} \end{array}$$

Example

$$\begin{array}{rcl} A & = & \begin{pmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{pmatrix} \\ B & = & \begin{pmatrix} 4 & 5 & 6 \\ 0 & 1 & 1 \end{pmatrix} \\ C = A + B & = & \begin{pmatrix} 5 & 7 & 9 \\ -1 & -1 & -2 \end{pmatrix} \end{array}$$

MATLAB: `C=A+B`

Properties

$$\begin{array}{rcl} A + B & = & B + A \\ A + (B + C) & = & (A + B) + C \\ A + 0 & = & A \\ k(A + B) & = & kA + kB \end{array}$$

Proof of the properties

We are not proving all properties, although all of them follow the same strategy. Let's see an example

$$k(A+B) = kA + kB$$

Proof
Let us develop the left hand side

$$\begin{array}{r|l} C = A + B & c_{ij} = a_{ij} + b_{ij} \\ D = kC = k(A+B) & d_{ij} = kc_{ij} = k(a_{ij} + b_{ij}) = ka_{ij} + kb_{ij} \end{array}$$

Now, the right hand side

$$\begin{array}{r|l} E = kA & e_{ij} = ka_{ij} \\ F = kB & f_{ij} = kb_{ij} \\ G = E + F = kA + kB & g_{ij} = e_{ij} + f_{ij} = ka_{ij} + kb_{ij} \end{array}$$

It is obvious that $d_{ij} = g_{ij}$, and consequently $k(A+B) = kA + kB$. (q.e.d.)

Definition 1.5 (Multiplication of two matrices)

We define the ***multiplication operator between two matrices*** *as:*

$$\begin{array}{rlcl} \cdot : & \mathcal{M}_{m\times n}\times\mathcal{M}_{n\times p} & \rightarrow & \mathcal{M}_{m\times p} \\ & \cdot(A,B) & \rightarrow & C = AB \quad \left| c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \right. \end{array}$$

If we consider the different columns of B, then we have

$$B = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_p \end{pmatrix} \Rightarrow AB = \begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_p \end{pmatrix}$$

That can be interpreted as "the j-th column of AB is a weighted sum of the columns of matrix A using the weights defined by the j-th column of B".

Example

MATLAB: `A*B`

Matrix operations

Example

Let $A = \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}$ and $B = \begin{pmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{pmatrix}$. Then,

$$\begin{aligned}
A\mathbf{b}_1 &= \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 11 \\ -1 \end{pmatrix} \\
A\mathbf{b}_2 &= \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 13 \end{pmatrix} \\
A\mathbf{b}_3 &= \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix} \begin{pmatrix} 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 21 \\ -9 \end{pmatrix} \\
AB &= \begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{pmatrix} = \begin{pmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{pmatrix}
\end{aligned}$$

To directly compute a specific entry, for instance, $(AB)_{23}$ we have to multiply the 2nd row of A and the third column of B

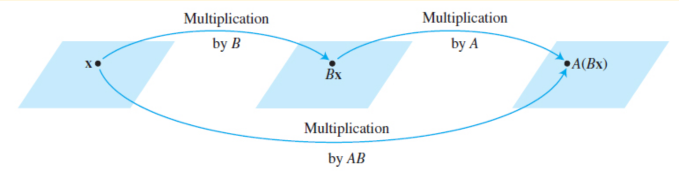
$$(AB)_{23} = \left[\begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix} \begin{pmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{pmatrix} \right] = 1 \cdot 6 + (-5) \cdot 3 = -9$$

Geometrical interpretation

Consider the linear transformations

$$T_A(\mathbf{x}) = A\mathbf{x}$$
$$T_B(\mathbf{x}) = B\mathbf{x}$$

that map any input vector using the matrix A or B, respectively. Now consider the sequential application of first T_B, and then T_A, as shown in the following figure:



Matrix multiplication helps us to define a single transformation such that we can transform the original vectors in a single step:

$$T_{AB}(\mathbf{x}) = (AB)\mathbf{x} = A(B\mathbf{x}) = T_A(T_B(\mathbf{x}))$$

Matrix operations

Property

$\mathrm{row}_i(AB) = \mathrm{row}_i(A)B$

Example (continued)

$$\mathrm{row}_1(AB) = \mathrm{row}_1(A)B = \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{pmatrix} = \begin{pmatrix} 11 & 0 & 21 \end{pmatrix}$$

More properties

$A(BC) = (AB)C$	Associativity
$A(B+C) = AB + AC$	Left distributivity
$(A+B)C = AC + BC$	Right distributivity
$r(AB) = (rA)B = A(rB)$	For any scalar r
$I_m A = A = A I_n$	For $A \in \mathcal{M}_{m\times n}$

Proof $A(BC)=(AB)C$

Let us consider the column decomposition of matrix C.

$$C=\begin{pmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_p\end{pmatrix}\Rightarrow$$
$$BC=\begin{pmatrix}B\mathbf{c}_1 & B\mathbf{c}_2 & \ldots & B\mathbf{c}_p\end{pmatrix}\Rightarrow$$
$$A(BC)=\begin{pmatrix}A(B\mathbf{c}_1) & A(B\mathbf{c}_2) & \ldots & A(B\mathbf{c}_p)\end{pmatrix}$$

But we have seen in the geometrical interpretation of matrix multiplication that $A(B\mathbf{c}_i)=(AB)\mathbf{c}_i$, therefore

$$A(BC)=\begin{pmatrix}(AB)\mathbf{c}_1 & (AB)\mathbf{c}_2 & \ldots & (AB)\mathbf{c}_p\end{pmatrix}=(AB)C$$

Warnings

- $AB\neq BA$, matrix multiplication is not commutative.
- $AB=AC\nRightarrow B=C$.
- $AB=0\nRightarrow B=0$ or $C=0$.

Matrix operations

Definition 1.6 (Power of a matrix)

If $A \in \mathcal{M}_{n\times n}$, then the k-th power of the matrix is defined as

$$A^k = \underbrace{A \cdot A \cdot \ldots \cdot A}_{k\ times}$$

Note: $A^0 = I_n$

Example

MATLAB: `A^k`

Matrix operations

Definition 1.7 (Transpose)

If $A \in \mathcal{M}_{m\times n}$, then the transpose of A (A^T) is a matrix in $\mathcal{M}_{n\times m}$ such that the rows of A are the columns of A^T, or more formally

$$(A^T)_{ij} = A_{ji}$$

Example

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

MATLAB: `A'`

Properties

$$(A^T)^T = A$$
$$(A+B)^T = A^T + B^T$$
$$(rA)^T = rA^T$$
$$(AB)^T = B^T A^T$$

Proof $(AB)^T = B^T A^T$

Let $A \in \mathcal{M}_{m\times n}$ and $B \in \mathcal{M}_{n\times p}$ By the definition of matrix multiplication we know that

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Let $B' = B^T$ and $A' = A^T$. For the same reason

$$(B^T A^T)_{ij} = (B'A')_{ij} = \sum_{k=1}^{n} b'_{ik} a'_{kj}$$

But $b'_{ik} = b_{ki}$ and $a'_{kj} = a_{jk}$, consequently

$$(B^T A^T)_{ij} = \sum_{k=1}^{n} b_{ki} a_{jk} = \sum_{k=1}^{n} a_{jk} b_{ki} = (AB)_{ji}$$

or what is the same

$$B^T A^T = (AB)^T$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 1:

- 2.1.3
- 2.1.10
- 2.1.12
- 2.1.18
- 2.1.19
- 2.1.20
- 2.1.22
- 2.1.23
- 2.1.24
- 2.1.25
- 2.1.26
- 2.1.27
- 2.1.39 (bring computer)
- 2.1.40 (bring computer)

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

Matrix inverse

Example

The inverse of a number is a clear concept

$$5\frac{1}{5} = 5 \cdot 5^{-1} = 1 = 5^{-1} \cdot 5$$

Definition 2.1 (Inverse of a matrix)

*A matrix $A \in \mathcal{M}_{n\times n}$ is **invertible** (or **non-singular**) if there exists another matrix $C \in \mathcal{M}_{n\times n}$ such that $AC = I_n = CA$. C is called the inverse of A and it is denoted as A^{-1}. If A is not invertible, it is said to be **singular**. (MATLAB: `inv(A)`)*

Properties

The inverse of a matrix is unique.

Proof

Let us assume that there exist two different inverses C_1 and C_2. Then,

$$C_2 = C_2 I = C_2(AC_1) = (C_2 A)C_1 = IC_1 = C_1$$

which is a contradiction and, therefore, the inverse must be unique. (q.e.d.)

Matrix inverse

Example

$$\text{Let } A = \begin{pmatrix} 2 & 5 \\ -3 & -7 \end{pmatrix} \text{ and } A^{-1} = \begin{pmatrix} -7 & -5 \\ 3 & 2 \end{pmatrix}$$

It can easily be verified that

$$AA^{-1} = A^{-1}A = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Theorem 2.1 (Inverse of a 2×2 matrix)

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *If* $ad - bc \neq 0$, *then* A *is invertible and its inverse is*

$$A^{-1} = \tfrac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Proof
It is easy to verify that $AA^{-1} = A^{-1}A = I_2$.

Matrix inverse

Theorem 2.2

If $A \in \mathcal{M}_{n\times n}$ is invertible, then for every $b \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution that is $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof

Proof $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution

If we substitute the solution in the equation we have

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = \mathbf{b} \text{ (q.e.d.)}$$

Proof $\mathbf{x} = A^{-1}\mathbf{b}$ is the unique solution

Let us assume that $\mathbf{x}' \neq \mathbf{x}$ is also a solution, then

$$A\mathbf{x}' = \mathbf{b}$$

If we multiply on the left by A^{-1}, we have

$$A^{-1}A\mathbf{x}' = A^{-1}\mathbf{b} \Rightarrow \mathbf{x}' = \mathbf{x}$$

which is obviously a contradiction and, therefore, $\mathbf{x} = A^{-1}\mathbf{b}$ must be the unique solution. (q.e.d.)

Matrix inverse

Example

EXAMPLE 3 A horizontal elastic beam is supported at each end and is subjected to forces at points 1, 2, 3, as shown in Fig. 1. Let $\mathbf{f}$ in $\mathbb{R}^3$ list the forces at these points, and let $\mathbf{y}$ in $\mathbb{R}^3$ list the amounts of deflection (that is, movement) of the beam at the three points. Using Hooke's law from physics, it can be shown that

$$\mathbf{y} = D\mathbf{f}$$

where D is a *flexibility matrix*. Its inverse is called the *stiffness matrix*. Describe the physical significance of the columns of D and D^{-1}.

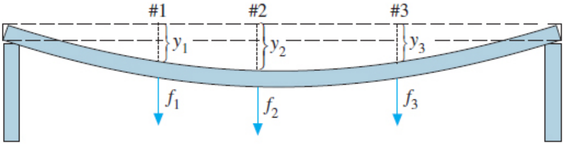


FIGURE 1 Deflection of an elastic beam.

Matrix inverse

Example (continued)

Consider the equation $\mathbf{y} = D\mathbf{f}$, $D = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 1 \end{pmatrix}$ and the fact that

$$D = DI = \begin{pmatrix} D\mathbf{e}_1 & D\mathbf{e}_2 & D\mathbf{e}_3 \end{pmatrix}$$

Therefore, the i-th column of D can be interpreted as the deflection at the different points when a unit force ($\mathbf{e}_i$) is applied onto the i-th point. In our example when we apply a unit force at point 1, the first column of D is $(1, \frac{1}{2}, \frac{1}{4})$ meaning that the first point displaces 1 m., the second point $\frac{1}{2}$ m., and the third point $\frac{1}{4}$ m.

Example (continued)

If we now consider that $\mathbf{f} = D^{-1}\mathbf{y}$, $D^{-1} = \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} & 0 \\ -\frac{2}{3} & \frac{4}{3} & -\frac{2}{3} \\ 0 & -\frac{2}{3} & \frac{4}{3} \end{pmatrix}$ and the fact that

$$D^{-1} = D^{-1}I = \begin{pmatrix} D^{-1}\mathbf{e}_1 & D^{-1}\mathbf{e}_2 & D^{-1}\mathbf{e}_3 \end{pmatrix}$$

Therefore, the i-th column of D^{-1} can be interpreted as the forces needed to be applied at the different points to produce a unit deformation ($\mathbf{e}_i$) at the i-th point. In our example, to produce a displacement of 1 m. in the first point and none at the other points ($\mathbf{e}_1 = (1, 0, 0)$, we need to push point 1 with a force of $\frac{4}{3}$ N., to pull point 2 with a force of $-\frac{2}{3}$ N., and we do not need to apply any force at point 3.

Matrix inverse

Theorem 2.3

1. *If A is invertible, then A^{-1} is also invertible and its inverse is A.*
2. *If A and B are invertible, then AB is also invertible and its inverse is $B^{-1}A^{-1}$*
3. *If A is invertible, then A^T is also invertible and its inverse is $(A^{-1})^T$.*

Proof 1)

The definition of A^{-1} is that it is a matrix such that

$$AA^{-1} = A^{-1}A = I$$

The inverse of A^{-1} must be a matrix C such that

$$CA^{-1} = A^{-1}C = I$$

If we compare this equation with the previous one, we easily see that $C = A$ is the inverse of A^{-1}.

Matrix inverse

Proof 2)

Let us check that $B^{-1}A^{-1}$ is actually the inverse of AB

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$

Proof 3)

Let us check that $(A^{-1})^T$ is actually the inverse of A^T

$$A^T(A^{-1})^T = [(AB)^T = B^TA^T] = (A^{-1}A)^T = I^T = I$$
$$(A^{-1})^TA^T = [(AB)^T = B^TA^T] = (AA^{-1})^T = I^T = I$$

Theorem 2.4

We may generalize the previous theorem and state that

$$(A_1A_2...A_p)^{-1} = A_p^{-1}A_{p-1}^{-1}...A_2^{-1}A_1^{-1}$$

Matrix inverse

Proof

Let's prove it by weak induction. That is, we know that the statement is true for $p = 2$ (by the previous theorem). Let us assume it is true for $p-1$, that is

$$(A_1 A_2 ... A_{p-1})^{-1} = A_{p-1}^{-1} ... A_2^{-1} A_1^{-1}$$

We wonder if it is also true for p. Let us define $B = A_1 A_2 ... A_{p-1}$. Then, we can rewrite the left hand side of the theorem as

$$(A_1 A_2 ... A_p)^{-1} = (B A_p)^{-1}$$

This is the inverse of the multiplication of two matrices. We know by the previous theorem that $(BA_p)^{-1} = A_p^{-1} B^{-1}$ But we presumed that

$$B^{-1} = (A_1 A_2 ... A_{p-1})^{-1} = A_{p-1}^{-1} ... A_2^{-1} A_1^{-1}$$

And consequently

$$(BA_p)^{-1} = A_p^{-1} A_{p-1}^{-1} ... A_2^{-1} A_1^{-1} \text{ (q.e.d.)}$$

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

Elementary matrices

The elementary operations we can perform on the rows of a matrix are

1. Multiply by a scalar
2. Swap two rows
3. Replace a row by a linear combination of two or several rows

All these operations can be represented as matrix multiplications.

Example

Consider the matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$

1. We can multiply the third row by a scalar r by multiplying on the left by the matrix

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & r \end{pmatrix} \Rightarrow E_1 A = \begin{pmatrix} a & b & c \\ d & e & f \\ rg & rh & ri \end{pmatrix}$$

Elementary matrices

Example (continued)

2. We can swap the first and second rows of the matrix by multiplying on the left by the matrix

$$E_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow E_2 A = \begin{pmatrix} d & e & f \\ a & b & c \\ g & h & i \end{pmatrix}$$

3. We can substitute the third row by $\mathbf{r}_3 + k_1 \mathbf{r}_1$ by multiplying on the left by the matrix

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k_1 & 0 & 1 \end{pmatrix} \Rightarrow E_3 A = \begin{pmatrix} a & b & c \\ d & e & f \\ g + k_1 a & h + k_1 b & i + k_1 c \end{pmatrix}$$

Definition 3.1 (Elementary matrix)

*An **elementary matrix** is one that differs from the identity matrix by one single, elementary row operation.*

Elementary matrices

Theorem 3.1

The inverse of an elementary matrix is another elementary matrix of the same type. That is, row operations can be undone.

Example (continued)

1. $E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{r} \end{pmatrix}$
2. $E_2^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
3. $E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -k_1 & 0 & 1 \end{pmatrix}$

MATLAB:

```
syms k1
E3=[1 0 0; 0 1 0; k1 0 1];
inv(E3)
```

Elementary matrices

Theorem 3.2

A matrix $A \in \mathcal{M}_{n\times n}$ is invertible iff it is row-equivalent to I_n. In this case, the sequence of operations that transforms A into I_n is also the one that transforms I_n into A^{-1}.

Proof $\Rightarrow$

If A is invertible, then by theorem 2.2 we know that the equation system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$. If it has a solution for every $\mathbf{b}$, then it must have a pivot in every row, that must be in the diagonal and, consequently the reduced echelon form of A must be I_n.

Proof $\Leftarrow$

If A is row-equivalent I_n, then there exists a sequence of elementary matrices that transform A into I_n

$$A \sim E_1A \sim E_2E_1A \sim ... \sim E_nE_{n-1}...E_2E_1A = I_n$$

$E = E_nE_{n-1}...E_2E_1$ is a candidate to be the inverse of A. Since each of the elementary matrices is invertible, and the product of invertible matrices is invertible, then E is invertible and A must be its (unique) inverse. Conversely, E is the inverse of A and A is invertible.

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

An algorithm to invert matrices

Algorithm

Algorithm: Reduce the augmented matrix $\left(\begin{array}{c|c} A & I \end{array}\right)$

If A is invertible, then $\left(\begin{array}{c|c} A & I \end{array}\right) \sim \left(\begin{array}{c|c} I & A^{-1} \end{array}\right)$.

If A is not invertible, then we will not be able to reduce A into I.

This algorithm is very much used in practice because it is numerically stable and rather efficient.

Example

Let $A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{pmatrix}$.

We construct the augmented matrix

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right)$$

Example (continued)

And now we transform it

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right)$$

$$\mathbf{r}_1 \leftrightarrow \mathbf{r}_2 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right)$$

$$\mathbf{r}_3 \leftarrow \mathbf{r}_3 - 4\mathbf{r}_1 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array}\right)$$

$$\mathbf{r}_3 \leftarrow \mathbf{r}_3 + 3\mathbf{r}_2 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right)$$

$$\mathbf{r}_3 \leftarrow \tfrac{1}{2}\mathbf{r}_3 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right)$$

An algorithm to invert matrices

Example (continued)

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right)$$

$$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - 2\mathbf{r}_3 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right)$$

$$\mathbf{r}_1 \leftarrow \mathbf{r}_1 - 3\mathbf{r}_3 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{array}\right)$$

Since A is row-equivalent to I_3, then A is invertible and its inverse is $A^{-1} = \begin{pmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{pmatrix}$. To finalize the exercise we should check that

$$AA^{-1} = A^{-1}A = I_3$$

An algorithm to invert matrices

A new interpretation of matrix inversion

By constructing the augmented matrix $\left(\begin{array}{c|c} A & I \end{array}\right)$ we are simultaneously solving multiple equation systems

$$A\mathbf{x} = \mathbf{e}_1 \quad A\mathbf{x} = \mathbf{e}_2 \quad A\mathbf{x} = \mathbf{e}_3 \quad \ldots$$

Example (continued)

$$\left(\begin{array}{ccc|c} 0 & 1 & 2 & 1 \\ 1 & 0 & 3 & 0 \\ 4 & -3 & 8 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} 0 & 1 & 2 & 0 \\ 1 & 0 & 3 & 1 \\ 4 & -3 & 8 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} 0 & 1 & 2 & 0 \\ 1 & 0 & 3 & 0 \\ 4 & -3 & 8 & 1 \end{array}\right)$$

This note is important because if we want to compute only the i-th column of A^{-1} it is enough to solve the equation system

$$A\mathbf{x} = \mathbf{e}_i$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 2:

- 2.2.7
- 2.2.11
- 2.2.13
- 2.2.17
- 2.2.19
- 2.2.21
- 2.2.25
- 2.2.36

Outline

3 Matrix algebra

- Matrix operations (a)
- Inverse of a matrix (b)
- Elementary matrices (b)
- An algorithm to invert matrices (b)
- Characterization of invertible matrices (c)
- Invertible linear transformations (c)
- Partitioned matrices (c)
- LU factorization (d)
- An application to computer graphics and image processing (d)
- Subspaces of $\mathbb{R}^n$ (e)
- Dimension and rank (e)

Characterization of invertible matrices

Theorem 5.1 (The invertible matrix theorem)

Let $A \in \mathcal{M}_{n\times n}$. The following statements are equivalent (either they are all true or they are all false):

i. *A is invertible.*

ii. *A is row-equivalent to I_n.*

iii. *A has n pivot positions.*

iv. *$A\mathbf{x} = \mathbf{0}$ only has the trivial solution $\mathbf{x} = \mathbf{0}$.*

v. *The columns of A are linearly independent.*

vi. *The transformation $T(\mathbf{x}) = A\mathbf{x}$ is injective.*

vii. *The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^n$.*

viii. *The columns of A span $\mathbb{R}^n$.*

ix. *The transformation $T(\mathbf{x}) = A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.*

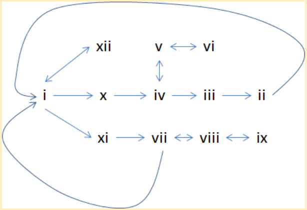
x. *There exists a matrix $C \in \mathcal{M}_{n\times n}$ such that $CA = I_n$.*

xi. *There exists a matrix $D \in \mathcal{M}_{n\times n}$ such that $AD = I_n$.*

xii. *A^T is an invertible matrix*

Characterization of invertible matrices

To prove the theorem we will follow the reasoning scheme below:



Proof $i \Rightarrow x$

If i is true, then x is true simply by doing $C = A^{-1}$.

Proof $x \Rightarrow iv$

See Exercise 2.1.23 in Lay.

Proof $iv \Rightarrow iii$

See Exercise 2.2.23 in Lay.

Proof $iii \Rightarrow ii$

If iii is true, then the n pivots have to be in the main diagonal and in this case, the reduced echelon form must be I_n.

Characterization of invertible matrices

Proof ii $\Rightarrow$ i

If ii is true, then i is true thanks to Theorem 3.2.

Proof i $\Rightarrow$ xi

If i is true, then xi is true simply by doing $D = A^{-1}$.

Proof xi $\Rightarrow$ vii

See Exercise 2.1.24 in Lay.

Proof vii $\Rightarrow$ i

See Exercise 2.2.24 in Lay.

Proof vii $\Leftrightarrow$ viii $\Leftrightarrow$ ix

See Theorems 3.2 and 8.2 in Chapter 2.

Proof iv $\Leftrightarrow$ v $\Leftrightarrow$ vi

See Theorems 3.2, 5.1 and 8.1 in Chapter 2.

Proof i $\Rightarrow$ xii

See Theorem 2.3.

Proof i $\Leftarrow$ xii

See Theorem 2.3 interchanging the roles of A and A^T.

The power of this theorem is that it connects equation systems to invertibility, linear independence and subspace bases.

Characterization of invertible matrices

Corollary

1. If A is invertible, then $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.
2. If $A, B \in \mathcal{M}_{n \times n}$ and $AB = I_n$, then A and B are invertible and $B = A^{-1}$ and $A = B^{-1}$.

Watch out that this corollary only applies to square matrices.

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

Invertible linear transformations

Consider the linear transformation

$$\begin{array}{rcl} T:\mathbb{R}^n & \rightarrow & \mathbb{R}^n \\ \mathbf{x} & \rightarrow & A\mathbf{x} \end{array}$$

Definition 6.1 (Inverse transformation)

T is invertible iff there exists $S:\mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $\forall \mathbf{x} \in \mathbb{R}^n$:

$$S(T(\mathbf{x})) = \mathbf{x} = T(S(\mathbf{x}))$$

Example

$T(\mathbf{x}) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}$ is invertible and its inverse is $S(\mathbf{x}) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}$.

Proof

$$\begin{aligned} S(T(\mathbf{x})) &= S\left(\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{x}\right) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{x} = \mathbf{x} \\ T(S(\mathbf{x})) &= T\left(\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{x}\right) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{x} = \mathbf{x} \end{aligned}$$

Invertible linear transformations

Example

$T(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{x}$ is not invertible because $T((1,0)) = T((1,1)) = (1,0)$, so given the "output" (1,0), we cannot recover the input vector that originated this output.

Theorem 6.1

If T is invertible, then it is surjective.

<u>Proof</u>

Consider any vector $\mathbf{b} \in \mathbb{R}^n$, we can always apply the transformation S to get a new vector $\mathbf{x} = S(\mathbf{b})$. And then, recover $\mathbf{b}$ making use of the fact that T is the inverse of S, that is, $\mathbf{b} = T(\mathbf{x})$. In other words, any vector $\mathbf{b}$ is in the range of T and, therefore, T is surjective.

Invertible linear transformations

Theorem 6.2

T is invertible iff A is invertible. If T is invertible, then the only function that satisfies the previous definition is

$$S(\mathbf{x}) = A^{-1}\mathbf{x}$$

Proof $\Rightarrow$

If T is invertible, then it is surjective (see previous Theorem). Then, A is invertible by Theorem 5.1 (items i and ix).

Proof $\Leftarrow$

If A is invertible, then we may construct the linear transformation $S = A^{-1}\mathbf{x}$. S is an inverse of T since

$$\begin{gathered} S(T(\mathbf{x})) = S(A\mathbf{x}) = A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = \mathbf{x} \\ T(S(\mathbf{x})) = T(A^{-1}\mathbf{x}) = A(A^{-1}\mathbf{x}) = (AA^{-1})\mathbf{x} = \mathbf{x} \end{gathered}$$

Invertible linear transformations

Proof uniqueness

Let us assume that there are two inverses $S_1(\mathbf{x}) = B_1\mathbf{x}$ and $S_2(\mathbf{x}) = B_2\mathbf{x}$ with $B_1 \neq B_2$. Let $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{v} = T(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^n$ (since T is invertible and, therefore, surjective, we are guaranteed that there exists at least one such $\mathbf{x}$). Now

$$\left.\begin{array}{l} S_1(\mathbf{v}) = B_1 A\mathbf{x} = \mathbf{x} = B_1\mathbf{v} \\ S_2(\mathbf{v}) = B_2 A\mathbf{x} = \mathbf{x} = B_2\mathbf{v} \end{array}\right\} \Rightarrow B_1\mathbf{v} = B_2\mathbf{v} \; [\forall \mathbf{v} \in \mathbb{R}^n] \; \Rightarrow B_1 = B_2$$

which is a contradiction and, consequently, there exists only one inverse (q.e.d.)

Definition 6.2 (Ill-conditioned matrix)

Informally, we say that a matrix A is ***ill-conditioned*** *if it is "nearly singular". In practice, this implies that the equation system $A\mathbf{x} = \mathbf{b}$ may have large variations in the solution ($\mathbf{x}$) when $\mathbf{b}$ varies slightly.*

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 3:

- 2.3.13
- 2.3.16
- 2.3.17
- 2.3.33
- 2.3.41

Outline

3 Matrix algebra

Partitioned matrices

Partitioned matrices sometimes help us to gain insight into the structure of the problem by identifying blocks within the matrix.

Example

$$A = \left(\begin{array}{ccc|cc|c} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right) = \left(\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \end{array}\right)$$

$A \in \mathcal{M}_{3\times 6}$,
$A_{11} \in \mathcal{M}_{2\times 3}$, $A_{12} \in \mathcal{M}_{2\times 2}$, $A_{13} \in \mathcal{M}_{2\times 1}$,
$A_{21} \in \mathcal{M}_{1\times 3}$, $A_{22} \in \mathcal{M}_{1\times 2}$, $A_{23} \in \mathcal{M}_{1\times 1}$.
MATLAB:

```
 A=[3 0 -1 5 9 -2; -5 2 4 0 -3 1; -8 -6 3 1 7 -4];
A11=A(1:2,1:3)
A12=A(1:2,4:5)
A13=A(1:2,6)
A21=A(3,1:3)
A22=A(3,4:5)
A23=A(3,6)
```

Partitioned matrices

Definition 7.1 (Sum of partitioned matrices)

Let A and B be two matrices partitioned in the same way. Then the blocks of $A+B$ are simply the sum of the corresponding blocks.

$$A+B=\left(\begin{array}{c|c|c} & & \\ \hline & A_{ij} & \\ \hline & & \end{array}\right)+\left(\begin{array}{c|c|c} & & \\ \hline & B_{ij} & \\ \hline & & \end{array}\right)=\left(\begin{array}{c|c|c} & & \\ \hline & A_{ij}+B_{ij} & \\ \hline & & \end{array}\right)$$

Definition 7.2 (Multiplication by scalar)

The multiplication by a scalar simply multiplies each one of the blocks independently

$$rA=r\left(\begin{array}{c|c|c} & & \\ \hline & A_{ij} & \\ \hline & & \end{array}\right)=\left(\begin{array}{c|c|c} & & \\ \hline & rA_{ij} & \\ \hline & & \end{array}\right)$$

Partitioned matrices

Definition 7.3 (Multiplication of partitioned matrices)

Multiply the different block as if they were scalars (but applying matrix multiplication).

Example

Let $A = \left(\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right) = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right)$

and $B = \left(\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right) = \left(\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right)$.

Then, $AB = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{pmatrix} = \begin{pmatrix} -5 & 4 \\ -6 & 2 \\ 2 & 1 \end{pmatrix}$

Partitioned matrices

Theorem 7.1 (Multiplication of matrices)

Let $A \in \mathcal{M}_{m\times n}$ *and* $B \in \mathcal{M}_{n\times p}$*, then*

$$AB = \sum_{k=1}^{n} \mathrm{column}_k(A)\mathrm{row}_k(B)$$

<u>*Proof*</u>

Let us analyze each one of the terms in the sum

$$\mathrm{column}_k(A)\mathrm{row}_k(B) = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \ldots \\ a_{mk} \end{pmatrix} \begin{pmatrix} b_{k1} & b_{k2} & \ldots & b_{kp} \end{pmatrix} =$$

$$\begin{pmatrix} a_{1k}b_{k1} & a_{1k}b_{k2} & \ldots & a_{1k}b_{kp} \\ a_{2k}b_{k1} & a_{2k}b_{k2} & \ldots & a_{2k}b_{kp} \\ \ldots & \ldots & \ldots & \ldots \\ a_{mk}b_{k1} & a_{mk}b_{k2} & \ldots & a_{mk}b_{kp} \end{pmatrix}$$

Partitioned matrices

In general, the *ij*-th term is

$$(\mathrm{column}_k(A)\mathrm{row}_k(B))_{ij} = a_{ik}b_{kj}$$

If we now analyze the *ij*-th element of the sum

$$\left(\sum_{k=1}^{n}\mathrm{column}_k(A)\mathrm{row}_k(B)\right)_{ij} = \sum_{k=1}^{n}(\mathrm{column}_k(A)\mathrm{row}_k(B))_{ij} = \sum_{k=1}^{n}a_{ik}b_{kj}$$

But this is the definition of matrix multiplication and, therefore,

$$\left(\sum_{k=1}^{n}\mathrm{column}_k(A)\mathrm{row}_k(B)\right)_{ij} = (AB)_{ij} \text{ (q.e.d.)}$$

Partitioned matrices

Definition 7.4 (Transpose of partitioned matrices)

Transpose the partitioned matrix as if it were composed of scalars, and transpose each one of the blocks.

Example

$$A = \left(\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array}\right) \Rightarrow A^T = \left(\begin{array}{c|c|c} A_{11}^T & A_{21}^T & A_{31}^T \\ \hline A_{12}^T & A_{22}^T & A_{32}^T \\ \hline A_{13}^T & A_{23}^T & A_{33}^T \end{array}\right)$$

Example

$$A = \left(\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right) \Rightarrow A^T = \left(\begin{array}{cc|c} 2 & 1 & 0 \\ -3 & 5 & -4 \\ 1 & -2 & -2 \\ \hline 0 & 3 & 7 \\ -4 & -1 & -1 \end{array}\right)$$

Partitioned matrices

Definition 7.5 (Inverse of partitioned matrices)

The formula for each one of the cases is worked out particularly for that case. Here go a couple of examples.

Example

Let $A = \left(\begin{array}{c|c|c} A_{11} & 0 & 0 \\ \hline 0 & A_{22} & 0 \\ \hline 0 & 0 & A_{33} \end{array}\right)$.

$A \in \mathcal{M}_{n\times n}$, $A_{11} \in \mathcal{M}_{p\times p}$, $A_{22} \in \mathcal{M}_{q\times q}$, $A_{33} \in \mathcal{M}_{r\times r}$ such that $p+q+r=n$.
We look for a matrix B such that

$$\left(\begin{array}{c|c|c} A_{11} & 0 & 0 \\ \hline 0 & A_{22} & 0 \\ \hline 0 & 0 & A_{33} \end{array}\right) \left(\begin{array}{c|c|c} B_{11} & B_{12} & B_{13} \\ \hline B_{21} & B_{22} & B_{23} \\ \hline B_{31} & B_{32} & B_{33} \end{array}\right) = \left(\begin{array}{c|c|c} I_p & 0 & 0 \\ \hline 0 & I_q & 0 \\ \hline 0 & 0 & I_r \end{array}\right) \Rightarrow$$

$$\left(\begin{array}{c|c|c} A_{11}B_{11} & A_{11}B_{12} & A_{11}B_{13} \\ \hline A_{22}B_{21} & A_{22}B_{22} & A_{22}B_{23} \\ \hline A_{33}B_{31} & A_{33}B_{32} & A_{33}B_{33} \end{array}\right) = \left(\begin{array}{c|c|c} I_p & 0 & 0 \\ \hline 0 & I_q & 0 \\ \hline 0 & 0 & I_r \end{array}\right)$$

Partitioned matrices

Example (continued)

For each one of the entries we have a set of equations:

$$\begin{aligned}
&\forall A_{11} \in \mathcal{M}_{p\times p}\ A_{11}B_{11} = I_p \Rightarrow B_{11} = A_{11}^{-1} \\
&\forall A_{11} \in \mathcal{M}_{p\times p}\ A_{11}B_{12} = 0 \Rightarrow B_{12} = 0 \\
&\forall A_{11} \in \mathcal{M}_{p\times p}\ A_{11}B_{13} = 0 \Rightarrow B_{13} = 0 \\
&\forall A_{22} \in \mathcal{M}_{q\times q}\ A_{22}B_{21} = 0 \Rightarrow B_{21} = 0 \\
&\forall A_{22} \in \mathcal{M}_{q\times q}\ A_{22}B_{22} = I_q \Rightarrow B_{22} = A_{22}^{-1} \\
&\forall A_{22} \in \mathcal{M}_{q\times q}\ A_{22}B_{23} = 0 \Rightarrow B_{23} = 0 \\
&\forall A_{33} \in \mathcal{M}_{r\times r}\ A_{33}B_{31} = 0 \Rightarrow B_{31} = 0 \\
&\forall A_{33} \in \mathcal{M}_{r\times r}\ A_{33}B_{32} = 0 \Rightarrow B_{32} = 0 \\
&\forall A_{33} \in \mathcal{M}_{r\times r}\ A_{33}B_{33} = I_r \Rightarrow B_{33} = A_{33}^{-1}
\end{aligned}$$

Finally,

$$B = \left(\begin{array}{c|c|c} A_{11}^{-1} & 0 & 0 \\ \hline 0 & A_{22}^{-1} & 0 \\ \hline 0 & 0 & A_{33}^{-1} \end{array}\right)$$

Partitioned matrices

Example

Let $A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right)$.

$A \in \mathcal{M}_{n\times n}$, $A_{11} \in \mathcal{M}_{p\times p}$, $A_{12} \in \mathcal{M}_{p\times q}$, $A_{22} \in \mathcal{M}_{q\times q}$ such that $p+q=n$.
We look for a matrix B such that

$$= \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right) \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right) = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & I_q \end{array}\right) \Rightarrow$$

$$\left(\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{22}B_{21} & A_{22}B_{22} \end{array}\right) = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & I_q \end{array}\right)$$

Partitioned matrices

Example (continued)

For each one of the entries we have a set of equations:

$$
\begin{array}{ll}
\forall A_{22} \in \mathcal{M}_{q\times q} \; A_{22}B_{21} = 0 \Rightarrow B_{21} = 0 & \\
\forall A_{22} \in \mathcal{M}_{q\times q} \; A_{22}B_{22} = I_q \Rightarrow B_{22} = A_{22}^{-1} & \\
\quad \forall A_{11} \in \mathcal{M}_{q\times q}, A_{12} \in \mathcal{M}_{p\times q} & A_{11}B_{11} + A_{12}B_{21} = I_p \Rightarrow [B_{21} = 0] \Rightarrow \\
 & A_{11}B_{11} = I_p \Rightarrow B_{11} = A_{11}^{-1} \\
\quad \forall A_{11} \in \mathcal{M}_{q\times q}, A_{12} \in \mathcal{M}_{p\times q} & A_{11}B_{12} + A_{12}B_{22} = 0 \Rightarrow [B_{22} = A_{22}^{-1}] \Rightarrow \\
 & A_{11}B_{12} + A_{12}A_{22}^{-1} = 0 \Rightarrow A_{11}B_{12} = -A_{12}A_{22}^{-1} \Rightarrow \\
 & B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}
\end{array}
$$

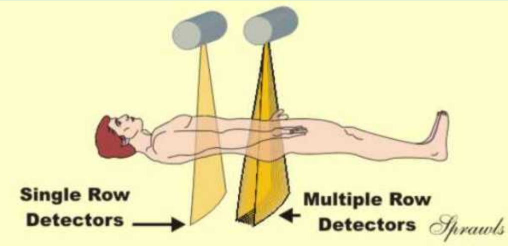
Finally,

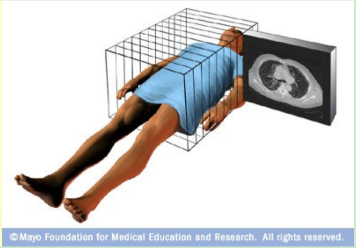
$$
B = \left(\begin{array}{c|c} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ \hline 0 & A_{22}^{-1} \end{array} \right)
$$

Partitioned matrices

Example

Computational Tomography (CT) with multiple rows gives a non-block structure for the system matrix that forces the problem to be solved in 3D. However, with a single row detector, the system matrix has a block structure so that the problem can be solved as a series of 2D problems strongly accelerating the process (on the other side the redundancy introduced by multiple row offers better resolution and robustness to noise).





Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 4:

- 2.4.15
- 2.4.16
- 2.4.18
- 2.4.19

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

LU factorization

Example

Let us presume that we have a collection of equation systems

$$A\mathbf{x} = \mathbf{b}_1$$
$$A\mathbf{x} = \mathbf{b}_2$$
$$\ldots$$

and A is not invertible, which could be an efficient way of solving all of them together? Factorize A as $A = LU$ (see below) and solve the equation system in two steps. In fact the method is so efficient it is even used to solve a single equation system.

LU factorization

Definition 8.1 (LU factorization)

Let $A \in \mathcal{M}_{m\times n}$ that can be reduced to a reduced echelon form without row permutations. We can factorize A as $A = LU$, where L is an invertible, lower triangular matrix (with 1s in the main diagonal) of size $m \times m$ and U is an upper triangular matrix of size $m \times n$.
MATLAB: `[L,U]=lu(A)`

Example

Let $A \in \mathcal{M}_{4\times 5}$. LU factorization will produce two matrices L and U may be of the following structure

$$A = LU = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \heartsuit & 1 & 0 & 0 \\ \heartsuit & \heartsuit & 1 & 0 \\ \heartsuit & \heartsuit & \heartsuit & 1 \end{pmatrix} \begin{pmatrix} \diamondsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ 0 & \diamondsuit & \heartsuit & \heartsuit & \heartsuit \\ 0 & 0 & 0 & \diamondsuit & \heartsuit \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

LU factorization

Solving a linear equation system using the LU decomposition

Consider the equation system $A\mathbf{x} = \mathbf{b}$, and assume we have decomposed A as $A = LU$. Then, we can solve the equation system in two steps:

$$A\mathbf{x} = \mathbf{b} \Rightarrow (LU)\mathbf{x} = L(U\mathbf{x}) = \mathbf{b} \Rightarrow \begin{cases} L\mathbf{y} = \mathbf{b} \\ U\mathbf{x} = \mathbf{y} \end{cases}$$

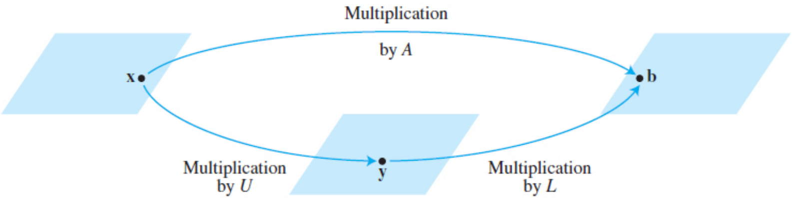


FIGURE 2 Factorization of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

LU factorization

Example

Consider

$$A=\begin{pmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{pmatrix}\begin{pmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and $\mathbf{b}=(-9,5,7,11)$. We first solve $L\mathbf{y}=\mathbf{b}$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{array}\right)\sim\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

and now we solve $U\mathbf{x}=\mathbf{y}$

$$\left(\begin{array}{cccc|c} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right)\sim\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

The trick is that, thanks to the triangular structure, solving these two equation systems is rather fast.

An algorithm to simple LU factorizations

Algorithm

Let us assume that A is row-equivalent to U only using row replacement only with the rows above the replaced row. Then, there must be a sequence of elementary matrices such that

$$A \sim U \Rightarrow E_p ... E_2 E_1 A = U \Rightarrow A = (E_p ... E_2 E_1)^{-1} U$$

By inspection, we note that $L = (E_p ... E_2 E_1)^{-1}$.

In the previous algorithm we are making using of the following theorem:

Theorem 8.1

1. *The product of two lower triangular matrices is lower triangular.*
2. *The inverse of a lower triangular matrix is lower triangular.*

An algorithm to simple LU factorizations

Example

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

$$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - \tfrac{1}{2}\mathbf{r}_1 \quad E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\mathbf{r}_3 \leftarrow \mathbf{r}_3 - \tfrac{2}{3}\mathbf{r}_2 \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{pmatrix} \quad U = \begin{pmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix}$$

Now, we calculate L as

$$L = (E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix}$$

An algorithm to simple LU factorizations

Example

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

$$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - \tfrac{1}{2}\mathbf{r}_1 \quad E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\mathbf{r}_3 \leftarrow \mathbf{r}_3 - \tfrac{2}{3}\mathbf{r}_2 \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{pmatrix} \quad U = \begin{pmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix}$$

Now, we calculate L as

$$L = (E_2E_1)^{-1} = E_1^{-1}E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix}$$

LDU factorization

Example (continued)

Note that the L and U matrices found so far are assymetric in the sense that L has 1s in its main diagonal, but U has not. We can extract the elements in the main diagonal of U to a separate matrix D by simply dividing the corresponding row of U by that element:

$$\begin{aligned} A &= LU = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \\ &= LDU = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

where D is always a diagonal matrix.

Other factorization examples

Other factorizations

There are many other possibilities to factorize a matrix $A \in \mathcal{M}_{m\times n}$. See http://en.wikipedia.org/wiki/Matrix_decomposition. Among the most important are:

- **QR:** $A = QR$ where $Q \in \mathcal{M}_{m\times m}$ is orthogonal ($Q^tQ = D$) and $R \in \mathcal{M}_{m\times n}$ is upper triangular.
- **SVD:** $A = UDV^t$ where $U \in \mathcal{M}_{m\times m}$ is unitary ($U^tU = I_m$), $D \in \mathcal{M}_{m\times n}$ is diagonal, and $V \in \mathcal{M}_{n\times n}$ is also unitary ($V^tV = I_n$).
- **Spectral:** $A = PDP^{-1}$ (only for square matrices) where $P \in \mathcal{M}_{n\times n}$ and $D \in \mathcal{M}_{n\times n}$ is diagonal.

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 5:

- 2.5.9
- 2.5.Practice problem

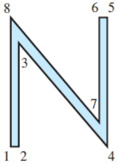
Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

An application to computer graphics and image processing

Example

In vectorial graphics, graphics are described as a set of connected points (whose coordinates are known).

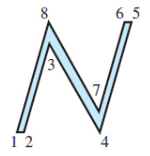


EXAMPLE 1 The capital letter N in Fig. 1 is determined by eight points, or *vertices*. The coordinates of the points can be stored in a data matrix, D.

$$\begin{array}{r} \\ \\ x\text{-coordinate} \\ y\text{-coordinate} \end{array} \begin{array}{c} \text{Vertex:} \\ \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0 & .5 & .5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix} \end{array} = D$$

In addition to D, it is necessary to specify which vertices are connected by lines, but we omit this detail. ■

We may produce "italic" fonts by shearing the standard coordinates $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{pmatrix} 1 & 0.25 \\ 0 & 1 \end{pmatrix}$.



An application to computer graphics and image processing

Example

Coordinate translations can be expressed as $T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$. But this is not a linear transformation:

$$\begin{aligned} T(\mathbf{u}) &= \mathbf{u} + \mathbf{x}_0 \\ T(\mathbf{v}) &= \mathbf{v} + \mathbf{x}_0 \\ T(\mathbf{u}+\mathbf{v}) &= \mathbf{u} + \mathbf{v} + \mathbf{x}_0 \\ T(\mathbf{u}) + T(\mathbf{v}) &= (\mathbf{u} + \mathbf{x}_0) + (\mathbf{v} + \mathbf{x}_0) = \mathbf{u} + \mathbf{v} + 2\mathbf{x}_0 \\ T(\mathbf{u}+\mathbf{v}) &\neq T(\mathbf{u}) + T(\mathbf{v}) \end{aligned}$$

We can solve this problem with homogeneous coordinates.

An application to computer graphics and image processing

Definition 9.1 (Homogeneous coordinates)

Given a point with coordinates $\mathbf{x}$ *we can construct its* ***homogeneous coordinates*** *as*

$$\tilde{\mathbf{x}} = \begin{pmatrix} h\mathbf{x} \\ h \end{pmatrix}$$

Or in other words, given the homogeneous coordinates $\tilde{\mathbf{u}} = \begin{pmatrix} \mathbf{u} \\ h \end{pmatrix}$*, they represent the point at* $\frac{\mathbf{u}}{h}$*. It is customary to use* $h = 1$ *(but it is not compulsory, and in certain applications it is better to use other* h*'s).*

Example

The 2D point $(1, 2)$ can be represented in homogeneous coordinates as $(1, 2, 1)$, as $(2, 4, 2)$ and, even, as $(-2, -4, -2)$. They all represent the same point.

An application to computer graphics and image processing

Example

Now, coordinate translations in homogeneous coordinates is a linear transformation. For instance, in 2D:

$$T(\tilde{\mathbf{x}}) = A\tilde{\mathbf{x}} = \begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + \Delta x \\ y + \Delta y \\ 1 \end{pmatrix}$$

2D transformations in homogeneous coordinates

In general, any 2D transformation of the form $T(\mathbf{x}) = A\mathbf{x}$ can be represented in homogeneous coordinates as

$$T(\tilde{\mathbf{x}}) = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \tilde{\mathbf{x}}$$

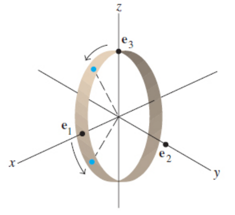
Example

An application in 3D graphics:
http://www.youtube.com/watch?v=EsNmiiKlRXQ

Example

Let's say we want to

1. Rotate a point 30°about the Y axis.



2. then, translate by $(-6, 4, 5)$

An application to computer graphics and image processing

Example (continued)

We need to use the transformation $T(\tilde{\mathbf{x}}) = \tilde{A}\tilde{\mathbf{x}}$ with

$$\tilde{A} = \begin{pmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(30^\circ) & 0 & \sin(30^\circ) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(30^\circ) & 0 & \cos(30^\circ) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\tilde{x} = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

An application to computer graphics and image processing

Example

Let's say we want to produce perspective projections. Let's imagine that the screen is on the XY plane and the viewer's eye is at $(0, 0, d)$ (the distance to the screen is d). Any object between the viewer and the screen is projected onto the screen as in the figure below

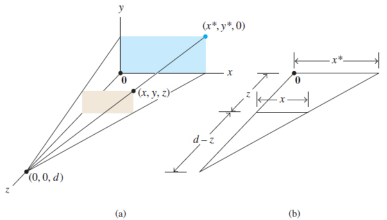


FIGURE 6 Perspective projection of (x, y, z) onto $(x^*, y^*, 0)$.

By similar triangles we have

$$\tan\alpha = \frac{x^*}{d} = \frac{x}{d-z} \Rightarrow x^* = \frac{x}{1-\frac{z}{d}}$$

An application to computer graphics and image processing

Example (continued)

Similarly, $y^* = \frac{y}{1-\frac{z}{d}}$. Using homogeneous coordinates we want that $(x, y, z, 1)$ maps onto $\left(\frac{x}{1-\frac{z}{d}}, \frac{y}{1-\frac{z}{d}}, 0, 1\right)$, or what is the same $\left(x, y, 0, 1 - \frac{z}{d}\right)$. We can achieve this with the perspective transformation:

$$\tilde{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{d} & 1 \end{pmatrix}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 7:

- 2.7.2
- 2.7.3
- 2.7.10
- 2.7.12
- 2.7.22

Outline

Subspace

Definition 10.1 (Subspace of $\mathbb{R}^n$)

$H \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ if:

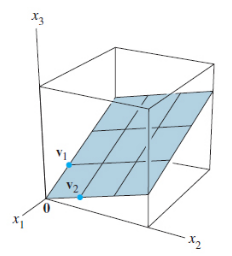
1. $\mathbf{0} \in H$
2. $\forall \mathbf{u}, \mathbf{v} \in H \quad \mathbf{u} + \mathbf{v} \in H$ *(H is closed under vector addition)*
3. $\forall \mathbf{u} \in H \, \forall r \in \mathbb{R} \quad r\mathbf{u} \in H$ *(H is closed under multiplication by a scalar)*

Example: Special subspaces

The following two sets are subspaces of $\mathbb{R}^n$:

1. $H = \{\mathbf{0}\}$
2. $H = \mathbb{R}^n$

Example: Plane



A plane is defined as

$H = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \{\mathbf{v} \in \mathbb{R}^n | \mathbf{v} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2\}$

This plane is a subspace of $\mathbb{R}^3$

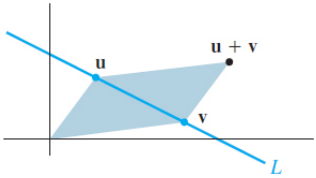
Proof

1. Proof $\mathbf{0} \in H$
   If $\lambda_1 = \lambda_2 = 0$, then $\mathbf{v} = \mathbf{0}$.
2. Proof $\mathbf{u} + \mathbf{v} \in H$
   $\mathbf{u} \in H \Rightarrow \mathbf{u} = \lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2$
   $\mathbf{v} \in H \Rightarrow \mathbf{v} = \lambda_{1v}\mathbf{v}_1 + \lambda_{2v}\mathbf{v}_2$
   $$\begin{aligned} \mathbf{u}+\mathbf{v} &= (\lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2) + (\lambda_{1v}\mathbf{v}_1 + \lambda_{2v}\mathbf{v}_2) \\ &= (\lambda_{1u} + \lambda_{1v})\mathbf{v}_1 + (\lambda_{2u} + \lambda_{2v})\mathbf{v}_2 \in H \end{aligned}$$
3. Proof $r\mathbf{u} \in H$
   $\mathbf{u} \in H \Rightarrow \mathbf{u} = \lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2$
   $$\begin{aligned} r\mathbf{u} &= r(\lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2) \\ &= r\lambda_{1u}\mathbf{v}_1 + r\lambda_{2u}\mathbf{v}_2 \in H \end{aligned}$$

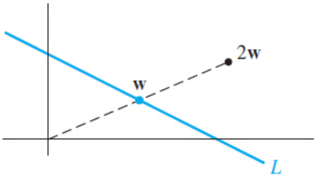
Example: Line not through the origin

A line (L) that does not pass through the origin is not a subspace, because

1. $\mathbf{0} \notin L$
2. If we take two points belonging to the line ($\mathbf{u}$ and $\mathbf{v}$), $\mathbf{u} + \mathbf{v} \notin L$.
3. If we take a point belonging to the line ($\mathbf{w}$), $2\mathbf{w} \notin L$.



u + v is not on L



2w is not on L

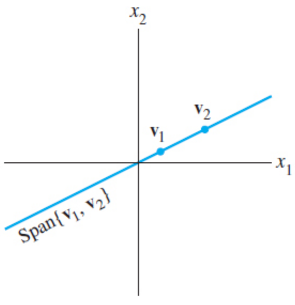
Subspace

Example: Line through the origin

Consider $\mathbf{v}_1$ and $\mathbf{v}_2 = k\mathbf{v}_1$. Then,

$$H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \text{Span}\{\mathbf{v}_1\}$$

is a line. It is easy to prove that this line is a subspace of $\mathbb{R}^n$.



$\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2 = k\mathbf{v}_1$.

Column space

Definition 10.2 (Column space of a matrix)

*Let $A \in \mathcal{M}_{m\times n}$. Let $\mathbf{a}_i \in \mathbb{R}^m$ be the columns of A. The **column space** of A is defined as*

$$\mathrm{Col}\{A\} = \mathrm{Span}\left\{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n\right\} \subseteq \mathbb{R}^m$$

Theorem 10.1

$\mathrm{Col}\{A\}$ *is a subspace of* $\mathbb{R}^m$.

Column space

Example

Let $A = \begin{pmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 3 \\ 3 \\ -4 \end{pmatrix}$.

Determine if $\mathbf{b}$ belongs to $\mathrm{Col}\{A\}$.

Solution

If $\mathbf{b} \in \mathrm{Col}\{A\}$ there must be some coefficients x_1, x_2 and x_3 such that

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$$

To find these coefficients we simply have to solve the equation system $A\mathbf{x} = \mathbf{b}$.

$$\left(\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ -4 & 6 & -2 & 3 \\ -3 & 7 & 6 & -4 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

In fact, there are infinite solutions to the equation system and, consequently, $\mathbf{b} \in \mathrm{Col}\{A\}$.

Null space

Definition 10.3 (Null space of a matrix)

*Let $A \in \mathcal{M}_{m\times n}$. The **null space** of A is defined as*

$$\mathrm{Nul}\{A\} = \{\mathbf{v} \in \mathbb{R}^n | A\mathbf{v} = \mathbf{0}\}$$

Theorem 10.2

$\mathrm{Nul}\{A\}$ *is a subspace of* $\mathbb{R}^n$.

Proof

1. *Proof* $\mathbf{0} \in \mathrm{Nul}\{A\}$
   $A\mathbf{0} = \mathbf{0} \Rightarrow \mathbf{0} \in \mathrm{Nul}\{A\}$ *(q.e.d.)*
2. *Proof* $\mathbf{u} + \mathbf{v} \in \mathrm{Nul}\{A\}$
   $\mathbf{u} \in \mathrm{Nul}\{A\} \Rightarrow A\mathbf{u} = \mathbf{0}$
   $\mathbf{v} \in \mathrm{Nul}\{A\} \Rightarrow A\mathbf{v} = \mathbf{0}$
   $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0} \Rightarrow \mathbf{u} + \mathbf{v} \in \mathrm{Nul}\{A\}$ *(q.e.d.)*
3. *Proof* $r\mathbf{u} \in \mathrm{Nul}\{A\}$
   $\mathbf{u} \in \mathrm{Nul}\{A\} \Rightarrow A\mathbf{u} = \mathbf{0}$
   $A(r\mathbf{u}) = rA\mathbf{u} = r\mathbf{0} = \mathbf{0} \Rightarrow r\mathbf{u} \in \mathrm{Nul}\{A\}$ *(q.e.d.)*

Basis of a subspace

Definition 10.4 (Basis of a subspace)

Let $H \subseteq \mathbb{R}^n$. The set of vectors B is a basis of H if:

1. *All vectors in B are linearly independent*
2. $H = \mathrm{Span}\{B\}$

Standard basis of $\mathbb{R}^n$

Let be the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \ldots \\ 0 \end{pmatrix} \quad \ldots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \ldots \\ 1 \end{pmatrix}$$

The set $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the **standard basis** of $\mathbb{R}^n$.

Example

Find a basis for the null space of $A = \begin{pmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{pmatrix}$.

Solution

The null space of A are all those vectors satisfying $A\mathbf{x} = \mathbf{0}$.

$$\left(\begin{array}{c|c} A & \mathbf{0} \end{array} \right) \sim \left(\begin{array}{ccccc|c} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

So the solution of the equation system is $\left. \begin{array}{c} x_1 = 2x_2 + x_4 - 3x_5 \\ x_3 = -2x_4 + 2x_5 \end{array} \right\} \Rightarrow$

$$\mathbf{x} = \begin{pmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_2 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$$

Null space and equation systems

Example (continued)

The set $B = \{(2,1,0,0,0),(1,0,-2,1,0),(-3,0,2,0,1)\}$ is a basis of $\mathrm{Nul}\{A\}$. By construction, we have chosen them to be linearly independent.

Example: Null space and equation systems

Consider $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

- $\{\mathbf{e}_3\}$ is a basis for $\mathrm{Nul}\{A\}$
- Consider $\mathbf{b} = (7,3,0)$. The general solution of $A\mathbf{x} = \mathbf{b}$ is of the form

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_{Nul}$$

where $\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$ that does not belong to $\mathrm{Nul}\{A\}$ and $\mathbf{x}_{Nul}$ belongs to $\mathrm{Nul}\{A\}$. In this particular case,

$$\mathbf{x} = (7,3,0) + x_3\mathbf{e}_3$$

Example: Null space and equation systems (continued)

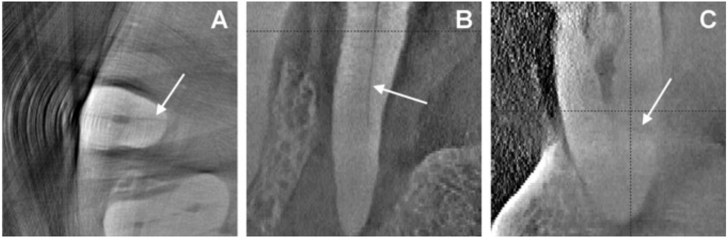
Let us prove that the general solution is actually a solution of $A\mathbf{x} = \mathbf{b}$

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{x}_{Nul}) = A\mathbf{x}_0 + A\mathbf{x}_{Nul} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

Intuititively we can say that the null space is the set of all solutions for which we have no measurements. The equation system only impose some constraints on those coefficients for which we have measurements. This is a problem in real situations as shown in the following slide.

Null space and equation systems

In this example, the authors describe how the exact location of a tooth fracture is uncertain (Fig. C) due to the artifacts introduced by the null space of the tomographic problem.



Mora, M. A.; Mol, A.; Tyndall, D. A., Rivera, E. M. *In vitro assessment of local computed tomography for the detection of longitudinal tooth fractures*.

Oral Surg Oral Med Oral Pathol Oral Radiol Endod, **2007**, 103, 825-829.

Basis of a subspace

Example

Find a basis for the column space of $B = \begin{pmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

Solution

From the columns with non-pivot positions of matrix B we learn that

$$\begin{aligned} \mathbf{b}_3 &= -3\mathbf{b}_1 + 2\mathbf{b}_2 \\ \mathbf{b}_4 &= 5\mathbf{b}_1 - \mathbf{b}_2 \end{aligned}$$

Then,

$$\begin{aligned} \mathrm{Col}\{B\} &= \left\{\mathbf{v} \in \mathbb{R}^4 | \mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3 + x_4\mathbf{b}_4 + x_5\mathbf{b}_5\right\} \\ &= \left\{\mathbf{v} \in \mathbb{R}^4 \,\middle|\, \begin{array}{rcl} \mathbf{v} & = & x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3(-3\mathbf{b}_1 + 2\mathbf{b}_2) + \\ & & x_4(5\mathbf{b}_1 - \mathbf{b}_2) + x_5\mathbf{b}_5 \end{array}\right\} \\ &= \left\{\mathbf{v} \in \mathbb{R}^4 | \mathbf{v} = x_1'\mathbf{b}_1 + x_2'\mathbf{b}_2 + x_5\mathbf{b}_5\right\} \end{aligned}$$

And, consequently, $\mathrm{Basis}\{\mathrm{Col}\{B\}\} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5\}$

Basis of a subspace

Example

Find a basis for the column space of $A = \begin{pmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{pmatrix}$.

Solution

It turns out that $A \sim B$ (B in the previous example). Since row operations do not affect linear dependence relations among the columns of the matrix, we should have

$$\mathbf{a}_3 = -3\mathbf{a}_1 + 2\mathbf{a}_2$$
$$\mathbf{a}_4 = 5\mathbf{a}_1 - \mathbf{a}_2$$

and $\mathrm{Basis}\{\mathrm{Col}\{A\}\} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$

Theorem 10.3

The pivot columns of A form a basis of $\mathrm{Col}\{A\}$.

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 1:

- 2.8.1
- 2.8.2
- 2.8.5

Outline

3. Matrix algebra
   - Matrix operations (a)
   - Inverse of a matrix (b)
   - Elementary matrices (b)
   - An algorithm to invert matrices (b)
   - Characterization of invertible matrices (c)
   - Invertible linear transformations (c)
   - Partitioned matrices (c)
   - LU factorization (d)
   - An application to computer graphics and image processing (d)
   - Subspaces of $\mathbb{R}^n$ (e)
   - Dimension and rank (e)

Coordinate system

Definition 11.1 (Coordinates of a vector in the basis B)

Suppose $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_p\}$ is a basis for the subspace $H \subseteq \mathbb{R}^n$. For each $\mathbf{x} \in H$, the coordinates of $\mathbf{x}$ relative to the basis B are the weights c_i such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + ... + c_p\mathbf{b}_p$$

*The **coordinates of $\mathbf{x}$ with respect to the basis** B is the vector in $\mathbb{R}^p$*

$$[\mathbf{x}]_B = \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_p \end{pmatrix}$$

Coordinate system

Example

Let $\mathbf{x} = (3, 12, 7)$, $\mathbf{v}_1 = (3, 6, 2)$, $\mathbf{v}_2 = (-1, 0, 1)$, $B = \{\mathbf{v}_1, \mathbf{v}_2\}$.

1. Show that B is a linearly independent set
2. Find the coordinates of $\mathbf{x}$ in the coordinate system B

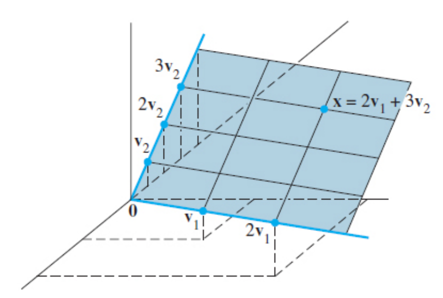
Solution

1. We need to prove that the only solution of the equation system $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$ is $c_1 = c_2 = 0$.
$$\left(\begin{array}{rr|r} 3 & -1 & 0 \\ 6 & 0 & 0 \\ 2 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$
And, therefore, the unique solution is $c_1 = c_2 = 0$ (q.e.d.)
2. We need to find c_1 and c_2 such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{x}$
$$\left(\begin{array}{rr|r} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{array}\right) \sim \left(\begin{array}{rr|r} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array}\right)$$
And, therefore, $[\mathbf{x}]_B = (2, 3)$.

Coordinate system

Example (continued)

The following figure shows how $\mathbf{x}$ is equal to $2\mathbf{v}_1 + 3\mathbf{v}_2$



Coordinate system

Theorem 11.1

The coordinates of a given vector with respect to a given basis are unique.
Proof
Let us assume they are not unique. Then, there must be two different sets of coordinates such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \ldots + c_p\mathbf{b}_p$$
$$\mathbf{x} = c_1'\mathbf{b}_1 + c_2'\mathbf{b}_2 + \ldots + c_p'\mathbf{b}_p$$

If we subtract both equations, we have

$$\mathbf{0} = (c_1 - c_1')\mathbf{b}_1 + (c_2 - c_2')\mathbf{b}_2 + \ldots + (c_p - c_p')\mathbf{b}_p$$

But because the basis is a linearly independent set of vectors, it must be

$$\begin{array}{c} c_1 - c_1' = 0 \Rightarrow c_1 = c_1' \\ c_2 - c_2' = 0 \Rightarrow c_2 = c_2' \\ \ldots \\ c_p - c_p' = 0 \Rightarrow c_p = c_p' \end{array}$$

This is a contradiction with the hypothesis that there were two different sets of coordinates, and therefore, the coordinates of the vector $\mathbf{x}$ *must be unique.*

Subspace dimension

Isomorphism to $\mathbb{R}^p$

For any given subspace H and its corresponding basis B, the mapping

$$\begin{array}{rcl} T : H & \rightarrow & \mathbb{R}^p \\ \mathbf{x} & \rightarrow & [\mathbf{x}]_B \end{array}$$

is a linear, injective transformation that makes H to behave as $\mathbb{R}^p$.

Definition 11.2 (Dimension)

The ***dimension of a subspace*** *H ($\dim\{H\}$) is the number of vectors of any of its basis.*
The dimension of $H = \{\mathbf{0}\}$ is 0.

Example (continued)

In our previous example in which $B = \{\mathbf{v}_1, \mathbf{v}_2\}$, the dimension is 2, in fact H behaves like a plane (see previous figure in the example).

Rank of a matrix

Definition 11.3 (Rank of a matrix)

*The **rank of a matrix** A is $\mathrm{rank}\{A\} = \dim\{\mathrm{Col}\{\mathrm{A}\}\}$, that is, the dimension of the column space of the matrix.*
MATLAB: `rank(A)`

Theorem 11.2

The rank of a matrix is the number of pivot columns it has.
Proof
Since the pivot columns form a basis of the column space of A, the number of pivot columns is the rank of the matrix.

Example

$$A = \begin{pmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore, the rank of A is 3.

Rank of a matrix

Theorem 11.3 (Rank theorem)

If A has n columns, then

$$\mathrm{Rank}\{A\} + \dim\{\mathrm{Nul}\{\mathrm{A}\}\} = n$$

Theorem 11.4 (Basis theorem)

Let H be a subspace of dimension p. Any linearly independent set of p vectors of H is a basis of H. Any set of p vectors that span H is a basis of H.

Characterization of invertible matrices (continued)

Theorem 11.5 (The invertible matrix theorem)

Let $A \in \mathcal{M}_{n\times n}$. The following statements are equivalent (either they are all true or they are all false):

xiii. *The columns of A form a basis of $\mathbb{R}^n$*

xiv. $\mathrm{Col}\{\mathrm{A}\} = \mathbb{R}^n$

xv. $\dim\{\mathrm{Col}\{\mathrm{A}\}\} = n$

xvi. $\mathrm{Rank}\{A\} = n$

xvii. $\mathrm{Nul}\{\mathrm{A}\} = \{\mathbf{0}\}$

xviii. $\dim\{\mathrm{Nul}\{\mathrm{A}\}\} = 0$

Characterization of invertible matrices

Proof $v \Rightarrow xiii$

This is true by the basis theorem.

Proof $xiii \Rightarrow xiv$

By the definition of basis.

Proof $xiii \Rightarrow xv$

By the definition of dimension.

Proof $xv \Rightarrow xvi$

By the definition of rank.

Proof $xvi \Rightarrow xviii$

By the rank theorem.

Proof $xvii \Rightarrow iv$

By the definition of null space.

Exercises

Exercises

From Lay (3rd ed.), Chapter 2, Section 9:

- 2.9.1
- 2.9.3
- 2.9.9
- 2.9.19
- 2.9.27

Outline

3 Matrix algebra

- Matrix operations (a)
- Inverse of a matrix (b)
- Elementary matrices (b)
- An algorithm to invert matrices (b)
- Characterization of invertible matrices (c)
- Invertible linear transformations (c)
- Partitioned matrices (c)
- LU factorization (d)
- An application to computer graphics and image processing (d)
- Subspaces of $\mathbb{R}^n$ (e)
- Dimension and rank (e)

Chapter 4. Determinant of a matrix

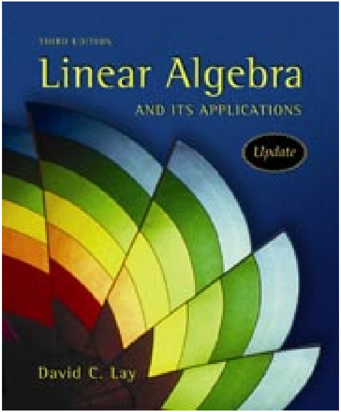
C.O.S. Sorzano

Biomedical Engineering

September 30, 2013

Outline

References



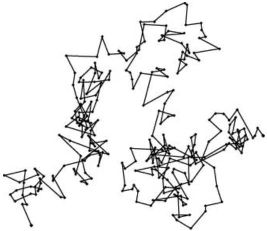
D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 3.

A little bit of history

The determinant of a matrix was first proposed by Seki Takakazu (1683) and Gottfried Leibniz (1693). Then Gabriel Cramer (1750) and Augustin Cauchy (1812) used them to solve problems in analytical geometry. Currently, they are not so much used in computational algebra, but they give important insights into the structure of a matrix.

Applications

The determinant plays an important role in the analysis of Brownian motion. It was first described by Robert Brown in 1827 (looking at pollen grains in water). Albert Einstein published in 1905 a paper in which he explained brownian motion as the result of the hitting molecules to bigger particles. This served as a theoretical basis for a posterior experiment by Jean Perrin that confirmed the existence of atoms. Jean Perrin was Nobel Prize in 1926.



See video at https://www.youtube.com/watch?v=hy-clLi8gHg

Outline

4 Determinant of a matrix
- Introduction
- Properties of determinants
- Cramer's rule
- Matrix inversion
- Areas and volumes

Cofactor

Definition 1.1 (Cofactor)

*The **cofactor** of the ij-th element of the matrix A is*

$$C_{ij} = (-1)^{i+j}\,|A_{ij}|$$

where A_{ij} is the matrix that results after eliminating the i-th row and the j-th column from matrix A.

Example

In the following example we calculate A_{32}

$$\begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

$$A_{32} = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

Determinant of a matrix

Definition 1.2 (Determinant of a matrix)

The ***determinant*** *of a square $n \times n$ matrix A ($|A|$ or $\det\{A\}$) is a mapping from $\mathcal{M}_{n\times n}$ onto $\mathbb{R}$ such that*

$$|A| = \begin{cases} A & n = 1 \\ a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n} & n \geq 2 \end{cases}$$

where a_{ij} is the ij-th element of matrix A.
MATLAB: `det(A)`

Example

$$\begin{aligned}
\det\left\{\begin{pmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{pmatrix}\right\} &= 1\det\left\{\begin{pmatrix} 4 & -1 \\ -2 & 0 \end{pmatrix}\right\} - 5\det\left\{\begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix}\right\} + 0\det\left\{\begin{pmatrix} 2 & 4 \\ 0 & -2 \end{pmatrix}\right\} \\
&= 1\cdot(-2) - 5\cdot 0 + 0\cdot(-4) = -2 \\
\det\left\{\begin{pmatrix} 4 & -1 \\ -2 & 0 \end{pmatrix}\right\} &= 4\det\{0\} - (-1)4\det\{-2\} = 4\cdot 0 - (-1)\cdot(-2) = -2 \\
\det\left\{\begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix}\right\} &= 2\det\{0\} - (-1)\det\{0\} = 2\cdot 0 - (-1)\cdot 0 = 0 \\
\det\left\{\begin{pmatrix} 2 & 4 \\ 0 & -2 \end{pmatrix}\right\} &= 2\det\{-2\} - 4\det\{0\} = 2\cdot(-2) - 4\cdot 0 = -4
\end{aligned}$$

Determinant of a matrix

Theorem 1.1

For $n \geq 2$, the determinant can be computed as a weighted sum of the cofactors along any row or column

$$|A| = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{i=1}^{n} a_{ij} C_{ij}$$

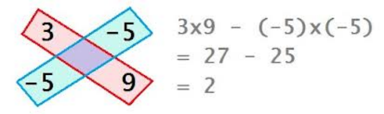
Example (continued)

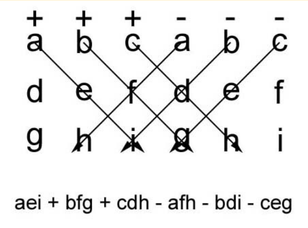
$$\det \left\{ \begin{pmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{pmatrix} \right\} = 0 \cdot C_{13} - 1 \cdot C_{23} + 0 \cdot C_{33} = -2$$

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 1 & 5 \\ 0 & -2 \end{vmatrix} = -(1|(-2)| - 5|(0)|) = 2$$

Theorem 1.2 (Useful particular cases)

- *For* $n = 2$,
  $|A| = a_{11}a_{22} - a_{12}a_{21}$
- *For* $n = 3$,
  $|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$





Determinant of a matrix

Theorem 1.3 (Useful particular cases (continued))

- *For triangular matrices,*

  $|A| = \prod_{i=1}^{n} a_{ii}$

Example

$$\begin{vmatrix} 1 & 4 & \frac{3}{5} & 2 \\ 0 & 1 & 2 & -10 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & 2 & -10 \\ 0 & 1 & 12 \\ 0 & 0 & 1 \end{vmatrix} = 1 \cdot 1 \begin{vmatrix} 1 & 12 \\ 0 & 1 \end{vmatrix} = 1 \cdot 1 \cdot 1 \, |1| = 1$$

Computing the determinant requires $O(n!)$ operations if we do it through the cofactor expansion. There are much faster algorithms $(O(n^3))$ that look for triangular matrices that have the same determinant as the original matrix and, then, they use this theorem that makes a much faster calculation.

Exercises

Exercises

From Lay (3rd ed.), Chapter 3, Section 1:

- 3.1.42
- 3.1.43 (with computer; MATLAB: `A=rand(4)`)
- 3.1.44 (with computer)
- 3.1.45 (with computer)
- 3.1.46 (with computer)

Outline

4 Determinant of a matrix

- Introduction
- Properties of determinants
- Cramer's rule
- Matrix inversion
- Areas and volumes

Properties of determinants

Theorem 2.1 (Determinant of the multiplication)

$\det\{AB\} = \det\{A\}\det\{B\}$
$\det\{kA\} = k^n\det\{A\}$

Note: In general, $\det\{A + B\} \neq \det\{A\} + \det\{B\}$

Theorem 2.2 (Determinant of row operations)

1. *If a multiple of one row of a matrix A is added to another row to obtain a matrix B, then $\det\{B\} = \det\{A\}$.*
2. *If two rows of a matrix A are interchanged to obtain a matrix B, then $\det\{B\} = -\det\{A\}$.*
3. *If a row of a matrix A is multiplied by k to obtain a matrix B, then $\det\{B\} = k\det\{A\}$.*

Properties of determinants

Example

Consider the following transformations that are of the form $B = EA$

1. $$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ k & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} A \Rightarrow |B| = |E||A| = 1|A|$$
2. $$B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} A \Rightarrow |B| = |E||A| = -1|A|$$
3. $$B = \begin{pmatrix} k & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} A \Rightarrow |B| = |E||A| = k|A|$$

Properties of determinants

Example

	$A = \begin{pmatrix} 2 & 4 & 6 \\ 3 & 5 & 7 \\ 1 & 2 & 3 \end{pmatrix}$	$\lvert A\rvert$
$\mathbf{r}_1 \leftarrow \frac{1}{2}\mathbf{r}_1$	$B_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 7 \\ 1 & 2 & 3 \end{pmatrix}$	$\lvert B_1\rvert = \frac{1}{2}\lvert A\rvert \Rightarrow \lvert A\rvert = 2\lvert B_1\rvert$
$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - 3\mathbf{r}_1$ $\mathbf{r}_2 \leftarrow \mathbf{r}_2 - \mathbf{r}_1$	$B_2 = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$	$\lvert B_2\rvert = \lvert B_1\rvert \Rightarrow$ $\lvert A\rvert = 2\lvert B_2\rvert = 2(1 \cdot (-1) \cdot 0) = 0$

Properties of determinants

Theorem 2.3

A is invertible iff $|A| \neq 0$. In that case, $|A^{-1}| = |A|^{-1}$.

Corollary

If $|A| = 0$, then the columns of A are not linearly independent.

Theorem 2.4

For any matrix $A \in \mathcal{M}_{n\times n}$, it is verified that $|A| = |A^T|$.

Corollary

The effect of column operations on the determinant is the same as the effect of row operations.

Exercises

Exercises

From Lay (3rd ed.), Chapter 3, Section 2:

- 3.2.14
- 3.2.15
- 3.2.18
- 3.2.19
- 3.2.24
- 3.2.31
- 3.2.32
- 3.2.33
- 3.2.45 (computer)

Outline

Cramer's rule

Cramer's rule is useful for a theoretical comprehension of what the determinant is and its properties, but it is not so useful for computational calculations.

Theorem 3.1 (Cramer's rule)

Let $A \in \mathcal{M}_{n\times n}$ be an invertible matrix. For every $\mathbf{b} \in \mathbb{R}^n$ the i-th entry of the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ is

$$x_i = \frac{\det\{A_i(\mathbf{b})\}}{\det\{A\}}$$

where $A_i(\mathbf{b})$ is the A matrix in which the i-th column has been substituted by $\mathbf{b}$, that is,

$$A_i(\mathbf{b}) = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \ldots & \mathbf{a}_n \end{pmatrix}$$

<u>Proof</u>
Let $\mathbf{e}_i$ $(i = 1, 2, ..., n)$ be the columns of the identity matrix I_n. Consider the product

$$\begin{aligned} AI_i(\mathbf{x}) &= \begin{pmatrix} A\mathbf{e}_1 & A\mathbf{e}_2 & \ldots & A\mathbf{e}_{i-1} & A\mathbf{x} & A\mathbf{e}_{i+1} & \ldots & A\mathbf{e}_n \end{pmatrix} = \\ &= \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \ldots & \mathbf{a}_n \end{pmatrix} = A_i(\mathbf{b}) \end{aligned}$$

Cramer's rule

Now we take the determinant on both sides

$$|A_i(\mathbf{b})| = |AI_i(\mathbf{x})| = |A||I_i(\mathbf{x})| = |A|x_i \Rightarrow x_i = \frac{|A_i(\mathbf{b})|}{|A|}$$

Example

Consider the equation system $\begin{pmatrix} 3s & -2 \\ -6 & s \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$. Its solution is given by

$$x_1 = \frac{\begin{vmatrix} 4 & -2 \\ 1 & s \end{vmatrix}}{\begin{vmatrix} 3s & -2 \\ -6 & s \end{vmatrix}} = \frac{4s+2}{3s^2-12} = \frac{4(s+\frac{1}{2})}{3(s-2)(s+2)}$$

$$x_2 = \frac{\begin{vmatrix} 3s & 4 \\ -6 & 1 \end{vmatrix}}{\begin{vmatrix} 3s & -2 \\ -6 & s \end{vmatrix}} = \frac{3s+24}{3s^2-12} = \frac{s+8}{(s-2)(s+2)}$$

Outline

4. Determinant of a matrix
   - Introduction
   - Properties of determinants
   - Cramer's rule
   - Matrix inversion
   - Areas and volumes

Matrix inversion

Algorithm to invert a matrix

We know that the inverse is a matrix such that $AA^{-1} = I_n$. If we call $\mathbf{x}_i$ to the i-th column of A^{-1}, then we have

$$AA^{-1} = A\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n\end{pmatrix} = \begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n\end{pmatrix}$$

i.e., we are solving simultaneously n equation systems of the form $A\mathbf{x}_j = \mathbf{e}_j$. The i-th entry of these columns is

$$x_{ij} = \frac{|A_i(\mathbf{e}_j)|}{|A|}$$

If we now calculate the determinant in the numerator by expanding by the j-th column, we have $|A_i(\mathbf{e}_j)| = (-1)^{i+j}|A_{ji}|$, where A_{ji} is the submatrix that results after eliminating the j-th row and the i-th column (or, what is the same, the cofactor of the ji-th element).

$$x_{ij} = \frac{(-1)^{i+j}|A_{ji}|}{|A|} = \frac{C_{ji}}{|A|}$$

Matrix inversion

Definition 4.1 (Adjoint (adjugate, *adjunta*) of a matrix)

Let $A \in \mathcal{M}_{n\times n}$ be a square matrix. The adjoint of A is another $n \times n$ matrix, denoted by A^* such that

$$A^*_{ij} = C_{ij}$$

Algorithm to invert a matrix (continued)

Finally we have

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} C_{11} & C_{21} & \ldots & C_{n1} \\ C_{12} & C_{22} & \ldots & C_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ C_{1n} & C_{2n} & \ldots & C_{nn} \end{pmatrix}$$

Watch out that the indexes of the cofactors are transposed with respect to the standard order. Consequently

$$A^{-1} = \frac{1}{|A|}(A^T)^*$$

Matrix inversion

Theorem 4.1

$$(A^T)^* = (A^*)^T$$

Example

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow$$

$$|A| = 1$$

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \quad C_{12} = (-1)^{1+2} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -1 \quad C_{13} = (-1)^{1+3} \begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix} = 0$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -1 \quad C_{22} = (-1)^{2+2} \begin{vmatrix} 2 & 0 \\ 0 & 1 \end{vmatrix} = 2 \quad C_{23} = (-1)^{2+3} \begin{vmatrix} 2 & 1 \\ 0 & 0 \end{vmatrix} = 0$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix} = 0 \quad C_{32} = (-1)^{3+2} \begin{vmatrix} 2 & 0 \\ 1 & 0 \end{vmatrix} = 0 \quad C_{33} = (-1)^{3+3} \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 1$$

$$A^* = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{|A|}(A^*)^T = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Outline

4 Determinant of a matrix
- Introduction
- Properties of determinants
- Cramer's rule
- Matrix inversion
- Areas and volumes

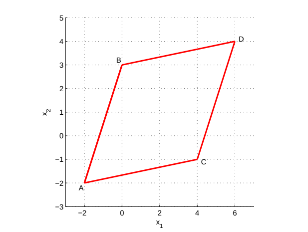
Areas and volumes

Theorem 5.1 (Area of a parallelogram, Volume of a parallelepiped)

If A is a 2×2 matrix, then $|\det\{A\}|$ is the area of the parallelogram formed by the columns of A. If A is a 3×3 matrix, then $|\det\{A\}|$ is the volume of the parallelepiped formed by the columns of A.

Example

Let be the parallelogram $ABCD$ ($A = (-2, -2)$, $B = (0, 3)$, $C = (4, -1)$, $D = (6, 4)$).



The area can be calculated as

$$\left|\det\begin{pmatrix} \mathbf{B}-\mathbf{A} & \mathbf{C}-\mathbf{A} \end{pmatrix}\right| =$$

$$\left|\det\begin{pmatrix} \begin{pmatrix} 0 \\ 3 \end{pmatrix} - \begin{pmatrix} -2 \\ -2 \end{pmatrix} & \begin{pmatrix} 4 \\ -1 \end{pmatrix} - \begin{pmatrix} -2 \\ -2 \end{pmatrix} \end{pmatrix}\right| =$$

$$\left|\det\begin{pmatrix} 2 & 6 \\ 5 & 1 \end{pmatrix}\right| = |-28| = 28$$

Areas and volumes

Theorem 5.2 (Area after a linear transformation)

Consider the transformation $T(\mathbf{x}) = A\mathbf{x}$.
If $A \in \mathcal{M}_{2\times 2}$ and S is a parallelogram in $\mathbf{R}^2$, then

$$\mathrm{Area}\{T(S)\} = |\det A|\mathrm{Area}\{S\}$$

If $A \in \mathcal{M}_{3\times 3}$ and S is a parallelepiped in $\mathbf{R}^3$, then the volume of $T(S)$ is

$$\mathrm{Volume}\{T(S)\} = |\det A|\mathrm{Volume}\{S\}$$

<u>Proof</u>
Let's prove it for the 2D case (the 3D one is analogous).
Consider the columns of A, $A = \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{pmatrix}$. Without loss of generality we may consider S to be at the origin with sides given by $\mathbf{b}_1$ and $\mathbf{b}_2$:

$$S = \left\{\mathbf{x} \in \mathbb{R}^2 | \mathbf{x} = s_1\mathbf{b}_1 + s_2\mathbf{b}_2 \; \forall s_1, s_2 \in [0,1]\right\}$$

Areas and volumes

The image of S by T is

$$T(S) = \left\{ \mathbf{y} \in \mathbb{R}^2 | \mathbf{y} = A\mathbf{x} = s_1 A\mathbf{b}_1 + s_2 A\mathbf{b}_2 \; \forall s_1, s_2 \in [0,1] \right\}$$

which is another parallelogram. Therefore, the area of $T(S)$ is

$$\begin{array}{rcl} \mathrm{Area}\{T(S)\} & = & \left|\det\begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{pmatrix}\right| = \left|\det\left\{A\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{pmatrix}\right\}\right| = \left|\det\{AB\}\right| \\ & = & |\det A||\det B| = |\det A| \mathrm{Area}\{S\} \end{array}$$

(q.e.d.)

Theorem 5.3

The previous theorem is valid for any closed region in $\mathbb{R}^2$ or $\mathbb{R}^3$ with finite area or volume.

Proof (hint)

We only need to divide the region into very small (infinitely small) parallelograms (or parallelepipeds) and apply the previous theorem to each one of the pieces.

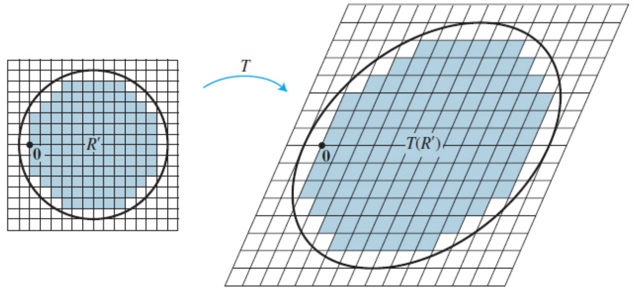


FIGURE 7 Approximating $T(R)$ by a union of parallelograms.

Areas and volumes

Example

Suppose that the unit disk defined as

$$D = \left\{ \mathbf{u} \in \mathbb{R}^2 | u_1^2 + u_2^2 \leq 1 \right\}$$

is transformed with the transformation

$$T(\mathbf{u}) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \mathbf{u}$$

to produce

$$E \equiv T(D) = \left\{ \mathbf{x} \in \mathbb{R}^2 \middle| \mathbf{x} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \mathbf{u} = \begin{pmatrix} au_1 \\ bu_2 \end{pmatrix} \right\}$$

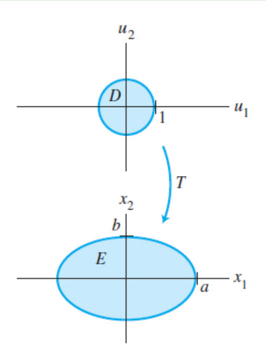
Exploiting the facts that $x_1 = au_1 \Rightarrow u_1 = \frac{x_1}{a}$, $x_2 = bu_2 \Rightarrow u_2 = \frac{x_2}{b}$ we may also characterize the transformed region as

$$E = \left\{ \mathbf{x} \in \mathbb{R}^2 \middle| \left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 \leq 1 \right\}$$

that is a solid ellipse.

Areas and volumes

Example (continued)



$$\begin{aligned} \mathrm{Area}\{E\} &= |\det A|\mathrm{Area}\{D\} = (ab)(\pi(1)^2) \\ &= \pi ab \end{aligned}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 3, Section 3:

- 3.3.1
- 3.3.7
- 3.3.11
- 3.3.21
- 3.3.25
- 3.3.26
- 3.3.29
- 3.3.32

Outline

4. Determinant of a matrix
   - Introduction
   - Properties of determinants
   - Cramer's rule
   - Matrix inversion
   - Areas and volumes

Chapter 5. Vector spaces

C.O.S. Sorzano

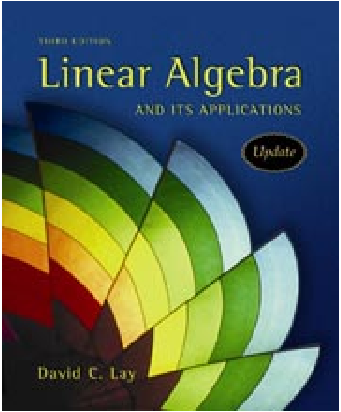
Biomedical Engineering

December 3, 2013

Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - Coordinate system (c)
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

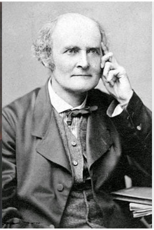
References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 4.

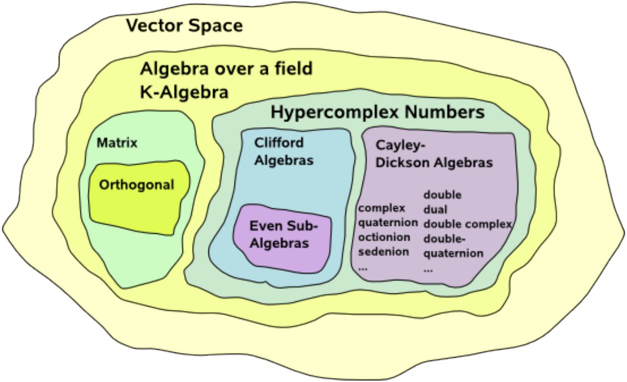
A little bit of history

Vectors were first used about 1636 in 2D and 3D to describe geometrical operations by René Descartes and Pierre de Fermat. In 1857 the notation of vectors and matrices was unified by Arthur Cayley. Giuseppe Peano was the firsst to give the modern definition of vector space in 1888, and Henri Lebesgue (about 1900) applied this theory to describe functional spaces as vector spaces.



Applications

It is difficult to think a mathematical tool with more applications than vector spaces. Thanks to them we may sum forces, control devices, model complex systems, denoise images, ... They underlie all these processes and it is thank to them that we can "nicely" operate with vectors. They are a mathemtical structure that generalizes many other useful structures.



Outline

5 Vector spaces
- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Vector space

Definition 1.1 (Vector space)

*A **vector space** is a non-empty set, V, of objects (called **vectors**) in which we define two operations: the sum among vectors and the multiplication by a scalar (an element of any field, $\mathbb{K}$), and that $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\forall c, d \in \mathbb{K}$ it is verified that*

1. $\mathbf{u} + \mathbf{v} \in V$
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. $\exists \mathbf{0} \in V | \mathbf{u} + \mathbf{0} = \mathbf{u}$
5. $\forall \mathbf{u} \in V \quad \exists! \mathbf{w} \in V | \mathbf{u} + \mathbf{w} = \mathbf{0}$ *(we normally write $\mathbf{w} = -\mathbf{u}$)*
6. $c\mathbf{v} \in V$
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$
10. $1\mathbf{u} = \mathbf{u}$

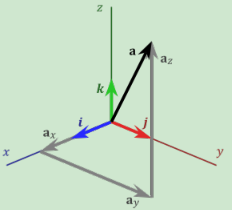
Vector space

Theorem 1.1 (Other properties)

11. $0\mathbf{u} = \mathbf{0}$
12. $c\mathbf{0} = \mathbf{0}$
13. $-\mathbf{u} = (-1)\mathbf{u}$

Watch out that 0 and 1 refer respectively to the neutral elements of the sum and multiplication in the field $\mathbb{K}$. -1 is the opposite number in $\mathbb{K}$ of 1 with respect to the sum of scalars.

Example: $\mathbb{R}^n$



$\mathbb{R}^n$ is a vector space of finite dimension for any n. As well as $\mathbb{C}^n$.

Example: Force fields in Physics

Consider V to be the set of all arrows (directed line segments) in 3D. Two arrows are regarded as equal if they have the same length and direction. Define the sum of arrows and the multiplication by a scalar as shown below:

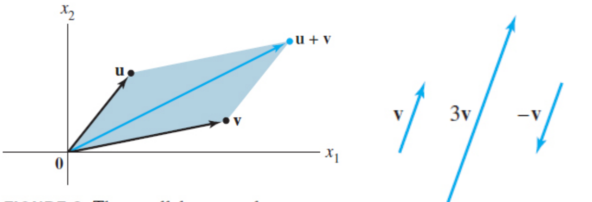


FIGURE 3 The parallelogram rule.

Example: Force fields in Physics (continued)

Here is an example of the application of some of the properties of vector spaces

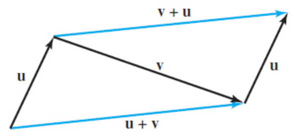
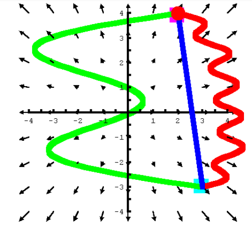


FIGURE 2 $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

FIGURE 3 $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

With a force field we may define at every point in 3D space, which is the force that is applied.

Conservative force field



Vector space

Example: Infinite sequences

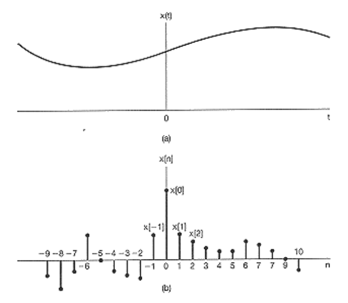
Let S be the set of all infinite sequences of numbers

$$\mathbf{u} = (..., u_{-2}, u_{-1}, u_0, u_1, u_2, ...)$$

Define the sum among two vectors and the multiplication by a scalar as

$$\mathbf{u}+\mathbf{v} = (..., u_{-2}+v_{-2}, u_{-1}+v_{-1}, u_0+v_0, u_1+v_1, u_2+v_2, ...)$$
$$c\mathbf{u} = (..., cu_{-2}, cu_{-1}, cu_0, cu_1, cu_2, ...)$$

Digital Signal Processing



Example: Polynomials of degree n ($\mathbb{P}_n$)

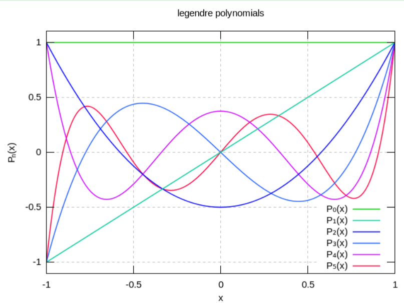
Let $\mathbb{P}_n$ be the set of all polynomials of degree n

$$u(x) = u_0 + u_1 x + u_2 x^2 + \ldots + u_n x^n$$

Define the sum among two vectors and the multiplication by a scalar as

$$(u+v)(x) = (u_0+v_0) + (u_1+v_1)x + (u_2+v_2)x^2 + \ldots + (u_n+v_n)x^n$$
$$(cu)(x) = cu_0 + cu_1 x + cu_2 x^2 + \ldots + cu_n x^n$$

Legendre polynomials



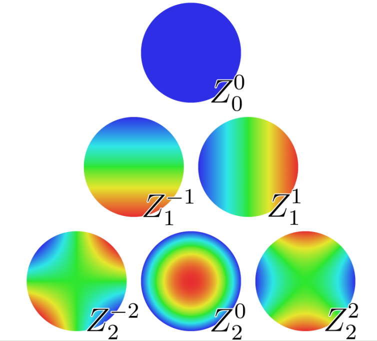
Vector space

Example: Set of real functions defined in some domain

Let $\mathbb{F}$ be the set of all real valued functions defined in some domain ($f : D \rightarrow \mathbb{R}$)
Define the sum among two vectors and the multiplication by a scalar as

$$(u+v)(x) = u(x) + v(x)$$
$$(cu)(x) = cu(x)$$

Ex: $u(x) = 3 + x$
Ex: $v(x) = \sin x$
Ex: Zernike polynomials



Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - Coordinate system (c)
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

Vector subspace

Sometimes we don't need to deal with the whole vector space, but only a part of it. It would be nice if it also has the space properties.

Definition 2.1 (Vector subspace)

*Let V be a vector space, and $H \subseteq V$ a part of it. H is **vector subspace** iff*

a) $\mathbf{0} \in H$

b) $\forall \mathbf{u}, \mathbf{v} \in H \quad \mathbf{u} + \mathbf{v} \in H$ *(H is closed with respect to sum)*

c) $\forall \mathbf{u} \in H$, $\forall c \in \mathbb{K} \quad c\mathbf{u} \in H$ *(H is closed with respect to scalar multiplication)*

Example

$H = \{\mathbf{0}\}$ is a subspace.

Example

The vector space of polynomials (of any degree), $\mathbb{P} \in \mathbb{F}(\mathbb{R})$, is a vector subspace of the vector space of real valued functions defined over $\mathbb{R}$ ($\mathbb{F}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R}\}$).

Vector subspace

Example

$H = \mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2 \not\subset \mathbb{R}^3$, for instance, the vector $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2$, but $\mathbf{u} \notin \mathbb{R}^3$.

Example

$H = \mathbb{R}^2 \times \{0\}$ is a subspace of $\mathbb{R}^3$ because all vectors of H are of the form $\mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \in \mathbb{R}^3$. It is obvious that H "looks like" $\mathbb{R}^2$. This resemblance is mathematically called **isomorphism**.

Example

Any plane in 3D passing through the origin is a subspace of $\mathbb{R}^3$.
Any plane in 3D <u>not</u> passing through the origin is <u>not</u> a subspace of $\mathbb{R}^3$, because $\mathbf{0}$ does not belong to the plane.

Vector subspace

Theorem 2.1

If H is a vector subspace, then H is a vector space.

Proof

a) $\Rightarrow$ *4*

$a \equiv \mathbf{0} \in H$

$4 \equiv \exists \mathbf{0} \in V | \mathbf{u} + \mathbf{0} = \mathbf{u}$

b) $\Rightarrow$ *1*

$b \equiv \forall \mathbf{u}, \mathbf{v} \in H \quad \mathbf{u} + \mathbf{v} \in H$

$1 \equiv \mathbf{u} + \mathbf{v} \in V$

Since $H \subset V$ and thanks to b) $\Rightarrow$ *2,3,7,8,9,10*

$2 \equiv \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

$3 \equiv (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

$7 \equiv c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

$8 \equiv (c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

$9 \equiv c(d\mathbf{u}) = (cd)\mathbf{u}$

$10 \equiv 1\mathbf{u} = \mathbf{u}$

Vector subspace

Proof (continued)

c) $\Rightarrow$ 6

$c \equiv \forall \mathbf{u} \in H, \ \forall c \in \mathbb{K} \quad c\mathbf{u} \in H$

$6 \equiv c\mathbf{v} \in V$

Proof of 5

Since H is a subset of V, we know that for every $\mathbf{u} \in H$ there exists a unique $\mathbf{w} \in V | \mathbf{u} + \mathbf{w} = \mathbf{0}$. The problem is whether or not $\mathbf{w}$ is in H. We also know that $\mathbf{w} = (-1)\mathbf{v}$, and by c), $\mathbf{w} \in H$.

(q.e.d.)

Outline

5 Vector spaces

- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Subspace spanned by a set of vectors

Example

Let $\mathbf{v}_1, \mathbf{v}_2 \in V$ be two vectors of a vector space, V. The subset

$$H = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

is a subspace of V.

Proof

Any vector of H is of the form $\mathbf{v} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2$ for some $\lambda_1, \lambda_2 \in \mathbb{K}$.

Proof a) $\mathbf{0} \in H$

Simply by setting $\lambda_1 = \lambda_2 = 0$, we get $\mathbf{0} \in H$

Proof b) $\mathbf{u} + \mathbf{v} \in H$

Let $\mathbf{u}, \mathbf{v} \in H \Rightarrow \left.\begin{array}{c} \mathbf{u} = \lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2 \\ \mathbf{v} = \lambda_{1v}\mathbf{v}_1 + \lambda_{2v}\mathbf{v}_2 \end{array}\right\} \Rightarrow$

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (\lambda_{1u}\mathbf{v}_1 + \lambda_{2u}\mathbf{v}_2) + (\lambda_{1v}\mathbf{v}_1 + \lambda_{2v}\mathbf{v}_2) \\ &= (\lambda_{1u} + \lambda_{1v})\mathbf{v}_1 + (\lambda_{2u} + \lambda_{2v})\mathbf{v}_2 \in H \end{aligned}$$

Subspace spanned by a set of vectors

Proof c) $c\mathbf{u} \in H$

Let $\mathbf{u} \in H \Rightarrow$

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \Rightarrow c\mathbf{u} = c(\lambda_u \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = c\lambda_u \mathbf{v}_1 + c\lambda_2 \mathbf{v}_2 \in H$$

Theorem 3.1

Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p \in V$ *be* p *vectors of a vector space,* V*. The subset*

$$H = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$$

is a subspace of V*.*

Proof

Analogous to the previous example.

Subspace spanned by a set of vectors

Example

Consider the set of vectors $\mathbb{R}^4 \supset H = \{(a-3b, b-a, a, b)\ \forall a, b \in \mathbb{R}\}$. Is it a vector subspace?

<u>*Solution*</u>

All vectors of H can be written as

$$H \ni \mathbf{u} = \begin{pmatrix} a-3b \\ b-a \\ a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Therefore, $H = \mathrm{Span}\{(1,-1,1,0), (-3,1,0,1)\}$ and by the previous theorem, it is a vector subspace.

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 1:

- 4.1.1
- 4.1.4
- 4.1.5
- 4.1.6
- 4.1.19
- 4.1.32
- 4.1.37 (computer)

Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - Coordinate system (c)
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

Null space of a matrix

Example

Consider the matrix

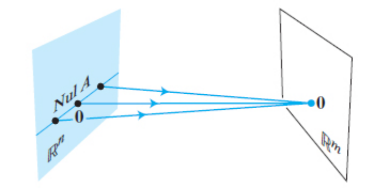
$$\begin{pmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{pmatrix}$$

The point $\mathbf{x} = (5, 3, -2)$ has the property that $A\mathbf{x} = \mathbf{0}$.

Definition 4.1 (Null space)

The ***null space of a matrix*** $A \in \mathcal{M}_{m\times n}$ *is the set of vectors*

$$\mathrm{Nul}\{A\} = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{0}\}$$



Null space of a matrix

Example (continued)

$$\left(\begin{array}{ccc|c} 1 & -3 & -2 & 0 \\ -5 & 9 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{5}{2} & 0 \\ 0 & 1 & \frac{3}{2} & 0 \end{array}\right)$$

Therefore

$$\mathrm{Nul}\{A\} = \{(-\tfrac{5}{2}x_3, -\tfrac{3}{2}x_3, x_3) \forall x_3 \in \mathbb{R}\}$$

The previous example ($\mathbf{x} = (5, 3, -2)$) is the point we obtain for $x_3 = -2$.

Null space of a matrix

Theorem 4.1

$\text{Nul}\{A\}$ *is a vector subspace of* $\mathbb{R}^n$.

Proof

It is obvious that $\text{Nul}\{A\} \subseteq \mathbb{R}^n$ *because* A *has* n *columns*

Proof a) $\mathbf{0} \in \text{Nul}\{A\}$

$A\mathbf{0}_n = \mathbf{0}_m \Rightarrow \mathbf{0}_n \in \text{Nul}\{A\}$

Proof b) $\mathbf{u} + \mathbf{v} \in \text{Nul}\{A\}$

Let $\mathbf{u}, \mathbf{v} \in \text{Nul}\{A\} \Rightarrow \left.\begin{array}{c} A\mathbf{u} = \mathbf{0} \\ A\mathbf{v} = \mathbf{0} \end{array}\right\} \Rightarrow$

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0} \Rightarrow \mathbf{u} + \mathbf{v} \in \text{Nul}\{A\}$$

Proof c) $c\mathbf{u} \in \text{Nul}\{A\}$

Let $\mathbf{u} \in H \Rightarrow$

$$A\mathbf{u} = \mathbf{0} \Rightarrow A(c\mathbf{u}) = c(A\mathbf{u}) = c\mathbf{0} = \mathbf{0} \Rightarrow c\mathbf{u} \in \text{Nul}\{A\}$$

Null space of a matrix

Example

Let $H = \left\{ (a, b, c, d) \in \mathbb{R}^4 \;\middle|\; \begin{array}{l} a - 2b + 5c = d \\ c - a = b \end{array} \right\}$. Is H a vector subspace of $\mathbb{R}^4$?

Solution

We may rewrite the conditions of belonging to H as

$$\begin{array}{l} a - 2b + 5c = d \\ c - a = b \end{array} \Rightarrow \begin{pmatrix} 1 & -2 & 5 & -1 \\ -1 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \mathbf{0}$$

and, thanks to the previous theorem, H is a vector subspace of $\mathbb{R}^4$.

Null space of a matrix

Example (continued)

We can even provide a basis for H

$$\begin{pmatrix} 1 & -2 & 5 & -1 \\ -1 & -1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

The solution of $A\mathbf{x} = \mathbf{0}$ are all points of the form

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} -c+d \\ 0 \\ c \\ d \end{pmatrix} = c \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + d \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Consequently $H = \mathrm{Span}\{(-1,0,1,0),(1,0,0,1)\}$.

Column space of a matrix

Definition 4.2 (Column space)

*Let $A \in \mathcal{M}_{m\times n}$ a matrix and $\mathbf{a}_i \in \mathbb{R}^m$ ($i = 1, 2, ...n$) its columns. The **column space of the matrix** A is defined as*

$$\mathrm{Col}\{A\} = \mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2, ...\mathbf{a}_n\} = \{\mathbf{b} \in \mathbb{R}^m | A\mathbf{x} = \mathbf{b} \textit{ for some } \mathbf{x} \in \mathbb{R}^n\}$$

Theorem 4.2

The column space of a matrix is a subspace of $\mathbb{R}^m$ Proof
$\mathrm{Col}\{A\}$ is a set generated by a number of vectors and by Theorem 3.1 it is a subspace of $\mathbb{R}^m$.

Column space of a matrix

Example

Find a matrix A such that $\mathrm{Col}\{A\} = \{(6a - b, a + b, -7a) \forall a, b \in \mathbb{R}\}$
Solution We can express the points in $\mathrm{Col}\{A\}$ as

$$\mathrm{Col}\{A\} \ni \mathbf{x} = \begin{pmatrix} 6a - b \\ a + b \\ -7a \end{pmatrix} = a \begin{pmatrix} 6 \\ 1 \\ -7 \end{pmatrix} + b \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

Therefore, $\mathrm{Col}\{A\} = \mathrm{Span}\{(6, 1, -7), (-1, 1, 0)\}$. That is, these must be the two columns of A

$$A = \begin{pmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{pmatrix}$$

Comparison between the Null and the Column spaces

Contrast Between Nul *A* and Col *A* for an *m* x *n* Matrix *A*

Nul A	Col A
1. Nul A is a subspace of $\mathbb{R}^n$.	**1.** Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2.** Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	**3.** It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4.** There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5.** A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6.** Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7.** Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8.** Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

Outline

5 Vector spaces

- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Linear transformation

We have said that $T(\mathbf{x}) = A\mathbf{x}$ is a linear transformation, but it is not the only one.

Definition 5.1 (Linear transformation)

The transformation $T : V \to W$ between two vectors spaces V and W is a rule that for each vector $\mathbf{v} \in V$ assigns a unique vector $\mathbf{w} = T(\mathbf{v}) \in W$, such that

1. $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) \quad \forall \mathbf{v}_1, \mathbf{v}_2 \in V$
2. $T(c\mathbf{v}) = cT(\mathbf{v}) \quad \forall \mathbf{v} \in V, \forall c \in \mathbb{K}$

Example

For a matrix $A \in \mathcal{M}_{m\times n}$, we have that

$$\begin{array}{rcl} T : \mathbb{R}^n & \to & \mathbb{R}^m \\ \mathbf{x} & \to & A\mathbf{x} \end{array}$$

is a linear transformation (we can easily verify that T meets the two required conditions).

Linear transformation

Example

Consider the space of all continuous, real-valued functions defined over $\mathbb{R}$ whose all derivatives are also continuous. We will refer to this space as $C^\infty(\mathbf{R})$. For instance, all polynomials belong to this space, as well as any sin, cos function. It can be proved that $C^\infty(\mathbf{R})$ is a vector space.
Consider the transformation that assigns to each function in $C^\infty(\mathbf{R})$ its derivative

$$\begin{array}{rcl} D: C^\infty(\mathbf{R}) & \to & C^\infty(\mathbf{R}) \\ f & \to & D(f) \end{array}$$

is a linear transformation.
Proof

1. $D(f+g) = D(f) + D(g)$
2. $D(cf) = cD(f)$

Kernel and range of transformation

Definition 5.2 (Kernel (Núcleo))

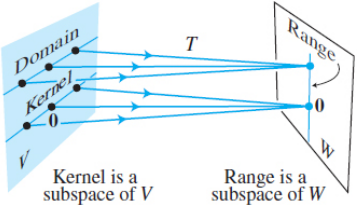
*The **kernel of a transformation** T is the set of all vectors such that*

$$\mathrm{Ker}\{T\} = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = \mathbf{0}\}$$

Definition 5.3 (Range (Imagen))

*The **range of a transformation** T is the set of all vectors such that*

$$\mathrm{Range}\{T\} = \{\mathbf{w} \in W \,|\exists \mathbf{v} \in V \;\; T(\mathbf{v}) = \mathbf{w}\}$$



Kernel and range of transformation

Example (continued)

$\mathrm{Ker}\{T\} = \mathrm{Nul}\{A\}$
$\mathrm{Ker}\{D\} = \{f(x) = c\}$ because $D(c) = 0$

Theorem 5.1

If $T(\mathbf{x}) = A\mathbf{x}$, then

$$\begin{aligned} \mathrm{Ker}\{T\} &= \mathrm{Nul}\{A\} \\ \mathrm{Range}\{T\} &= \mathrm{Col}\{A\} \end{aligned}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 2:

- 4.2.3
- 4.2.9
- 4.2.11
- 4.2.30
- 4.2.31

Outline

5 Vector spaces
- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Linear independence

Definition 6.1 (Linear independence)

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ is ***linearly independent*** *iff the only solution to the equation*

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p = \mathbf{0}$$

is the trivial solution ($c_1 = c_2 = ... = c_p = 0$). The set is linearly dependent if there exists another solution to the equation.

Watch out that we cannot simply put all vectors as columns of a matrix A and solve $A\mathbf{c} = \mathbf{0}$ because this is only valid for vectors in $\mathbb{R}^n$, but it is not valid for any vector space.

Linear independence

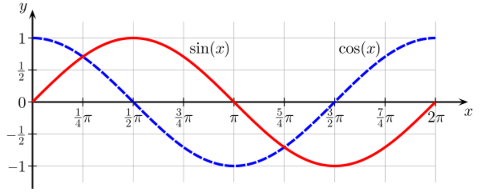
Example

- $\{\mathbf{v}_1\}$ is linearly dependent if $\mathbf{v}_1 = \mathbf{0}$.
- $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if $\mathbf{v}_2 = c\mathbf{v}_1$.
- $\{\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ is linearly dependent.

Example

In the vector space of continuous functions over $\mathbb{R}$, $C(\mathbb{R})$, the vectors $f_1(x) = \sin x$ and $f_2(x) = \cos x$ are independent because

$$f_2(x) \neq cf_1(x)$$



Linear independence

Theorem 6.1

*A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$, with $\mathbf{v}_1 \neq \mathbf{0}$ is **linearly dependent** if any of the vectors $\mathbf{v}_j$ ($j > 1$) is linearly dependent on the previous ones $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{j-1}\}$.*

Example

In the vector space of polynomials, consider the vectors $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = 4 - x$. The set $\{p_0(x), p_1(x), p_2(x)\}$ is linearly dependent because

$$p_2(x) = 4p_0(x) - p_1(x) \Rightarrow p_1(x) - 4p_0(x) + p_2(x) = 0$$

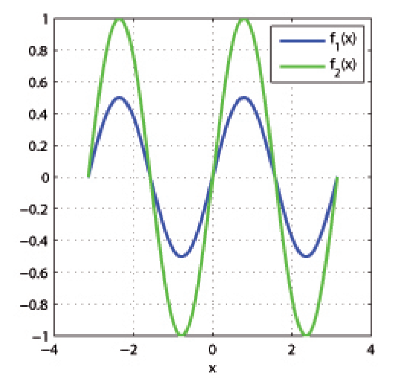
Linear independence

Example

In the vector space of continuous functions, consider the vectors $f_1(x) = \sin(x)\cos(x)$ and $f_2(x) = \sin(2x)$. The set $\{f_1(x), f_2(x)\}$ is linearly dependent because $f_2(x) = 2f_1(x)$

MATLAB:

```
x=[-pi:0.001:pi]
f1=sin(x).*cos(x);
f2=sin(2*x);
plot(x,f1,x,f2)
```



Basis of a subspace

Definition 6.2 (Basis of a subspace)

A set of vectors $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ *is a basis of the vector subspace* H *iff*

1. B *is a linearly independent set of vectors*
2. $H = \mathrm{Span}\{B\}$

In other words, a basis is a non-redundant set of vectors that span H*.*

Example

Let A be an invertible matrix. By Theorem 5.1 and 11.5 of Chapter 3 (the invertible matrix theorem), we know that the columns of A span $\mathbb{R}^n$ and that they are linearly independent. Consequently, the columns of A are a basis of $\mathbb{R}^n$.

Basis of a subspace

Example

The standard basis of $\mathbb{R}^n$ are the columns of I_n

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix} \quad \dots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$$

Example

Let $\mathbf{v}_1 = (3, 0, -6)$, $\mathbf{v}_2 = (-4, 1, 7)$, $\mathbf{v}_3 = (-2, 1, 5)$. Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis of $\mathbb{R}^3$?

Solution

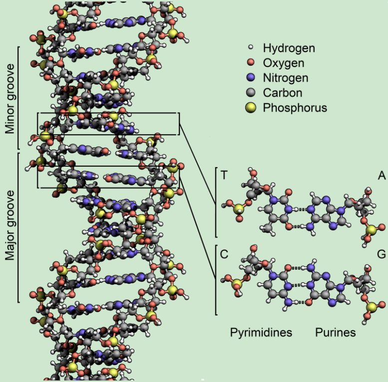
This question is the same as whether A is invertible with

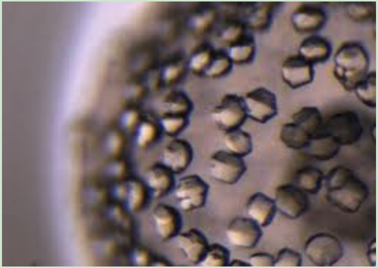
$$A = \begin{pmatrix} 3 & -4 & -2 \\ 0 & 1 & 1 \\ -6 & 7 & 5 \end{pmatrix} \Rightarrow |A| = 6 \Rightarrow \exists A^{-1}$$

Because A is invertible, we have that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis of $\mathbb{R}^3$.

Example: DNA Structure

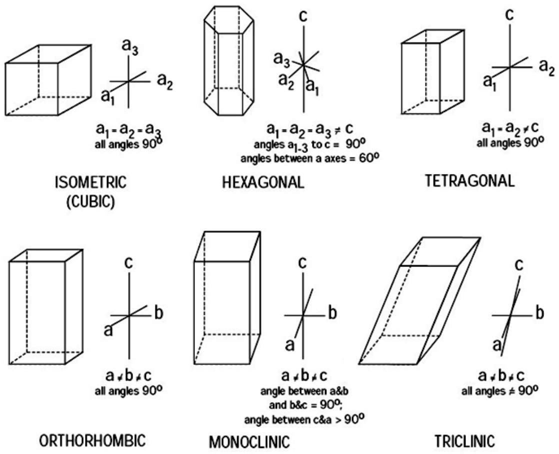
In 1953, Rosalind Franklin, James Watson and Francis Crick determined the 3D structure of DNA using data coming from X-ray diffraction of crystallized DNA. Watson and Crick received the Nobel prize in physiology and medicine in 1962 (Franklin died 1958).





Example: DNA Structure (continued)

Three-dimensional crystals repeat a certain motif all over the space following a crystal lattice. The vectors that define the crystal lattice are a basis of $\mathbb{R}^3$



Basis of a subspace

Example

$B = \{1, x, x^2, x^3, ...\}$ is the standard basis of the vector space of polynomials $\mathbb{P}$.
Proof

1. B is linearly independent:

$$\forall x \in \mathbb{R} \quad c_0 1 + c_1 x + c_2 x^2 + c_3 x^3 + ... = 0 \Rightarrow c_0 = c_1 = c_2 = ... = 0$$

   The only way that a polynomial of degree whichever is 0 for all values of x is that the coefficients of the polynomial are all 0.

2. $\mathbb{P} = \mathrm{Span}\{B\}$:
   It is obvious that any polynomial can be written as a linear combination of elements of B (in fact, this is they way we normally do).

Basis of a subspace

Example

$H = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ with $\mathbf{v}_1 = (0, 2, -1)$, $\mathbf{v}_2 = (2, 2, 0)$, $\mathbf{v}_3 = (6, 16, -5)$. Find a basis of H

Solution

All vectors in H are of the form:

$$H \ni \mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

We realize that $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, therefore, $\mathbf{v}_3$ is redundant:

$$\begin{aligned} H \ni \mathbf{x} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3(5\mathbf{v}_1 + 3\mathbf{v}_2) \\ &= (c_1 + 5c_3)\mathbf{v}_1 + (c_2 + 3c_3)\mathbf{v}_2 \\ &= c_1'\mathbf{v}_1 + c_2'\mathbf{v}_2 \end{aligned}$$

It suffices to construct our basis with $\mathbf{v}_1$ and $\mathbf{v}_2$.

Basis of a subspace

Theorem 6.2 (Spanning set theorem (conjunto generador))

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ be a set of vectors and $H = \mathrm{Span}\{S\}$. Then,

1. *If $\mathbf{v}_k$ is a linear combination of the rest, then the set $S - \{\mathbf{v}_k\}$ still generates H.*
2. *If $H \neq \{\mathbf{0}\}$, then some subset of S is a basis of H.*

Proof

1. *Assume that the linear combination that explains $\mathbf{v}_k$ is*

$$\mathbf{v}_k = a_1\mathbf{v}_1 + ... + a_{k-1}\mathbf{v}_{k-1} + a_{k+1}\mathbf{v}_{k+1} + ... + a_p\mathbf{v}_p$$

Consider any vector in H

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p \\ &= (c_1 + a_1)\mathbf{v}_1 + ... + (c_{k-1} + a_{k-1})\mathbf{v}_{k-1} + \\ &\quad (c_{k+1} + a_{k+1})\mathbf{v}_{k+1} + ... + (c_p + a_p)\mathbf{v}_p \end{aligned}$$

That is we can express $\mathbf{x}$ not using $\mathbf{v}_k$.

2. *Step 1: If S is a linearly independent set, then S is the basis of H.*
Step 2: If S is not, using the previous point we can remove a vector to produce S' that still generates H (go to Step 1).

Outline

5 Vector spaces
- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Basis for $\mathrm{Nul}\{A\}$

Example

Let $A = \begin{pmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{pmatrix}$

We solve the equation system $A\mathbf{x} = \mathbf{0}$ to find

$$(A|\mathbf{0}) \sim \left(\begin{array}{ccccc|c} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

we have coloured the pivot columns from which learn

$$\begin{array}{l} x_1 = 2x_2 + x_4 - 3x_5 \\ x_3 = -2x_4 + 2x_5 \end{array} \Rightarrow \mathrm{Nul}\{A\} \ni \mathbf{x} = \begin{pmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{pmatrix}$$

Basis for $\mathrm{Nul}\{A\}$

Example (continued)

$$\mathrm{Nul}\{A\} \ni \mathbf{x} = \begin{pmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_2 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$$

Finally the basis for $\mathrm{Nul}\{A\}$ is

$$\mathrm{Nul}\{A\} = \mathrm{Span} \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Basis for $\mathrm{Col}\{A\}$

Example

Consider A as in the previous example. We had

$$A \sim \begin{pmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = B$$

Let's call this latter matrix B. Non-pivot columns of B can be written as a linear combination of the pivot columns:

$$\begin{aligned} \mathbf{b}_2 &= -2\mathbf{b}_1 \\ \mathbf{b}_4 &= -\mathbf{b}_1 + 2\mathbf{b}_3 \\ \mathbf{b}_5 &= 3\mathbf{b}_1 - 2\mathbf{b}_3 \end{aligned}$$

Basis for $\mathrm{Col}\{A\}$

Example (continued)

Since row operations do not change the linear dependences among matrix columns, we can derive the same relationships for matrix A

$$\begin{aligned}\mathbf{a}_2 &= -2\mathbf{a}_1 \\ \mathbf{a}_4 &= -\mathbf{a}_1 + 2\mathbf{a}_3 \\ \mathbf{a}_5 &= 3\mathbf{a}_1 - 2\mathbf{a}_3\end{aligned}$$

Finally, the basis of $\mathrm{Col}\{A\}$ is $\{\mathbf{a}_1, \mathbf{a}_3\}$.

$$\mathrm{Col}\{A\} = \mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_3\} = \mathrm{Span}\left\{\begin{pmatrix}-3\\1\\2\end{pmatrix}, \begin{pmatrix}-1\\2\\5\end{pmatrix}\right\}$$

Basis for $\mathrm{Col}\{A\}$

Theorem 7.1

The pivot columns of A constitute a basis for $\mathrm{Col}\{A\}$.

Proof

Let B the reduced echelon form of A.

1. *The pivot columns of B form a linearly independent set because none of its elements can be expressed as a linear combination of the elements before each one of them.*
2. *The dependence relationships among columns are not affected by row operations. Therefore, the corresponding pivot columns of A are also linearly independent and, consequently, a basis of* $\mathrm{Col}\{A\}$.

Two views of a basis

As small as possible, as large as possible

1. The Spanning Set Theorem states that the basis is as small as possible as long as it spans the required subspace.
2. The basis has the maximum amount of vectors spanning the required subspace. If we add one more, the new set is not linearly independent.

Example

- $\{(1,0,0),(2,3,0)\}$ is a set of 2 linearly independent vectors. But it cannot span $\mathbb{R}^3$ because for this we need 3 vectors.
- $\{(1,0,0),(2,3,0),(4,5,6)\}$ is a set of 3 linearly independent vectors that spans $\mathbb{R}^3$, so it is a basis of $\mathbb{R}^3$.
- $\{(1,0,0),(2,3,0),(4,5,6),(7,8,9)\}$ is a set of 4 linearly dependent vectors that spans $\mathbb{R}^3$, so it cannot be a basis.

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 3:

- 4.3.1
- 4.3.2
- 4.3.8
- 4.3.12
- 4.3.24
- 4.3.31
- 4.3.32
- 4.3.33
- 4.3.37 (computer)

Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - **Coordinate system (c)**
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

Coordinate system

An important reason to assign a basis to a vector space V is that it makes V to "behave" as $\mathbb{R}^n$ through, what is called, a coordinate system.

Theorem 8.1 (The unique representation theorem)

Let $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$ a basis of the vector space V, and consider any vector $\mathbf{v} \in V$. There exists a unique set of scalars such that

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + ... + c_n\mathbf{b}_n$$

Proof

Let assume that there exists another set of scalars such that

$$\mathbf{v} = c_1'\mathbf{b}_1 + c_2'\mathbf{b}_2 + ... + c_n'\mathbf{b}_n$$

Subtracting both equations we have

$$\mathbf{0} = (c_1 - c_1')\mathbf{b}_1 + (c_2 - c_2')\mathbf{b}_2 + ... + (c_n - c_n')\mathbf{b}_n$$

But since the vectors $\mathbf{b}_i$ form a basis and are linearly independent, it must be

$$(c_1 - c_1') = (c_2 - c_2') = (c_n - c_n') = 0$$

Coordinate system

Proof (continued)

Finally, $c_1 = c'_1$, $c_2 = c'_2$, ..., $c_n = c'_n$ which is a contradiction with the hypothesis that there were two different sets of scalars representing the vector. Consequently, the set of scalars must be unique.

Definition 8.1 (Coordinates)

*Let $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$ a basis of the vector space V, and consider any vector $\mathbf{v} \in V$. The **coordinates** of $\mathbf{v}$ in B are the c_i coefficients such that*

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + ... + c_n\mathbf{b}_n \Rightarrow [\mathbf{v}]_B = \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{pmatrix}$$

*The transformation $T : V \to \mathbb{R}^n$ such that $T(\mathbf{x}) = [\mathbf{x}]_B$ is called the **coordinate mapping**.*

Coordinate system

Example

Let $B = \{(1,0),(1,2)\}$ be a basis of $\mathbb{R}^2$ and $[\mathbf{x}]_B = (-2,3)$, then

$$\mathbf{x} = -2\mathbf{b}_1 + 3\mathbf{b}_2 = -2\begin{pmatrix}1\\0\end{pmatrix} + 3\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1\\6\end{pmatrix}$$

In fact $(1,6)$ are the coordinates of $\mathbf{x}$ in the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$

$$\mathbf{x} = 1\mathbf{e}_1 + 6\mathbf{e}_2 = 1\begin{pmatrix}1\\0\end{pmatrix} + 6\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\6\end{pmatrix}$$

That is, the point $\mathbf{x}$ does not change, but depending on the coordinate system employed, we "see" it with different coordinates.

Coordinate system

Example (continued)

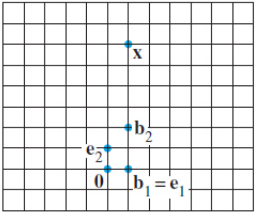


FIGURE 1 Standard graph paper.

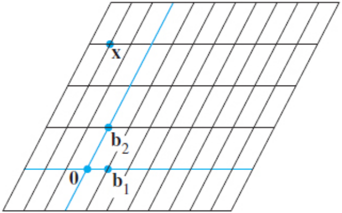
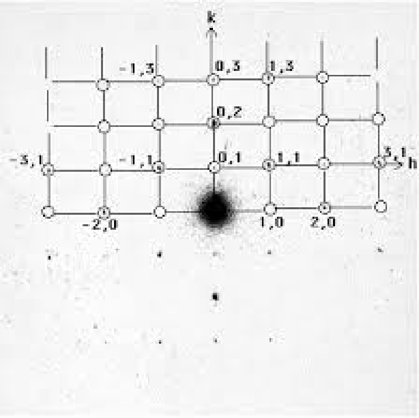


FIGURE 2 $\mathcal{B}$-graph paper.

Coordinate system

Example: X-ray diffraction

In ths figure we see how a X-ray diffraction pattern of a crystal is "indexed".



Coordinates in $\mathbb{R}^n$

If we have a point $\mathbf{x}$ in $\mathbb{R}$ we can easily find its coordinates in any basis, as in the following example.

Example

Let $\mathbf{x} = (4, 5)$ and the basis $B = \{(2, 1), (-1, 1)\}$. We need to find c_1 and c_2 such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 \Rightarrow \begin{pmatrix} 4 \\ 5 \end{pmatrix} = c_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

From which we can easily derive that $c_1 = 3$ and $c_2 = 2$.

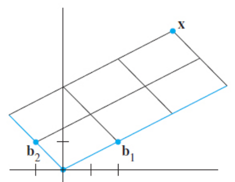


FIGURE 4

The $\mathcal{B}$-coordinate vector of **x** is $(3, 2)$.

Change of basis

Change from the standard basis to an arbitrary basis

Note that the previous equation system is of the form

$$\mathbf{x} = P_B[\mathbf{x}]_B$$

where P_B is called the **change-of-coordinates matrix** and its columns are the vectors of the basis B (consequently, it is invertible). We find the coordinates of the vector $\mathbf{x}$ in the basis B as

$$[\mathbf{x}]_B = P_B^{-1}\mathbf{x}$$

Change between two arbitrary bases

Let's say we know the coordinates of a point in some basis, B_1, and we want to know its coordinates in some other basis, B_2. We may use

$$\mathbf{x} = P_{B_1}[\mathbf{x}]_{B_1} = P_{B_2}[\mathbf{x}]_{B_2} \Rightarrow [\mathbf{x}]_{B_2} = P_{B_2}^{-1}P_{B_1}[\mathbf{x}]_{B_1}$$

Coordinate mapping

Theorem 8.2 (The coordinate mapping is an isomorphism between V and $\mathbb{R}^n$)

The coordinate mapping is a bijective, linear transformation.

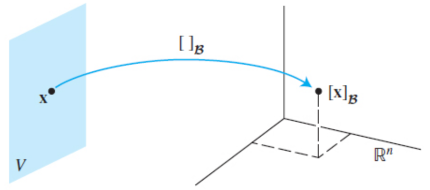


FIGURE 5 The coordinate mapping from V onto $\mathbb{R}^n$.

Corollary

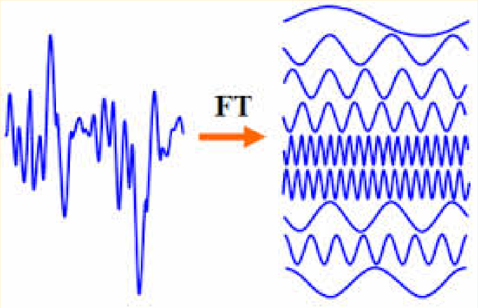
Since the coordinate mapping is a linear transformation it extends to linear combinations

$$[a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \dots + a_p\mathbf{u}_p]_B = a_1[\mathbf{u}_1]_B + a_2[\mathbf{u}_2]_B + \dots + a_p[\mathbf{u}_p]_B$$

Coordinate mapping

Consequences

Any operation in V can be performed in $\mathbb{R}^n$ and then go back to V.
For spaces of functions, this opens a new door to analyze functions (signals, images, ...) in $\mathbb{R}^n$ using the appropriate basis: Fourier transform, wavelet transform, Discrete Cosine Transform, ...



Coordinate mapping

Example

Consider the space of polynomials of degree 2, $\mathbb{P}_2$. any polynomial in this space is of the form

$$p(t) = a_0 + a_1 t + a_2 t^2$$

If we choose the standard basis in $\mathbb{P}_2$ that is

$$B = \{1, t, t^2\}$$

Then, we have the coordinate mapping

$$T(p(t)) = [p]_B = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

that is an isomorphism from $\mathbb{P}_2$ onto $\mathbb{R}^3$.

Coordinate mapping

Example (continued)

Now we can perform any reasoning in $\mathbb{P}_2$ by studying an analogous problem in $\mathbb{R}^3$. For instance, let's study if the following polynomials are linearly independent

$$\begin{array}{rcl} p_1(t) & = & 1+2t^2 & \Rightarrow & [p_1(t)]_B=(1,0,2) \\ p_2(t) & = & 4+t+5t^2 & \Rightarrow & [p_2(t)]_B=(4,1,5) \\ p_3(t) & = & 3+2t & \Rightarrow & [p_3(t)]_B=(3,2,0) \end{array}$$

We simply need to see if the corresponding coordinates in $\mathbb{R}^3$ are linearly independent

$$\begin{pmatrix} 1 & 4 & 3 \\ 0 & 1 & 2 \\ 2 & 5 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Looking at the non-pivot columns we learn that

$$p_3(t) = -5p_1(t) + 2p_2(t)$$

Finally, we conclude that the 3 polynomials are not linearly independent.

Coordinate mapping

Example

Consider $\mathbf{v}_1 = (3, 6, 2)$, $\mathbf{v}_2 = (-1, 0, 1)$, $B = \{\mathbf{v}_1, \mathbf{v}_2\}$, and $H = \mathrm{Span}\{B\}$. H is isomorphic to $\mathbb{R}^2$ (because its points have only 2 coordinates). For instance, the coordinates of $\mathbf{x} = (3, 12, 7) \in H$ are $[\mathbf{x}]_B = (2, 3)$.

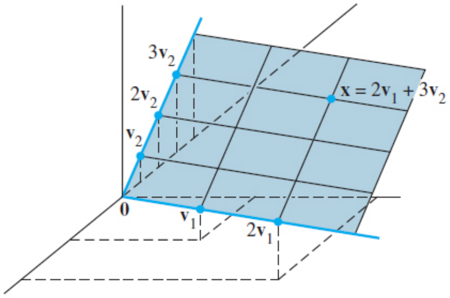


FIGURE 7 A coordinate system on a plane H in $\mathbb{R}^3$.

Coordinate mapping

Example

Consider $\mathbf{v}_1 = (3, 6, 2)$, $\mathbf{v}_2 = (-1, 0, 1)$, $B = \{\mathbf{v}_1, \mathbf{v}_2\}$, and $H = \mathrm{Span}\{B\}$. H is isomorphic to $\mathbb{R}^2$ (because its points have only 2 coordinates). For instance, the coordinates of $\mathbf{x} = (3, 12, 7) \in H$ are $[\mathbf{x}]_B = (2, 3)$.

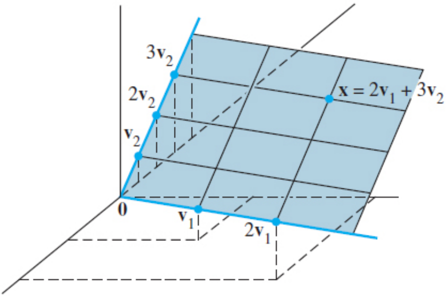


FIGURE 7 A coordinate system on a plane H in $\mathbb{R}^3$.

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 4:

- 4.4.3
- 4.4.8
- 4.4.9
- 4.4.13
- 4.4.17
- 4.4.19
- 4.4.22
- 4.4.24
- 4.4.25

Outline

5 Vector spaces

- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Dimension of a vector space

We have just said that if the basis of a vector space V has n elements, then V is isomorphic to $\mathbb{R}^n$. n is a characteristic number of each space called the dimension.

Theorem 9.1

Let V be a vector space with a basis given by $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$. Then, any subset of V with more than n elements is linearly dependent.

Proof

Let S be a subset of V with $p > n$ vectors

$$S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$$

We now consider the set of coordinates of these vectors.

$$\{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_p]_B\}$$

They are $p > n$ vectors in $\mathbb{R}^n$ and, therefore, necessarily linearly dependent. That is, there exist $c_1, c_2, ..., c_p$, not all of them 0, such that

$$c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + c_p[\mathbf{v}_p]_B = \mathbf{0} \in \mathbb{R}^n$$

Dimension of a vector space

Proof (continued)

If we now exploit the fact that the coordinate mapping is linear, then we have

$$[c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_p\mathbf{v}_p]_B = \mathbf{0} \in \mathbb{R}^n$$

Finally, we make use of the fact that the coordinate mapping is bijective

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_p\mathbf{v}_p = \mathbf{0} \in V$$

And, consequently, we have shown that the p vectors in S are linearly dependent.

Theorem 9.2

If a basis of a vector space has n vectors, then all other bases also have n vectors.

Proof

Let B_1 be a basis with n vectors of a vector space V. Let B_2 another basis of V. By the previous theorem, B_2 has at most n vectors. Let us assume now that B_2 has less than n vectors, then by the previous theorem B_1 would not be a basis. This is a contradiction with the fact that B_1 is a basis and, consequently, B_2 cannot have less than n vectors.

Dimension of a vector space

Definition 9.1

*If the vector space V is spanned by a finite set of vectors, then V is **finite-dimensional** and its **dimension** ($\dim\{V\}$) is the number of elements of any of its bases. The dimension of $V = \{\mathbf{0}\}$ is 0. If V is not generated by a finite set of vectors, then it is **infinite-dimensional**.*

Example

$\dim\{\mathbb{R}^n\} = n$
$\dim\{\mathbb{P}_2\} = 3$ because one of its bases is $\{1, t, t^2\}$
$\dim\{\mathbb{P}\} = \infty$
$\dim\{\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}\} = 2$

Dimension of a vector space

Example: in $\mathbb{R}^3$

There is a single subspace of dimension 0 ($\{\mathbf{0}\}$)
There are infinite subspaces of dimension 1 (all lines going through the origin)
There are infinite subspaces of dimension 2 (all planes going through the origin)
There is a single subspace of dimension 3 ($\mathbb{R}^3$)

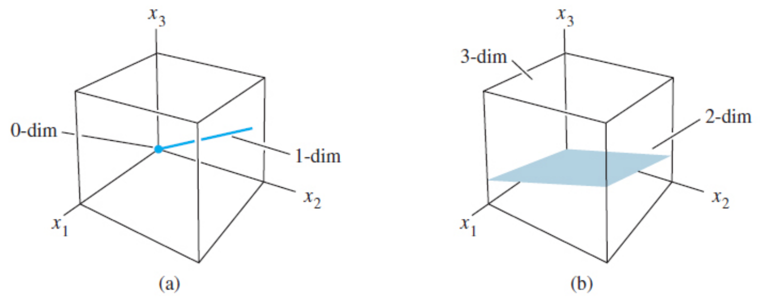


FIGURE 1 Sample subspaces of $\mathbb{R}^3$.

Dimension of a vector space

Theorem 9.3

Let $H \subseteq V$ be a vector subspace of a vector space V. Then,

$$\dim\{H\} \leq \dim\{V\}$$

Theorem 9.4

Let V a n-dimensional vector space ($n \geq 1$).

- *Any linearly independent subset of V with n elements is a basis.*
- *Any subset of V with n elements that span V is a basis.*

Dimension of a vector space

Theorem 9.5

Consider any matrix $A \in \mathcal{M}_{m\times n}$.

- $\dim\{\mathrm{Nul}\{A\}\}$ *is the number of free variables in the equation* $A\mathbf{x} = \mathbf{0}$.
- $\dim\{\mathrm{Col}\{A\}\}$ *is the number of pivot columns of* A.

Example

$$A = \begin{pmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The number of pivot columns of A is $2 = \dim\{\mathrm{Col}\{A\}\}$ (in blue), while the number of free variables is $3 = \dim\{\mathrm{Nul}\{A\}\}$ (the free variables are x_2, x_4 and x_5).

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 5:

- 4.5.1
- 4.5.13
- 4.5.21
- 4.5.25
- 4.5.26
- 4.5.27
- 4.5.28
- 4.5.31
- 4.5.32

Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - Coordinate system (c)
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

Rank of a matrix

The **rank of a matrix** is the number of linearly independent rows of that matrix. It can also be defined as the number of linearly independent columns of that matrix because both definitions yield the same number. We'll see a more formal definition below.

Definition 10.1 (Row space of a matrix)

*Given a matrix $A \in \mathcal{M}_{m\times n}$, the **row space** of A is the space spanned by all rows of A ($\mathrm{Row}\{A\} \subseteq \mathbb{R}^n$).*

Theorem 10.1

$$\mathrm{Row}\{A\} = \mathrm{Col}\{A^T\}$$

Rank of a matrix

Theorem 10.2

If a matrix A is row equivalent to another matrix B, then $\mathrm{Row}\{A\} = \mathrm{Row}\{B\}$. If B is in a reduced echelon form, then the non-null rows of B form a basis of $\mathrm{Row}\{A\}$

Proof

Proof $\mathrm{Row}\{A\} \supseteq \mathrm{Row}\{B\}$

Since the rows of B are obtained by row operations on the rows of A, then any linear combination of the rows of B can be obtained as linear combinations of the rows of A.

Proof $\mathrm{Row}\{A\} \subseteq \mathrm{Row}\{B\}$

Since the row operations are reversible, then any linear combination of the rows of A can be obtained as linear combinations of the rows of B.

Proof non-null rows of B form a basis

They are linearly independent because any non-null row of B cannot be obtained as a linear combination of the rows below (because it is in echelon form and there are numbers in early columns that have 0s below)

Rank of a matrix

Example

$$A = \begin{pmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{pmatrix} \sim B = \begin{pmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & -4 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Pivot columns have been highlighted in blue. At this point we can already construct a basis for the row and column spaces of A

$$\begin{aligned} \mathbb{R}^5 \supset \mathrm{Row}\{A\} &= \mathrm{Span}\{(1,3,-5,1,5),(0,1,-2,2,-7),(0,0,0,-4,20)\} \\ \mathbb{R}^4 \supset \mathrm{Col}\{A\} &= \mathrm{Span}\{(-2,1,3,1),(-5,3,11,7),(0,1,7,5)\} \end{aligned}$$

To calculate the null space of A we need the reduced echelon form

$$A \sim \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Example (continued)

$$A \sim \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow$$

$$\begin{array}{rcl} x_1 & = & -x_3 - x_5 \\ x_2 & = & 2x_3 - 3x_5 \\ x_4 & = & 5x_5 \end{array} \Rightarrow \mathrm{Nul}\{A\} \ni \mathbf{x} = x_3 \begin{pmatrix} -1 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -1 \\ -3 \\ 0 \\ 5 \\ 1 \end{pmatrix}$$

Finally,

$$\mathbb{R}^5 \supset \mathrm{Nul}\{A\} \quad = \quad \mathrm{Span}\{(-1, 2, 1, 0, 0), (-1, -3, 0, 5, 1)\}$$

Rank of a matrix

Definition 10.2 (Rank of a matrix)

$$\mathrm{Rank}\{A\} = \dim\{\mathrm{Col}\{A\}\}$$

That is, by definition, $\mathrm{Rank}\{A\}$ *is the number of pivot columns of* A.

Rank of a matrix

Theorem 10.3 (Rank theorem)

For any matrix $A \in \mathcal{M}_{m\times n}$

1. $\dim\{\mathrm{Row}\{A\}\} = \dim\{\mathrm{Col}\{A\}\}$
2. $\mathrm{Rank}\{A\} + \dim\{\mathrm{Nul}\{A\}\} = n$

Proof

1. *Let B be the reduced echelon form of A. By definition $\mathrm{Rank}\{A\}$ is the number of pivot columns in A (that is the same as the number of pivot columns in B). Since B is in reduced echelon form, each of its non-zero rows has a column pivot and, consequently, the number of non-zero rows coincides with the number of pivot columns. The basis of $\mathrm{Row}\{B\} = \mathrm{Row}\{A\}$ must have as many elements as pivot columns.*
2. *From Theorem 9.5 we know that $\mathrm{Null}\{A\}$ is the number of free variables in $A\mathbf{x} = \mathbf{0}$, that is, the number of non-pivot columns of B. Consequently, we have*

$$\dim\{\mathrm{Col}\{A\}\} + \dim\{\mathrm{Nul}\{A\}\} = n$$

But by definition, $\mathrm{Rank}\{A\} = \dim\{\mathrm{Col}\{A\}\}$, which proves the theorem.

Rank of a matrix

Example

Let $A \in \mathcal{M}_{7\times 9}$. We know $\dim\{\mathrm{Nul}\{A\}\} = 2$. What is $\mathrm{Rank}\{A\}$?
According to the previous theorem

$$\mathrm{Rank}\{A\} = n - \dim\{\mathrm{Nul}\{A\}\} = 9 - 2 = 7$$

Example

Let $A \in \mathcal{M}_{6\times 9}$. Can it be $\dim\{\mathrm{Nul}\{A\}\} = 2$?
Let us presume that it can be $\dim\{\mathrm{Nul}\{A\}\} = 2$, then

$$\mathrm{Rank}\{A\} = n - \dim\{\mathrm{Nul}\{A\}\} = 9 - 2 = 7$$

But since A has only 6 rows, the maximum rank can only be 6 (not 7), and therefore, it must be $\dim\{\mathrm{Nul}\{A\}\} \geq 3$.

Rank of a matrix

Example

$$A = \begin{pmatrix} 3 & 0 & -1 \\ 3 & 0 & -1 \\ 4 & 0 & 5 \end{pmatrix} \Rightarrow \begin{array}{l} \mathrm{Nul}\{A\} = \{(0, x_2, 0) \quad \forall x_2 \in \mathbb{R}\} \\ \mathrm{Row}\{A\} = \{(x_1, 0, x_3) \quad \forall x_1, x_3 \in \mathbb{R}\} \\ \mathrm{Col}\{A\} = \{(x_2, x_2, x_3) \quad \forall x_2, x_3 \in \mathbb{R}\} \\ \mathrm{Nul}\{A^T\} = \{(x_1, -x_1, 0) \quad \forall x_1 \in \mathbb{R}\} \end{array}$$

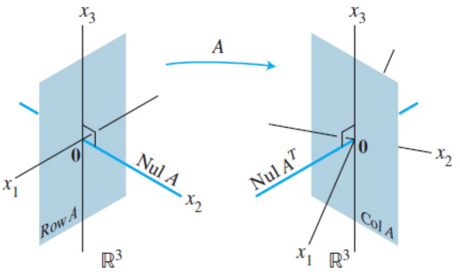


FIGURE 1 Subspaces determined by a matrix A.

Theorem 10.4 (The invertible matrix theorem (continued))

The following statements are equivalent to those in Theorems 5.1 and 11.5 of Chapter 3 (the invertible matrix theorem). Let $A \in \mathcal{M}_{n\times n}$

xix. *The columns of A form a basis of $\mathbb{R}^n$.*

xx. $\mathrm{Col}\{A\} = \mathbb{R}^n$.

xxi. $\dim\{\mathrm{Col}\{A\}\} = n$

xxii. $\mathrm{Rank}\{A\} = n$

xxiii. $\mathrm{Nul}\{A\} = \{\mathbf{0}\}$.

xxiv. $\dim\{\mathrm{Nul}\{A\}\} = 0$.

<u>Proof vii $\Leftrightarrow$ xx</u>

vii$\equiv$The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^n$.

But $\mathrm{Col}\{A\}$ is the set of all $\mathbf{b}$'s for which $A\mathbf{x} = \mathbf{b}$ has a solution. Therefore, vii $\Rightarrow$ xx.

<u>Proof xx $\Leftrightarrow$ xxi $\Leftrightarrow$ xxii</u>

Because of the definition of rank.

Rank of a matrix

Proof v,viii $\Leftrightarrow$ xix
v$\equiv$The columns of A are linearly independent.
viii$\equiv$The columns of A span $\mathbb{R}^n$.
But both together are the definition of a basis for $\mathbb{R}^n$.
Proof xxi $\Leftrightarrow$ xxiv
Knowing xxi and thanks to the rank theorem 10.3, we can infer that
$\dim\{\mathrm{Nul}\{A\}\} = n - n = 0$
Proof xxiv $\Leftrightarrow$ xxiii
The only subset with null dimension is $\{\mathbf{0}\}$.

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 6:

- 4.6.1
- 4.6.13
- 4.6.15
- 4.6.19
- 4.6.26
- 4.6.28
- 4.6.29
- 4.6.33
- 4.6.35

Outline

5 Vector spaces
- Definition (a)
- Vector subspace (a)
- Subspace spanned by a set of vectors (a)
- Null space and column space of a matrix (b)
- Kernel and range of a linear transformation (b)
- Linearly independent sets and bases (b)
- Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
- Coordinate system (c)
- Dimension of a vector space (d)
- Rank of a matrix (d)
- Change of basis (d)

Example

Let us assume we have a vector **x** that has two different coordinates in two different coordinate systems B and C.

$$[\mathbf{x}]_B = (3, 1) \text{ and } [\mathbf{x}]_C = (6, 4)$$

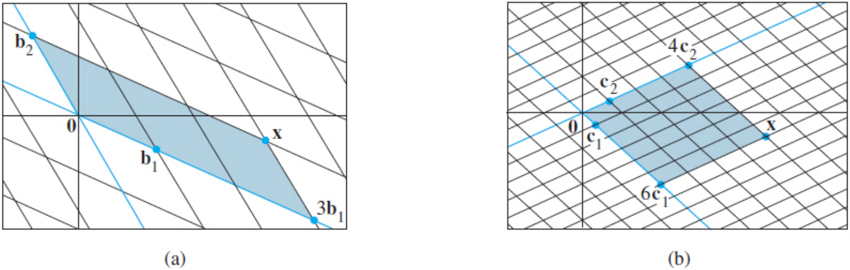


FIGURE 1 Two coordinate systems for the same vector space.

Change of basis

Example (continued)

Presume that for our example

$$\begin{array}{rcl} \mathbf{b}_1 & = & 4\mathbf{c}_1 + \mathbf{c}_2 \\ \mathbf{b}_2 & = & -6\mathbf{c}_1 + \mathbf{c}_2 \end{array}$$

We can calculate the coordinates of the basis vectors B in the C coordinate system as

$$\begin{array}{rcl} [\mathbf{b}_1]_C & = & (4, 1) \\ [\mathbf{b}_2]_C & = & (-6, 1) \end{array}$$

The coordinates of $\mathbf{x}$ in the basis B tell us

$$\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2$$

If we now apply the coordinate mapping transformation we have

$$[\mathbf{x}]_C = 3[\mathbf{b}_1]_C + [\mathbf{b}_2]_C = 3\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} -6 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 & -6 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

Change of basis

Example (continued)

Note that the columns of the matrix

$$\begin{pmatrix} 4 & -6 \\ 1 & 1 \end{pmatrix}$$

are the coordinates of each one of the elements of the basis B expressed in the coordinate system C, and that the overall change of coordinates has the form

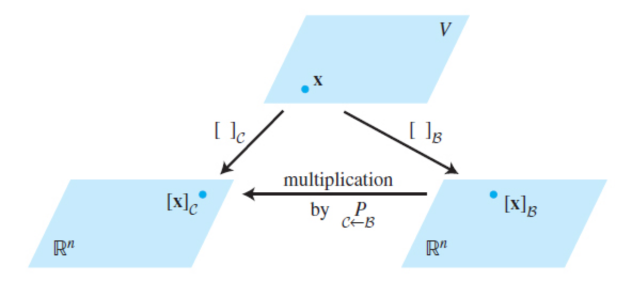
$$[\mathbf{x}]_C = \begin{pmatrix} 4 & -6 \\ 1 & 1 \end{pmatrix} [\mathbf{x}]_B$$

Change of basis

Theorem 11.1 (Change of basis)

Let $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ be two bases of the vector space V. We can transform coordinates from one coordinate system to the other by multiplying by a single, invertible $n \times n$ matrix, called $P_{C \leftarrow B}$ whose columns are the coordinates of the vectors of B in the basis C.

$$[\mathbf{x}]_C = P_{C \leftarrow B}[\mathbf{x}]_B$$



Change of basis

Corollary

To convert from C coordinates back to B coordinates we simply have to invert the transformation.

$$P_{B \leftarrow C} = P_{C \leftarrow B}^{-1}$$

Corollary

Consider the standard base in V given by $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$. The matrix to convert the coordinates from B to E is simply

$$P_{E \leftarrow B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{pmatrix}$$

Consequently, we have that for two different bases

$$\mathbf{x} = P_{E \leftarrow B}[\mathbf{x}]_B = P_{E \leftarrow C}[\mathbf{x}]_C$$

Finally,

$$[\mathbf{x}]_C = P_{E \leftarrow C}^{-1} P_{E \leftarrow B}[\mathbf{x}]_B$$

Change of basis

Numerical trick

Given the two basis B and C we can easily find the coordinates of B in the basis C in the following way. Let us define two matrices $\mathcal{B}$ and $\mathcal{C}$ whose columns are the elements of the basis. Then

$$(\mathcal{C}|\mathcal{B}) \sim (I_n|P_{C\leftarrow B})$$

Example

Let's say we are given $\mathbf{b}_1 = (-9, 1)$, $\mathbf{b}_2 = (-5, -1)$, $\mathbf{c}_1 = (1, -4)$, $\mathbf{c}_2 = (3, -5)$.

$$\left(\begin{array}{rr|rr} 1 & 3 & -9 & 5 \\ -4 & -5 & 1 & -1 \end{array}\right) \sim \left(\begin{array}{rr|rr} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & 3 \end{array}\right)$$

Then, $P_{C\leftarrow B} = \begin{pmatrix} 6 & 4 \\ -5 & 3 \end{pmatrix}$.

Exercises

Exercises

From Lay (3rd ed.), Chapter 4, Section 7:

- 4.7.1
- 4.7.9

Outline

5. Vector spaces
   - Definition (a)
   - Vector subspace (a)
   - Subspace spanned by a set of vectors (a)
   - Null space and column space of a matrix (b)
   - Kernel and range of a linear transformation (b)
   - Linearly independent sets and bases (b)
   - Bases for $\mathrm{Nul}\{A\}$ and $\mathrm{Col}\{A\}$ (c)
   - Coordinate system (c)
   - Dimension of a vector space (d)
   - Rank of a matrix (d)
   - Change of basis (d)

Chapter 6. Eigenvalues and eigenvectors

C.O.S. Sorzano

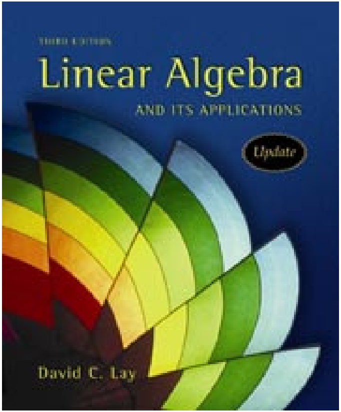
Biomedical Engineering

December 3, 2013

Outline

6. Eigenvalues and eigenvectors
   - Definition (a)
   - Characteristic equation (a)
   - Diagonalization (b)
   - Eigenvectors and linear transformations (b)
   - Complex eigenvalues (c)

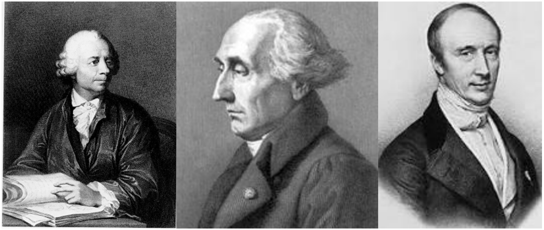
References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 5.

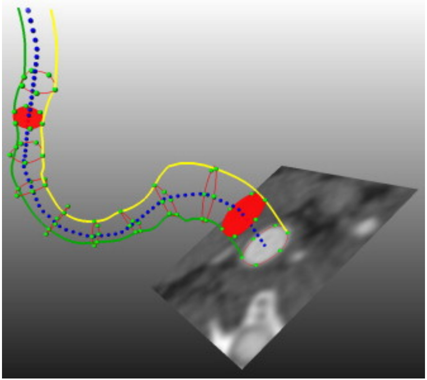
A little bit of history

Eigenvalues (or "proper values") were first used in the study of the motion of rigid bodies through the inertia matrix by Leonhard Euler and Joseph-Louis Lagrange in the mid of XVIIIth century. Then Augustin-Louis Cauchy used it to analyze quadratic surfaces and conic sections in the early XIXth. Since then, they have found applications in most scientific problems.



Applications

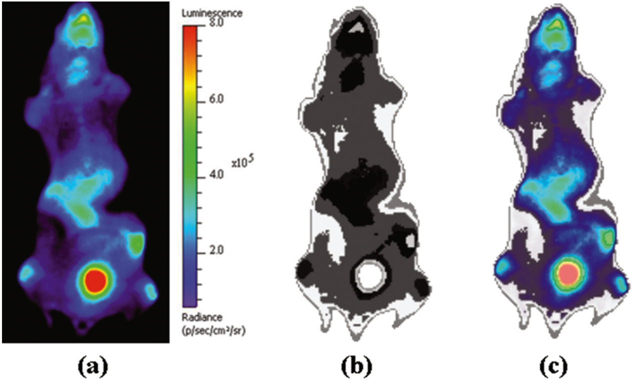
In this example eigenvalues are used to estimate the size of carotid in a volumetric image.



Hameeteman, K.; Zuluaga, M. A.; et al. *Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading*. Med Image Anal, **2011**, 15, 477-488.

Applications

In this example eigenvalues were used as a part of another technique (Principal Component Analysis) to automatically analyze luminiscent images.



Spinelli, A.E., Boschi, F. *Unsupervised analysis of small animal dynamic Cerenkov luminescence imaging*. J Biomed Opt, **2011**, 16, 120506

Outline

6 Eigenvalues and eigenvectors
- Definition (a)
- Characteristic equation (a)
- Diagonalization (b)
- Eigenvectors and linear transformations (b)
- Complex eigenvalues (c)

Eigenvalues and eigenvectors

Example

Consider the linear transformation $T(\mathbf{x}) = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \mathbf{x}$ on the vectors $\mathbf{u} = (-1, 1)$ and $\mathbf{v} = (2, 1)$

$$\begin{aligned} T(\mathbf{u}) &= \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ -1 \end{pmatrix} \\ T(\mathbf{v}) &= \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \end{aligned}$$

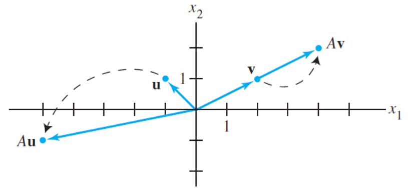


FIGURE 1 Effects of multiplication by A.

u is changing its direction and module, but **v** is only changing its module.

Eigenvalues and eigenvectors

Definition 1.1 (Eigenvalue and eigenvector)

Given the matrix $A \in \mathcal{M}_{n\times n}$, λ is an ***eigenvalue*** *of A if there exists a non-trivial solution $\mathbf{v} \in \mathbb{R}^n$ of the equation*

$$A\mathbf{v} = \lambda\mathbf{v}$$

The solution $\mathbf{v}$ is the ***eigenvector*** *associated to the eigenvalue λ.*

Example (continued)

In the previous example, $\mathbf{v}$ was an eigenvector with eigenvalue 2 (because $(2,1) \to (4,2)$, while $\mathbf{u}$ was not an eigenvector.

Eigenvalues and eigenvectors

Example

Show that $\lambda = 7$ is an eigenvalue of $A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}$.

Solution

We must find a solution of the equation $A\mathbf{v} = \lambda\mathbf{v}$, or what is the same

$$A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{v} = \mathbf{0}$$
$$\left(\begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} - 7\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -6 & 6 \\ 5 & -5 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Any vector of the form $\mathbf{v} = (v_1, v_1)$ satisfies the previous equation

Theorem 1.1

In general, eigenvectors are solution of the equation

$$(A - \lambda I)\mathbf{v} = \mathbf{0}$$

That is, all eigenvectors belong to $\mathrm{Nul}\{A - \lambda I\}$. *This is called the* ***eigenspace***.

Eigenvalues and eigenvectors

Example (continued)

We see that we have a whole set of vectors associated to $\lambda = 7$, this is a subspace of the eigenspace:

$$\text{Eigenspace}\{7\} = \{(v_1, v_1)\ \forall v_1 \in \mathbb{R}\}$$

It is a line passing through the origin with the direction $(1, 1)$.
The other eigenvalue of matrix A is $\lambda = -4$

$$\text{Eigenspace}\{-4\} = \{(v_1, -\tfrac{5}{6}v_1)\ \forall v_1 \in \mathbb{R}\}$$

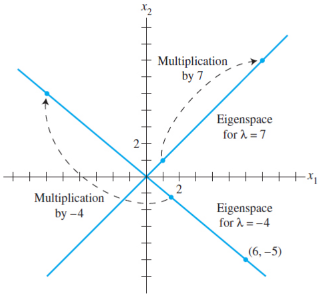


FIGURE 2 Eigenspaces for $\lambda = -4$ and $\lambda = 7$.

Eigenvalues and eigenvectors

Example

Knowing that $\lambda = 2$ is an eigenvalue of $A = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}$, find a basis of its eigenspace.

Solution

$$A - 2I = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix} - \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So any vector fulfilling this equation must satisfy

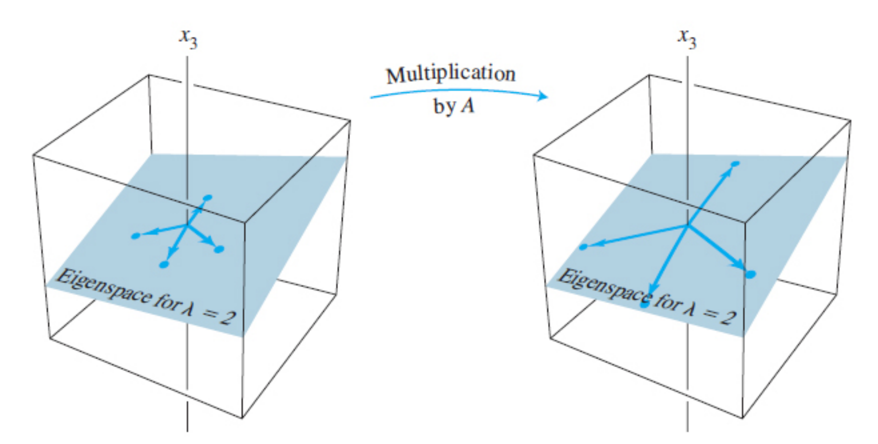
$$x_1 = \frac{1}{2}x_2 - 3x_3 \Rightarrow \mathrm{Eigenspace}\{2\} \ni \mathbf{x} = x_2 \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}$$

Finally the basis is formed by the vectors $(\frac{1}{2}, 1, 0)$ and $(-3, 0, 1)$.

Eigenvalues and eigenvectors

Example (continued)

Within the eigenspace, A acts as a dilation.



Eigenvalues and eigenvectors

Theorem 1.2

The eigenvalues of a triangular matrix A are the elements of the main diagonal (a_{ii}, $i = 1, 2, ..., n$).

<u>Proof</u>

Consider the matrix $A - \lambda I$

$$\begin{pmatrix} a_{11} - \lambda & a_{12} & a_{13} & ... & a_{1n} \\ 0 & a_{22} - \lambda & a_{23} & ... & a_{2n} \\ 0 & 0 & a_{33} - \lambda & ... & a_{3n} \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & a_{nn} - \lambda \end{pmatrix}$$

The equation system $A - \lambda I = \mathbf{0}$ has a non-trivial solution if at least 1 of the entries in the diagonal is 0. Therefore, it must be $\lambda = a_{ii}$ for some i. Varying i from 1 to n we obtain that all the elements in the main diagonal are the n eigenvalues of the matrix A.

Eigenvalues and eigenvectors

Example

The eigenvalues of $A = \begin{pmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{pmatrix}$ are $\lambda = 3, 0, 2$.

Theorem 1.3

Let $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_r$ be r eigenvectors associated to r different eigenvalues. Then, the set $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ is linearly independent.

Proof

Let us assume that S is linearly dependent. Without loss of generality, we may assume that the first p ($p < r$) are linearly independent, and that the $p+1$-th vector is dependent on the precedent vectors. Then, there must exist $c_1, c_2, ..., c_p$ not all of them zero such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p \tag{1}$$

Eigenvalues and eigenvectors

If we multiply both sides of the equation by A, then we have

$$\begin{array}{rcl} A\mathbf{v}_{p+1} & = & c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + \ldots + c_p A\mathbf{v}_p \\ \lambda_{p+1}\mathbf{v}_{p+1} & = & c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \ldots + c_p\lambda_p\mathbf{v}_p \end{array} \tag{2}$$

If we multiply Eq. (1) by λ_{p+1} and subtract from Eq. (2), we have

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{p+1})\mathbf{v}_2 + \ldots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p$$

Since the first p vectors are linearly independent it must be for $i = 1, 2, \ldots, p$

$$c_i(\lambda_i - \lambda_{p+1}) = 0$$

Because all eigenvalues are different, then it must be $c_i = 0$ $(i = 1, 2, \ldots, p)$. But this is a contradiction with the initial hypothesis that not all of them were 0. Consequently, the set S must be linearly independent. (q.e.d.)

Difference equations

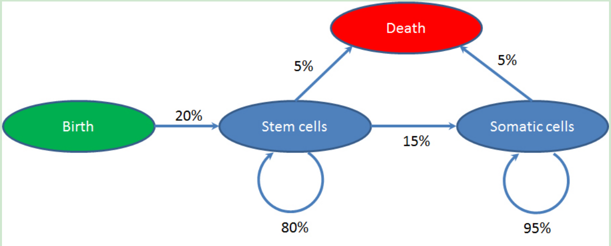
Let us assume we have two populations of cells: stem cells and mature cells. Everyday we measure the number of them and we observe that:

Stem cells:

- 80% of them have remained as stem cells
- 15% of them have differentiated into somatic cells
- 5% of them have died
- There are 20% new stem cells.

Somatic cells:

- 95% of them have remained as somatic cells
- 5% of them have died



Eigenvalues and eigenvectors

Difference equations (continued)

If we call $x_{stem}^{(k)}$ the number of stem cells on the day k, and $x_{somatic}^{(k)}$ the number of somatic cells the same day, then the following equation reflects the dynamics of the system:

$$\begin{pmatrix} x_{stem}^{(k+1)} \\ x_{somatic}^{(k+1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.15 & 0.95 \end{pmatrix} \begin{pmatrix} x_{stem}^{(k)} \\ x_{somatic}^{(k)} \end{pmatrix}$$

Let us assume that the day 0, there are $10,000$ stem cells, and 0 somatic cells. Then, the evolution over time is

$$\begin{pmatrix} x_{stem}^{(1)} \\ x_{somatic}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.15 & 0.95 \end{pmatrix} \begin{pmatrix} x_{stem}^{(0)} \\ x_{somatic}^{(0)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.15 & 0.95 \end{pmatrix} \begin{pmatrix} 10,000 \\ 0 \end{pmatrix} = \begin{pmatrix} 10,000 \\ 1,500 \end{pmatrix}$$

$$\begin{pmatrix} x_{stem}^{(2)} \\ x_{somatic}^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.15 & 0.95 \end{pmatrix} \begin{pmatrix} x_{stem}^{(1)} \\ x_{somatic}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.15 & 0.95 \end{pmatrix} \begin{pmatrix} 10,000 \\ 1,500 \end{pmatrix} = \begin{pmatrix} 10,000 \\ 2,925 \end{pmatrix}$$

...

Eigenvalues and eigenvectors

Difference equations

The previous model is of the form

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$$

The simplest way of constructing a solution of the previous equation is by taking an eigenvector $\mathbf{x}_1$ and its corresponding eigenvalue, λ:

$$\mathbf{x}^{(k)} = \lambda_1^k \mathbf{x}_1$$

This is actually a solution because:

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)} = A(\lambda_1^k \mathbf{x}_1) = \lambda_1^k (A\mathbf{x}_1) = \lambda_1^k(\lambda_1 \mathbf{x}_1) = \lambda_1^{k+1}\mathbf{x}_1$$

It turns out that any linear combination of eigenvectors is also a solution

$$\mathbf{x}^{(k)} = c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + ... + c_n\lambda_n^k\mathbf{x}_n$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 5, Section 1:

- 5.1.1
- 5.1.3
- 5.1.9
- 5.1.17
- 5.1.19
- 5.1.23
- 5.1.25
- 5.1.26
- 5.1.27

Outline

6. Eigenvalues and eigenvectors
   - Definition (a)
   - Characteristic equation (a)
   - Diagonalization (b)
   - Eigenvectors and linear transformations (b)
   - Complex eigenvalues (c)

Characteristic equation

Example

Find the eigenvalues of $A = \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix}$

Solution

We need to find scalar values λ such that the equation

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

has non-trivial solutions. By the Invertible Matrix theorem we know that this problem is equivalent to that of finding λ values such that

$$|A - \lambda I| = 0$$

In this case

$$\left| \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = 0$$

$$\begin{vmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{vmatrix} = (2-\lambda)(-6-\lambda) - 9 = \lambda^2 + 4\lambda - 21 = 0$$

Characteristic equation

Example (continued)

$$\lambda^2 + 4\lambda - 21 = 0 \Rightarrow \lambda = \frac{-4 \pm \sqrt{4^2 - 4 \cdot 1 \cdot (-21)}}{2 \cdot 1} = \begin{cases} -7 \\ 3 \end{cases}$$

Theorem 2.1 (The invertible matrix theorem (continued))

This theorem adds to the Theorems 5.1, 11.5 of Chapter 3 and 10.4 of Chapter 5.

xxv. $|A| \neq 0$.

xxvi. *0 is not an eigenvalue of A.*

Definition 2.1 (Characteristic equation)

A scalar λ is an eigenvalue of a matrix $A \in \mathcal{M}_{n \times n}$ iff it is solution of the ***characteristic equation***

$$|A - \lambda I| = 0$$

The determinant of $A - \lambda I$ is called the ***characteristic polynomial***.

Characteristic equation

Example

Let us calculate the eigenvalues of $A = \begin{pmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

$$|A - \lambda I| = \begin{vmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{vmatrix} = (5-\lambda)^2, (3-\lambda)(1-\lambda) = 0$$

whose solutions are $\lambda = 5$ (with multiplicity 2), $\lambda = 3$, and $\lambda = 1$.

Example

Let us find the eigenvalues of a matrix whose characteristic polynomial is

$$|A - \lambda I| = \lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2) = 0$$

whose solutions are $\lambda = 0$ (with multiplicity 4), $\lambda = 6$, and $\lambda = -2$.

Characteristic equation

Definition 2.2 (Similarity between matrices)

*Given two matrices $A, B \in \mathcal{M}_{n\times n}$, A is **similar** to B iff there exists an invertible matrix $P \in \mathcal{M}_{n\times n}$ such that*

$$B = P^{-1}AP$$

Watch out that *similarity* is not the same as *row equivalence* (A and B are row equivalent if there exists a E such that $B = EA$ being E invertible and the product of row operation matrices).

Characteristic equation

Theorem 2.2

If A is similar to B, then B is similar to A.

Proof

It suffices to take the definition of A similar to B and solve for B. If we multiply by P on the right

$$B = P^{-1}AP \Rightarrow PB = AP$$

Now, we multiply by P on the left (P^{-1} exists because P is invertible)

$$PB = AP \Rightarrow PBP^{-1} = A$$

and this is the definition of B being similar to A.

Characteristic equation

Theorem 2.3

If A and B are similar matrices, then they have the same characteristic polynomial.

Proof

If A is similar to B, then there exists an invertible matrix P such that

$$B = P^{-1}AP$$

If we subtract on both sides λI we have

$$B - \lambda I = P^{-1}AP - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(A - \lambda I)P$$

Now taking the determinant of both sides

$$|B - \lambda I| = |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P| = |P|^{-1}|A - \lambda I||P| = |A - \lambda I|$$

Characteristic equation

Theorem 2.4

If A and B are similar matrices, then they have the same characteristic polynomial.

Proof

If A is similar to B, then there exists an invertible matrix P such that

$$B = P^{-1}AP$$

If we subtract on both sides λI we have

$$B - \lambda I = P^{-1}AP - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(A - \lambda I)P$$

Now taking the determinant of both sides

$$|B - \lambda I| = |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P| = |P|^{-1}|A - \lambda I||P| = |A - \lambda I|$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 5, Section 2:

- 5.2.1
- 5.2.9
- 5.2.18
- 5.2.19
- 5.2.20
- 5.2.23
- 5.2.24
- 5.2.28 (computer)

Outline

Diagonalization

Definition 3.1 (Diagonalization)

*$A \in \mathcal{M}_{n\times n}$ is **diagonalizable** if there exists $P, D \in \mathcal{M}_{n\times n}$ (with P invertible and D diagonal) such that*

$$A = PDP^{-1}$$

Diagonalization simplifies the calculation of powers of A (A^k), is used to decouple dynamic systems, and in multivariate statistics to produce uncorrelated random variables.

Example

$$D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \quad D^2 = \begin{pmatrix} 5^2 & 0 \\ 0 & 3^2 \end{pmatrix} \; D^3 = \begin{pmatrix} 5^3 & 0 \\ 0 & 3^3 \end{pmatrix}$$

Diagonalization

Example

Let us assume that $A = PDP^{-1}$. Let us calculate calculate now the different powers of A

$$
\begin{aligned}
&A^2 = A \cdot A = (PDP^{-1})(PDP^{-1}) = (PD)(P^{-1}P)(DP^{-1}) = PDDP^{-1} = PD^2P^{-1} \\
&A^3 = A^2 \cdot A = (PD^2P^{-1})(PDP^{-1}) = PD^3P^{-1} \\
&\ldots \\
&A^k = PD^kP^{-1}
\end{aligned}
$$

Let us particularize this result for $A = \begin{pmatrix} 7 & 2 \\ -4 & 1 \end{pmatrix}$ that can be factorized with $P = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}$ and $D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$ as $A = PDP^{-1}$.

$$
A^k = PD^kP^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 5^k & 0 \\ 0 & 3^k \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{pmatrix}
$$

Diagonalization

Theorem 3.1 (Diagonalization theorem)

*$A \in \mathcal{M}_{n\times n}$ is **diagonalizable** iff A has n linearly independent eigenvectors. In this case, we may construct P by stacking the n eigenvectors, and D as a diagonal matrix with the corresponding eigenvalues.*

Proof

Consider the columns of $P = \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \ldots & \mathbf{p}_n \end{pmatrix}$ *and* $D = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & d_n \end{pmatrix}$

Let us assume that $A = PDP^{-1}$ *and we multiply by* P *on the right*

$$\begin{aligned} AP &= PD \\ A\begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \ldots & \mathbf{p}_n \end{pmatrix} &= \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \ldots & \mathbf{p}_n \end{pmatrix} \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & d_n \end{pmatrix} \\ \begin{pmatrix} A\mathbf{p}_1 & A\mathbf{p}_2 & \ldots & A\mathbf{p}_n \end{pmatrix} &= \begin{pmatrix} d_1\mathbf{p}_1 & d_2\mathbf{p}_2 & \ldots & d_n\mathbf{p}_n \end{pmatrix} \end{aligned}$$

Diagonalization

This implies that

$$\begin{aligned} A\mathbf{p}_1 &= d_1\mathbf{p}_1 \\ A\mathbf{p}_2 &= d_2\mathbf{p}_2 \\ &\ldots \\ A\mathbf{p}_n &= d_n\mathbf{p}_n \end{aligned}$$

But this is the definition of eigenvector, so the columns of P ($\mathbf{p}_i$) must be eigenvectors of A and d_i its corresponding eigenvalue. Since P is invertible, its columns must be linearly independent.

Example

Diagonalize $A = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}$.

Step 1: Find the eigenvalues of A

$$|A - \lambda I| = 0 \Rightarrow -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2 = 0$$

whose solutions are $\lambda = 1$ and $\lambda = -2$ (double).

Step 2: Find a linearly independent set of eigenvectors

$\underline{\lambda = 1}$

$$A - \lambda I = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{pmatrix} \sim \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Diagonalization

Example (continued)

Step 2: Find a linearly independent set of eigenvectors

$\underline{\lambda = 1}$

$$A - \lambda I \sim \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \Rightarrow \begin{array}{l} x_1 = -x_2 \\ x_3 = -x_2 \end{array} \Rightarrow \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

$\underline{\lambda = -2}$

$$A - \lambda I = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix} - \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{pmatrix} \sim$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow x_1 = -x_2 - x_3 \Rightarrow \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Diagonalization

Example (continued)

Step 3: Construct P and D

$$P = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Step 4: Check everything is correct

P is invertible $|P| \neq 0$

$$|P| = 1$$

$A = PDP^{-1} \Rightarrow AP = PD$

$$AP = \begin{pmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{pmatrix} \quad PD = \begin{pmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{pmatrix}$$

Diagonalization

Example (continued)

Step 4: Check everything is correct

$\underline{P \text{ is invertible } |P| \neq 0}$

MATLAB:

```
P=[1 -1 -1; -1 1 0; 1 0 1];
det(P)
```

$\underline{A = PDP^{-1} \Rightarrow AP = PD}$

MATLAB:

```
A=[1 3 3; -3 -5 3; 3 3 1];
P=[1 -1 -1; -1 1 0; 1 0 1];
D=[1 0 0; 0 -2 0; 0 0 -2];
A*P
P*D
```

Diagonalization

Example

Diagonalize $A = \begin{pmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{pmatrix}$.

Step 1: Find the eigenvalues of A

$$|A - \lambda I| = 0 \Rightarrow -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2 = 0$$

whose solutions are $\lambda = 1$ and $\lambda = -2$ (double). (Same eigenvalues as in the previous example)

Step 2: Find a linearly independent set of eigenvectors

$\underline{\lambda = 1}$

$$A - \lambda I = \begin{pmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 \\ -4 & -7 & -3 \\ 3 & 3 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Diagonalization

Example (continued)

Step 2: Find a linearly independent set of eigenvectors

$\underline{\lambda = 1}$

$$A - \lambda I \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{array}{c} x_1 = x_3 \\ x_2 = -x_3 \end{array} \Rightarrow \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

(The same eigenspace as in the previous example).

$\underline{\lambda = -2}$

$$A - \lambda I = \begin{pmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{pmatrix} - \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 3 \\ -4 & -4 & -3 \\ 3 & 3 & 3 \end{pmatrix} \sim$$

$$\begin{pmatrix} 1 & 1 & \frac{3}{4} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \Rightarrow \begin{array}{c} x_1 = -x_2 - \frac{3}{4}x_3 \\ \frac{1}{4}x_3 = 0 \end{array} \Rightarrow \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

(A cannot be diagonalized because there are not 3 linearly independent vectors)

Theorem 3.2

If a $n \times n$ matrix has n different eigenvalues, then it is diagonalizable.

Proof

Let $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ be the n eigenvectors corresponding to the n different eigenvalues. The set

$$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$$

is linearly independent by Theorem 1.3 and A is diagonalizable by Theorem 3.1.

Example

Is $A = \begin{pmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{pmatrix}$ diagonalizable?

Solution

A is a triangular matrix and its eigenvalues are 5, 0 and -2, all of them distinct, and by the previous theorem A is diagonalizable.

Diagonalization

Theorem 3.3

Let $A \in \mathcal{M}_{n\times n}$ with $p \leq n$ different eigenvalues. Let d_k be the dimension associated to the eigenvalue λ_k. Then,

1. *d_k is smaller or equal the multiplicity of λ_k.*
2. *A is diagonalizable iff d_k is equal to the multiplicity of λ_k. In this case,*
$$\sum_{k=1}^{p} d_k = n$$
3. *If A is diagonalizable and B_k are the bases of each one of the eigenspaces, then $\{B_1, B_2, ..., B_p\}$ is a basis of $\mathbb{R}^n$.*

Diagonalization

Example

Let $A = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 1 & 4 & -3 & 0 \\ -1 & -2 & 0 & 3 \end{pmatrix}$. Let's factorize it as $A = PDP^{-1}$. The eigenvalues and associated eigenvectors are

$$\lambda_1 = 5 \quad \leftrightarrow \quad \mathbf{v}_1 = \begin{pmatrix} -8 \\ 4 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} -16 \\ 4 \\ 0 \\ 1 \end{pmatrix} \qquad P = \begin{pmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow$$

$$\lambda_2 = -3 \quad \leftrightarrow \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad D = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 5, Section 3:

- 5.3.1
- 5.3.23
- 5.3.27
- 5.3.28
- 5.3.29
- 5.3.31
- 5.3.32
- 5.3.33 (computer)

Outline

The matrix of a linear transformation

The objective of this section is to show that if A is diagonalizable ($A = PDP^{-1}$), then the transformation $T_A(\mathbf{x}) = A\mathbf{x}$ is essentially the same as $T_D(\mathbf{u}) = D\mathbf{u}$.

Definition 4.1 (The matrix of a linear transformation)

Consider a linear transformation between two vectors spaces $T : U \to V$. Let B be a basis of V, and C be a basis of W. Let $\mathbf{x} \in V$ and consider its coordinates $[\mathbf{x}]_B = (r_1, r_2, ..., r_n)$.

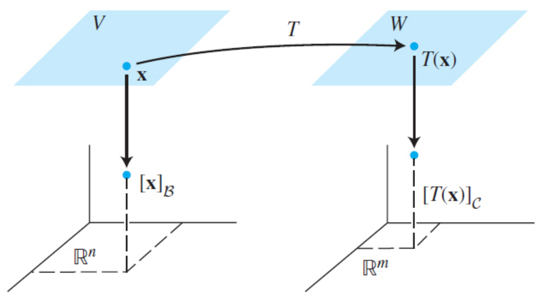


FIGURE 1 A linear transformation from V to W.

The matrix of a linear transformations

Let's analyze $\mathbf{x}$ and $T(\mathbf{x})$

$$\begin{array}{rcl} \mathbf{x} & = & r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + \ldots + r_n\mathbf{b}_n \Rightarrow \\ T(\mathbf{x}) & = & T(r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + \ldots + r_n\mathbf{b}_n) \quad \text{[T is linear]} \\ & = & r_1 T(\mathbf{b}_1) + r_2 T(\mathbf{b}_2) + \ldots + r_n T(\mathbf{b}_n) \end{array}$$

Now, let us consider the coordinates in C of the transformed vector

$$[T(\mathbf{x})]_C = r_1[T(\mathbf{b}_1)]_C + r_2[T(\mathbf{b}_2)]_C + \ldots + r_n[T(\mathbf{b}_n)]_C$$

We can write this equation in matrix form as

$$[T(\mathbf{x})]_C = M[\mathbf{x}]_B$$

where $M \in \mathcal{M}_{m\times n}$ is a matrix formed by the transformations of each one of the basis vectors in B

$$M = \begin{pmatrix} [T(\mathbf{b}_1)]_C & [T(\mathbf{b}_2)]_C & \ldots & [T(\mathbf{b}_n)]_C \end{pmatrix}$$

Matrix M is called the **matrix of T relative to the bases B and C**.

The matrix of a linear transformations

$$
\begin{tikzcd}
\mathbf{x} \arrow[r, "T"] \arrow[d] & T(\mathbf{x}) \arrow[d] \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } M"'] & {[T(\mathbf{x})]_{\mathcal{C}}}
\end{tikzcd}
$$

Example

Let $B = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ and

$$
\begin{array}{rcl}
T(\mathbf{b}_1) & = & 3\mathbf{c}_1 - 2\mathbf{c}_2 + 5\mathbf{c}_3 \\
T(\mathbf{b}_2) & = & 4\mathbf{c}_1 + 7\mathbf{c}_2 - \mathbf{c}_3
\end{array}
\Rightarrow M = \begin{pmatrix} 3 & 4 \\ -2 & 7 \\ 5 & -1 \end{pmatrix}
$$

Transformations from V into V

Definition 4.2 (B-matrix for T)

*If T is a transformation from V into V and B is a basis of V, then the matrix M is called the B-**matrix of** T.*

Example

Consider in the vector space of polynomials of degree 2 ($\mathbb{P}_2$), the derivative transformation

$$\begin{aligned} T: \quad & \mathbb{P}_2 \to \mathbb{P}_2 \\ & T(a_0 + a_1 t + a_2 t^2) = a_1 + 2a_2 t \end{aligned}$$

Consider the standard basis of $\mathbb{P}_2$, $B = \{1, t, t^2\}$.

Example (continued)

Which is the B-transformation matrix?

Solution

$$\begin{aligned}
T(1) = 0 \quad &\rightarrow \quad [T(1)]_B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
T(t) = 1 \quad &\rightarrow \quad [T(t)]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \Rightarrow M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \\
T(t^2) = 2t \quad &\rightarrow \quad [T(t^2)]_B = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}
\end{aligned}$$

Transformations from V into V

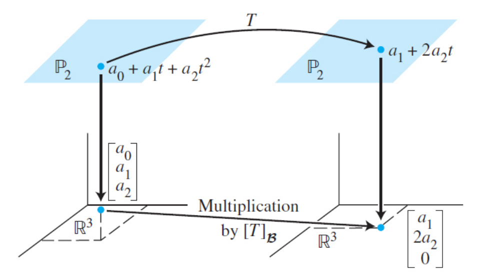
Example (continued)

Verify that $[T(\mathbf{x})]_B = M[\mathbf{x}]_B$

Solution

Given any polynomial $p(t) = a_0 + a_1 t + a_2 t^2$ its coordinates are $[p(t)]_B = (a_0, a_1, a_2)$. The derivative of $p(t)$ is $T(p(t)) = a_1 + 2a_2 t$, then

$$[T(p(t))]_B = \begin{pmatrix} a_1 \\ 2a_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$



Theorem 4.1 (Diagonal matrix representation)

Suppose matrix A is diagonalizable ($A = PDP^{-1}$). If B is the basis of $\mathbb{R}^n$ formed by the columns of P, then D is the B-matrix of the linear transformation $T(\mathbf{x}) = A\mathbf{x}$.

Proof

Let $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n$ be the columns of P so that $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$ is a basis. We know that for any basis in $\mathbb{R}^n$

$$\mathbf{x} = P[\mathbf{x}]_B \Rightarrow [\mathbf{x}]_B = P^{-1}\mathbf{x}$$

Let $[T]_B$ be the transformation matrix in the basis B. We know that by definition

$$\begin{array}{rcll} [T]_B & = & \left([T(\mathbf{b}_1)]_B \quad [T(\mathbf{b}_2)]_B \quad \ldots \quad [T(\mathbf{b}_n)]_B\right) & (T(\mathbf{x}) = A\mathbf{x}) \\ & = & \left([A\mathbf{b}_1]_B \quad [A\mathbf{b}_2]_B \quad \ldots \quad [A\mathbf{b}_n]_B\right) & \textit{(change of coordinates)} \\ & = & \left(P^{-1}A\mathbf{b}_1 \quad P^{-1}A\mathbf{b}_2 \quad \ldots \quad P^{-1}A\mathbf{b}_n\right) & \textit{(matrix multiplication)} \\ & = & P^{-1}A\left(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_n\right) & \textit{(definition of } P\textit{)} \\ & = & P^{-1}AP = D & \end{array}$$

Transformations from $\mathbb{R}^n$ into $\mathbb{R}^n$

Example

Let $T(\mathbf{x}) = \begin{pmatrix} 7 & 2 \\ -4 & 1 \end{pmatrix} \mathbf{x}$. Find a basis B in which the B-matrix of T is diagonal.

Solution

We diagonalize A as $A = PDP^{-1}$, with $P = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}$ and $D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$. We may change vectors $\mathbf{x}$ to the basis $B = \{(1,-1),(1,-2)\}$ by applying

$$\mathbf{u} = P^{-1}\mathbf{x}$$

Then, in this new basis, T can be applied as

$$T(\mathbf{u}) = D\mathbf{u} = DP^{-1}\mathbf{x}$$

If we now, come back to the original basis

$$T(\mathbf{x}) = PT(\mathbf{u}) = PDP^{-1}\mathbf{x} = A\mathbf{x}$$

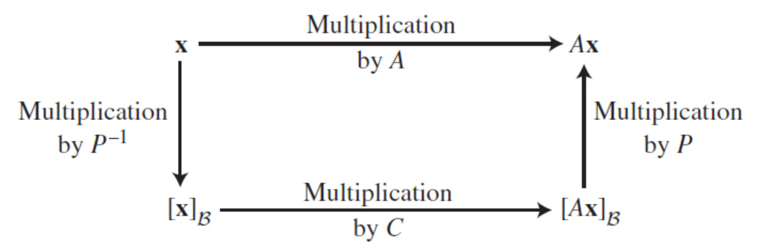
Understanding D as the transformation matrix in some basis gives us insight on its effect (in this example, an anisotropic dilation).

Similar matrices

Definition 4.3 (Similar matrices)

*A and C are **similar matrices** iff there exists another matrix P such that $A = PCP^{-1}$. Given the transformation $T(\mathbf{x}) = A\mathbf{x}$, C is the B-matrix of the transformation T, when B is the basis defined by the columns of the matrix P.*

Conversely, if B is any basis and P is the matrix formed by the vectors in the basis B, then the B-matrix of the transformation T is $P^{-1}AP$.



Similar matrices

Example

Let $A = \begin{pmatrix} 4 & -9 \\ 4 & 8 \end{pmatrix}$, $T(\mathbf{x}) = A\mathbf{x}$ and $\mathbf{b}_1 = (3,2)$, $\mathbf{b}_2 = (2,1)$. A is not diagonalizable but the basis $B = \{\mathbf{b}_1, \mathbf{b}_2\}$ has the property that $[T]_B$ is triangular (it is said to be in Jordan form). According to the previous definition, the B-matrix of the transformation T is

$$[T]_B = P^{-1}AP = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} 4 & -9 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 0 & -2 \end{pmatrix}$$

Numerical note

An easy way to compute $P^{-1}AP$ once we have AP is to find a row equivalent matrix

$$\left(\begin{array}{c|c} P & AP \end{array} \right) \sim \left(\begin{array}{c|c} I & P^{-1}AP \end{array} \right)$$

Exercises

From Lay (3rd ed.), Chapter 5, Section 4:

- 5.4.1
- 5.4.3
- 5.4.5
- 5.4.13
- 5.4.18
- 5.4.22
- 5.4.23
- 5.4.25
- 5.4.27 (computer)

Outline

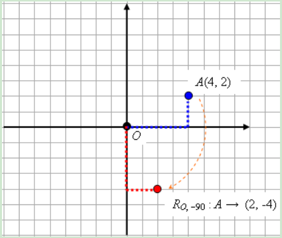
6. Eigenvalues and eigenvectors
   - Definition (a)
   - Characteristic equation (a)
   - Diagonalization (b)
   - Eigenvectors and linear transformations (b)
   - Complex eigenvalues (c)

Complex eigenvalues

Complex eigenvalues are always related to a rotation around a certain axis.

Example

Consider the linear transformation $T(\mathbf{x}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{x}$ is a rotation of 90°.



Obviously, there cannot be any real eigenvector since all the vectors are rotating. All eigenvalues are complex:

$$|A - \lambda I| = 0 = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$$

Complex eigenvalues

Example (continued)

Let's see what happens if we allow applying the transformation on complex vectors:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = i \begin{pmatrix} 1 \\ -i \end{pmatrix}$$
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} = -i \begin{pmatrix} 1 \\ i \end{pmatrix}$$

Complex eigenvalues

Example

Find the eigenvalues and eigenvectors of $A = \begin{pmatrix} \frac{1}{2} & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} \end{pmatrix}$.

Solution

To find the eigenvalues we solve the characteristic equation:

$$0 = |A - \lambda I| = \begin{vmatrix} \frac{1}{2} - \lambda & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} - \lambda \end{vmatrix} = \lambda^2 - \frac{8}{5}\lambda + 1 \Rightarrow \lambda = \frac{4}{5} \pm \frac{3}{5} i$$

MATLAB: `A=[1/2 -3/5; 3/4 11/10]; l=eigs(A)`

Complex eigenvalues

Example (continued)

$\underline{\lambda_1 = \frac{4}{5} - \frac{3}{5}i}$

$$
\begin{aligned}
A - \lambda_1 I &= \begin{pmatrix} \frac{1}{2} - (\frac{4}{5} - \frac{3}{5}i) & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} - (\frac{4}{5} - \frac{3}{5}i) \end{pmatrix} = \begin{pmatrix} -\frac{3}{10} + \frac{3}{5}i & -\frac{3}{5} \\ \frac{3}{4} & \frac{3}{10} + \frac{3}{5}i \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & \frac{2}{5} + \frac{4}{5}i \\ 0 & 0 \end{pmatrix} \Rightarrow x_1 = -(\frac{2}{5} + \frac{4}{5}i)x_2 \Rightarrow \mathbf{v}_1 = \begin{pmatrix} -2 - 4i \\ 5 \end{pmatrix}
\end{aligned}
$$

MATLAB:

```
A_lI=A-l(1)*eye(2);
A_lI(1,:)=A_lI(1,:)/A_lI(1,1)
A_lI(2,:)=A_lI(2,:)-A_lI(1,:)*A_lI(2,1)
```

$\underline{\lambda_2 = \frac{4}{5} + \frac{3}{5}i = \lambda_1^*}$

$$
A - \lambda_2 I \sim \begin{pmatrix} 1 & \frac{2}{5} - \frac{4}{5}i \\ 0 & 0 \end{pmatrix} \Rightarrow x_1 = -(\frac{2}{5} - \frac{4}{5}i)x_2 \Rightarrow \mathbf{v}_2 = \begin{pmatrix} -2 + 4i \\ 5 \end{pmatrix} = \mathbf{v}_1^*
$$

Complex eigenvalues

Example (continued)

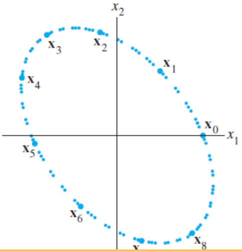
The application of A on $\mathbb{R}^2$ is a rotation. To see this, we may start with $\mathbf{x}_0 = (2, 0)$ and calculate

$$\mathbf{x}_1 = A\mathbf{x}_0 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix}$$

$$\mathbf{x}_2 = A\mathbf{x}_1 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}\begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} = \begin{bmatrix} -.4 \\ 2.4 \end{bmatrix}$$

$$\mathbf{x}_3 = A\mathbf{x}_2, \ldots$$

Figure 1 shows $\mathbf{x}_0, \ldots, \mathbf{x}_8$ as larger dots. The smaller dots are the locations of $\mathbf{x}_9, \ldots, \mathbf{x}_{100}$. The sequence lies along an elliptical orbit. ■



Complex eigenvalues

Definition 5.1 (Conjugate of a vector and matrix)

*The **conjugate of a vector** is defined as*

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix} \Rightarrow \mathbf{v}^* = \begin{pmatrix} v_1^* \\ v_2^* \\ \dots \\ v_n^* \end{pmatrix}$$

*In the same way, the **conjugate of a matrix** is defined as*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \Rightarrow A^* = \begin{pmatrix} a_{11}^* & a_{12}^* & \dots & a_{1n}^* \\ a_{21}^* & a_{22}^* & \dots & a_{2n}^* \\ \dots & \dots & \dots & \dots \\ a_{m1}^* & a_{m2}^* & \dots & a_{mn}^* \end{pmatrix}$$

Theorem 5.1 (Properties)

$(r\mathbf{v})^* = r^*\mathbf{v}^*$

$(A\mathbf{v})^* = A^*\mathbf{v}^*$

$(AB)^* = A^*B^*$

$(rA)^* = r^*A^*$

Eigenanalysis of a real matrix that acts on $\mathbb{C}^n$

Theorem 5.2

Let $A \in \mathcal{M}_{n\times n}$ be a matrix with real coefficients. If λ is an eigenvalue of A, then λ^* is also an eigenvalue. If $\mathbf{v}$ is an eigenvector associated to λ, then $\mathbf{v}^*$ is an eigenvector associated to λ^*.

Proof

If λ is an eigenvalue and $\mathbf{v}$ one of its eigenvectors, then we know that

$$A\mathbf{v} = \lambda\mathbf{v}$$

If we now conjugate both sides

$$(A\mathbf{v})^* = (\lambda\mathbf{v})^* \Rightarrow A\mathbf{v}^* = \lambda^*\mathbf{v}^*$$

(Remind that A has real coefficients and that's why $A^* = A$).
The previous equation means that $\mathbf{v}^*$ is also an eigenvector of A and that λ^* is its eigenvalue.

Eigenanalysis of a real matrix that acts on $\mathbb{C}^n$

Example

Let $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Its eigenvalues are $\lambda = a \pm bi$ and the corresponding eigenvectors $\mathbf{v} = \begin{pmatrix} 1 \\ \pm i \end{pmatrix}$.

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} a+bi \\ b-ai \end{pmatrix} = (a+bi) \begin{pmatrix} 1 \\ -i \end{pmatrix}$$
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} a-bi \\ b+ai \end{pmatrix} = (a-bi) \begin{pmatrix} 1 \\ i \end{pmatrix}$$

In particular if $a = \cos(\phi)$ and $b = \sin(\phi)$, then we have a rotation matrix whose eigenvalues are

$$\begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \Rightarrow \lambda = \cos(\phi) \pm \sin(\phi) i = e^{\pm i\phi}$$

Eigenanalysis of a real matrix that acts on $\mathbb{C}^n$

Example on Slide 60 (continued)

Let $A = \begin{pmatrix} \frac{1}{2} & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} \end{pmatrix}$. Consider $\lambda_1 = \frac{4}{5} - \frac{3}{5}i$ and its corresponding eigenvector $\mathbf{v}_1 = (-2-4i, 5)$. Now, we construct the matrix

$$P = \begin{pmatrix} \mathrm{Re}\{\mathbf{v}_1\} & \mathrm{Im}\{\mathbf{v}_1\} \end{pmatrix} = \begin{pmatrix} -2 & -4 \\ 5 & 0 \end{pmatrix}$$

and make a change of basis to the basis whose vectors are the columns of P:

$$C = P^{-1}AP = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{pmatrix} = \begin{pmatrix} \cos(36.87^\circ) & -\sin(36.87^\circ) \\ \sin(36.87^\circ) & \cos(36.87^\circ) \end{pmatrix}$$

That is, C is a pure rotation and thanks to the change of basis we obtain an elliptical rotation as shown in Slide 62.

Eigenanalysis of a real matrix that acts on $\mathbb{C}^n$

Theorem 5.3

Let A be a real, 2×2 matrix with complex eigenvalue $\lambda = a - bi$ $(b \neq 0)$ and an associated eigenvector in $\mathbb{C}^2$. Then

$$A = PCP^{-1}$$

where

$$P = \begin{pmatrix} \mathrm{Re}\{\mathbf{v}\} & \mathrm{Im}\{\mathbf{v}\} \end{pmatrix}$$

and

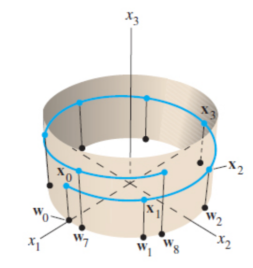
$$C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Proof
It makes use of

$$\mathrm{Re}\{A\mathbf{v}\} = A\mathrm{Re}\{\mathbf{v}\}$$
$$\mathrm{Im}\{A\mathbf{v}\} = A\mathrm{Im}\{\mathbf{v}\}$$

Example: Rotations extend to higher dimensions

Consider $A = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 1.07 \end{pmatrix}$. This is the rotation previously described in the XY plane plus a scaling in the Z direction. Any point in the XY (for instance, $\mathbf{w}_0 = (2,0,0)$) plane rotates within the plane. Any point outside the plane (for instance, $\mathbf{x}_0 = (2,0,1)$ rotates in XY and shifts along Z). The following figure shows the successive application of A on $\mathbf{w}_0$ and $\mathbf{x}_0$.



Exercises

Exercises

From Lay (3rd ed.), Chapter 5, Section 5:

- 5.5.1
- 5.5.7
- 5.5.13
- 5.5.23
- 5.5.24
- 5.5.25
- 5.5.26
- 5.5.27

Outline

6. Eigenvalues and eigenvectors
    - Definition (a)
    - Characteristic equation (a)
    - Diagonalization (b)
    - Eigenvectors and linear transformations (b)
    - Complex eigenvalues (c)

Chapter 7. Orthogonality and least squares

C.O.S. Sorzano

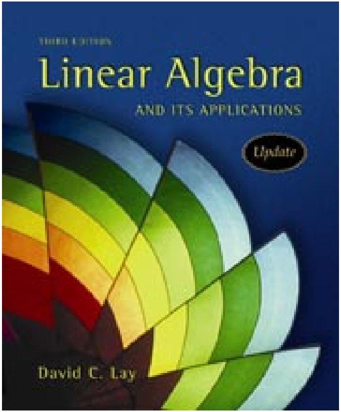
Biomedical Engineering

December 3, 2013

Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

References

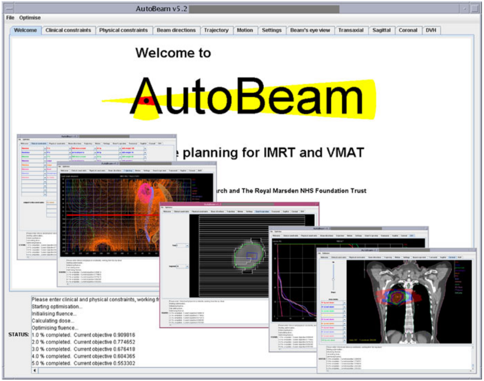


D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 6.

A little bit of history

Least squares was first used to solve problems in geodesy (Andrien-Marie Legendre, 1805) and astronomy (Carl Friedrich Gauss, 1809). Gauss made the connection of this method to the distribution of measurement errors. Currently it is one of the best understood and most widely spread methods.

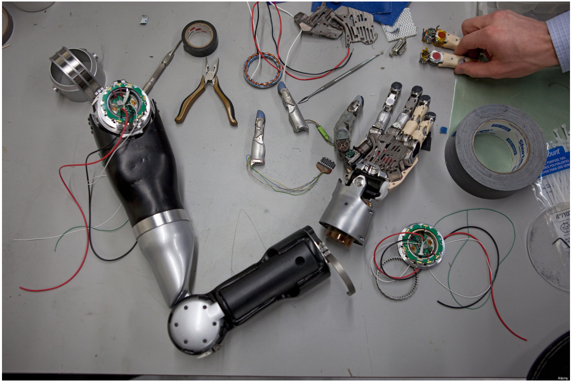
In this example Least Squares are used to plan a radiation therapy.



Bedford, J. L. *Sinogram analysis of aperture optimization by iterative least-squares in volumetric modulated arc therapy*. Physics in Medicine and Biology, **2013**, 58, 1235-1250

Applications

Traditionally, control applications were formulated in a least-squares setup. Currently, they have found more sophisticated goal functions that can be regarded as evolved versions of least squares.



Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Inner product

Definition 1.1 (Inner product or dot product)

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be two vectors. The ***inner product*** *or* ***dot product*** *between these two vectors is defined as*

$$\mathbf{u} \cdot \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle \triangleq \sum_{i=1}^{n} u_i v_i$$

Theorem 1.1

If we considered $\mathbf{u}$ and $\mathbf{v}$ to be column vectors ($\in \mathcal{M}_{n\times 1}$), then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

Example

Let $\mathbf{u} = (2, -5, -1)$ and $\mathbf{v} = (3, 2, -3)$.

$$\mathbf{u} \cdot \mathbf{v} = 2 \cdot 3 + (-5) \cdot 2 + 1 \cdot (-3) = -1$$

Inner product

Theorem 1.2

For any three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *and any scalar* $r \in \mathbb{R}$ *it is verified that*

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $(r\mathbf{u}) \cdot \mathbf{v} = r(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (r\mathbf{v})$
4. $\mathbf{u} \cdot \mathbf{u} \geq 0$
5. $\mathbf{u} \cdot \mathbf{u} = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$

Corollary

$$(r_1\mathbf{u}_1 + r_2\mathbf{u}_2 + ... + r_p\mathbf{u}_p) \cdot \mathbf{v} = r_1(\mathbf{u}_1 \cdot \mathbf{v}) + r_2(\mathbf{u}_2 \cdot \mathbf{v}) + ... + r_p(\mathbf{u}_p \cdot \mathbf{v})$$

Length

Definition 1.2 (Length of a vector)

Given any vector $\mathbf{v}$, *its length is defined as*

$$\|\mathbf{v}\| \triangleq \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Theorem 1.3

Given any vector $\mathbf{v} \in \mathbb{R}^n$

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

Example

The length of $\mathbf{v} = (1, -2, 2, 0)$ is

$$\|\mathbf{v}\| = \sqrt{1^2 + (-2)^2 + 2^2 + 0^2} = 3$$

Length

Theorem 1.4

For any vector $\mathbf{v}$ *and any scalar* r *it is verified that*

$$\|r\mathbf{v}\| = |r|\|\mathbf{v}\|$$

Proof
It will be given only for $\mathbf{v} \in \mathbb{R}^n$*:*

$$\begin{aligned} \|r\mathbf{v}\| &= \sqrt{(rv_1)^2 + (rv_2)^2 + \ldots + (rv_n)^2} = \sqrt{r^2(v_1^2 + v_2^2 + \ldots + v_n^2)} \\ &= \sqrt{r^2}\sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} = |r|\|\mathbf{v}\| \end{aligned} \quad \text{(q.e.d.)}$$

Example (continued)

Find a vector of unit length that has the same direction as $\mathbf{v} = (1, -2, 2, 0)$.
Solution

$$\mathbf{u_v} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \left(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}, 0\right) \Rightarrow \|\mathbf{u_v}\| = \sqrt{\frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0} = 1$$

Distance

Definition 1.3 (Distance in $\mathbb{R}$)

The distance between any two numbers $a, b \in \mathbb{R}$ can be defined as

$$d(a, b) = |a - b|$$

Example

Calculate the distance between 2 and 8 as well as between -3 and 4.

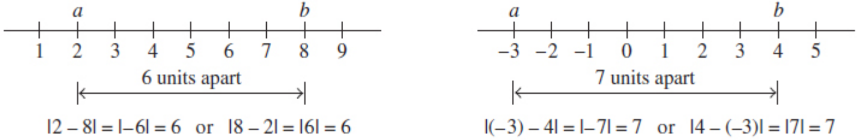


FIGURE 3 Distances in $\mathbb{R}$.

Distance

Definition 1.4 (Distance in $\mathbb{R}^n$)

The distance between any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *can be defined as*

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Example

Calculate the distance between $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$

$$d(\mathbf{u}, \mathbf{v}) = \|(7, 1) - (3, 2)\| = \|(4, -1)\| = \sqrt{4^2 + 1^2} = \sqrt{17}$$

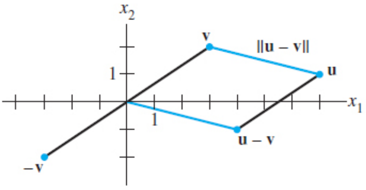


FIGURE 4 The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$.

Distance

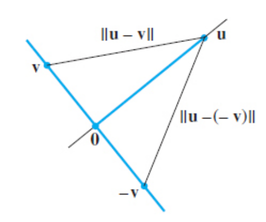
Example

For any two vectors in $\mathbb{R}^3$, $\mathbf{u}$ and $\mathbf{v}$, the distance can be calculated through

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = \|(u_1-v_1,u_2-v_2,u_3-v_3)\| = \sqrt{(u_1-v_1)^2+(u_2-v_2)^2+(u_3-v_3)^2}$$

Example

Any two vectors in $\mathbb{R}^2$, $\mathbf{u}$ and $\mathbf{v}$, are orthogonal if $d(\mathbf{u},\mathbf{v}) = d(\mathbf{u},-\mathbf{v})$



$$d^2(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|^2 = (\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) = \mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}-2\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|^2+\|\mathbf{v}\|^2-2\mathbf{u}\cdot\mathbf{v}$$
$$d^2(\mathbf{u},-\mathbf{v}) = \|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}+2\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|^2+\|\mathbf{v}\|^2+2\mathbf{u}\cdot\mathbf{v}$$

$$d^2(\mathbf{u},\mathbf{v}) = d^2(\mathbf{u},-\mathbf{v}) \Rightarrow -2\mathbf{u}\cdot\mathbf{v} = 2\mathbf{u}\cdot\mathbf{v} \Rightarrow \mathbf{u}\cdot\mathbf{v} = 0$$

Orthogonality

Definition 1.5 (Orthogonality between two vectors)

Any two different vectors, $\mathbf{u}$ *and* $\mathbf{v}$*, in a vector space* V *are* ***orthogonal*** *iff*

$$\mathbf{u} \cdot \mathbf{v} = 0$$

Corollary

$\mathbf{0}$ is orthogonal to any other vector.

Theorem 1.5 (Pythagorean theorem)

Any two vectors, $\mathbf{u}$ *and* $\mathbf{v}$*, in a vector space* V *are orthogonal iff*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Orthogonality

Definition 1.6 (Orthogonality between vector and vector space)

Let $\mathbf{u}$ *be a vector in a vector space* V *and* W *a vector subspace of* V*.* $\mathbf{u}$ *is* ***orthogonal*** *to* W *if* $\mathbf{u}$ *is orthogonal to all vectors in* W*. The set of all vectors orthogonal to* W *is denoted as* $W^{\perp}$ *(the* ***orthogonal complement*** *of* W*).*

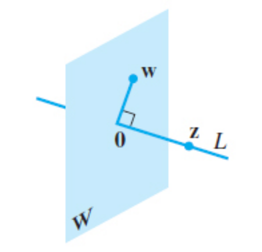
Example

Let W be a plane in $\mathbb{R}^3$ passing through the origin and L be a line, passing through the origin and perpendicular to W. For any vector $\mathbf{w} \in W$ and any vector $\mathbf{z} \in L$ we have

$$\mathbf{w} \cdot \mathbf{z} = 0$$

Therefore,

$$L = W^{\perp} \Leftrightarrow W = L^{\perp}$$



Orthogonality

Theorem 1.6

Let W be a vector subspace of a vector space V.

1. $\mathbf{x} \in W^{\perp}$ *iff* $\mathbf{x}$ *is orthogonal to every vector in a set that spans* W.
2. $W^{\perp}$ *is a vector subspace of* V.

Theorem 1.7

Let $A \in \mathcal{M}_{m\times n}$, then

1. $(\mathrm{Row}\{A\})^{\perp} = \mathrm{Nul}\{A\}$
2. $(\mathrm{Col}\{A\})^{\perp} = \mathrm{Nul}\{A^T\}$

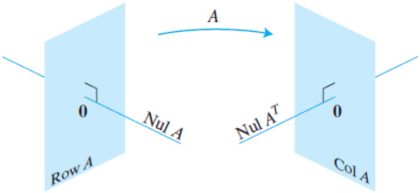


FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Orthogonality

Proof $\mathrm{Nul}\{A\} \subseteq (\mathrm{Row}\{A\})^\perp$

Consider the rows of A, $\mathbf{a}_i$ ($i = 1, 2, ..., m$) as column vectors, then for any vector $\mathbf{x} \in \mathrm{Nul}\{A\}$ we know

$$A\mathbf{x} = \mathbf{0} \Rightarrow \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \ldots \\ \mathbf{a}_m^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T \mathbf{x} \\ \mathbf{a}_2^T \mathbf{x} \\ \ldots \\ \mathbf{a}_m^T \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \mathbf{a}_2 \cdot \mathbf{x} \\ \ldots \\ \mathbf{a}_m \cdot \mathbf{x} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}$$

Consequently, $\mathbf{x}$ is orthogonal to all the rows of A, which span $\mathrm{Row}\{A\}$ and by the previous theorem, $\mathbf{x} \in (\mathrm{Row}\{A\})^\perp$

Proof $\mathrm{Nul}\{A\} \supseteq (\mathrm{Row}\{A\})^\perp$

Conversely, let $\mathbf{x} \in (\mathrm{Row}\{A\})^\perp$, then by the previous theorem we know that

$$\mathbf{a}_i \cdot \mathbf{x} \quad \text{for } i = 1, 2, ..., m \Rightarrow A\mathbf{x} = \mathbf{0}$$

So, $\mathbf{x} \in \mathrm{Nul}\{A\}$

Proof $(\mathrm{Col}\{A\})^\perp = \mathrm{Nul}\{A^T\}$

Let's define $B = A^T$. By the first part of this theorem, we know

$$(\mathrm{Row}\{B\})^\perp = \mathrm{Nul}\{B\} \Rightarrow (\mathrm{Row}\{A^T\})^\perp = \mathrm{Nul}\{A^T\} \Rightarrow (\mathrm{Col}\{A\})^\perp = \mathrm{Nul}\{A^T\}$$

Theorem 1.8

For any two vectors $\mathbf{u}$ *and* $\mathbf{v}$ *in a vector space* V*, the angle between the two can be measured through the dot product:*

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 1:

- 6.1.15
- 6.1.22
- 6.1.24
- 6.1.26
- 6.1.28
- 6.1.30
- 6.1.32 (computer)

Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Orthogonal sets

Definition 2.1 (Orthogonal set)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ be a set of vectors. S is an orthogonal set iff

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0 \quad \forall i, j \in \{1, 2, ..., p\}\ i \neq j$$

Example

Let $\mathbf{u}_1 = (3, 1, 1)$, $\mathbf{u}_2 = (-1, 2, 1)$, $\mathbf{u}_3 = (-\frac{1}{2}, -2, \frac{7}{2})$. Check whether the set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthogonal.

Solution

$$\begin{aligned}
\mathbf{u}_1 \cdot \mathbf{u}_2 &= 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0 \\
\mathbf{u}_1 \cdot \mathbf{u}_3 &= 3 \cdot (-\tfrac{1}{2}) + 1 \cdot (-2) + 1 \cdot (\tfrac{7}{2}) = 0 \\
\mathbf{u}_2 \cdot \mathbf{u}_3 &= (-1) \cdot (-\tfrac{1}{2}) + 2 \cdot (-2) + 1 \cdot (\tfrac{7}{2}) = 0
\end{aligned}$$

Orthogonal sets

Theorem 2.1

If S is an orthogonal set of non-null vectors, then S is linearly independent and, consequently, it is a basis of the subspace spanned by S.

Proof

Let $\mathbf{u}_i$ $(i = 1, 2, ..., p)$ be the elements of S. Let us assume that S is linearly dependent. Then, there exists coefficients c_1, c_2, ..., c_p not all of them null such that

$$\mathbf{0} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p$$

Now, we compute the inner product with $\mathbf{u}_1$

$$\begin{gathered}\mathbf{0} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ 0 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + ... + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) = c_1\|\mathbf{u}_1\|^2 \Rightarrow c_1 = 0\end{gathered}$$

Multiplying by $\mathbf{u}_i$ $(i = 2, 3, ..., p)$ we can show that all c_i's are 0, and, therefore, the set S is linearly independent.

Orthogonal basis

Definition 2.2 (Orthogonal basis)

A set of vectors B is an ortohogonal basis of a vector space V if it is an ortohogonal set and it is a basis of V.

Theorem 2.2

Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ be an orthogonal basis for a vector space V, for each $\mathbf{x} \in V$ we have

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u_1}}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u_2}}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 + ... + \frac{\mathbf{x} \cdot \mathbf{u_p}}{\|\mathbf{u}_p\|^2}\mathbf{u}_p$$

Proof

If $\mathbf{x}$ is in V, then it can be expressed as a linear combination of the vectors in a basis of V

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p$$

Now, we calculate the dot product with $\mathbf{u}_1$

$$\mathbf{x} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1\|\mathbf{u}_1\|^2 \Rightarrow c_1 = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}$$

Orthogonal basis

Example

Let $\mathbf{u}_1 = (3, 1, 1)$, $\mathbf{u}_2 = (-1, 2, 1)$, $\mathbf{u}_3 = (-\frac{1}{2}, -2, \frac{7}{2})$, and $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be an orthogonal basis of $\mathbb{R}^3$. Let $\mathbf{x} = (6, 1, -8)$. The coordinates of $\mathbf{x}$ in B are given by

$$\begin{array}{ccc} \mathbf{x} \cdot \mathbf{u}_1 = 11 & \mathbf{x} \cdot \mathbf{u}_2 = -12 & \mathbf{x} \cdot \mathbf{u}_1 = -33 \\ \|\mathbf{u}_1\|^2 = 11 & \|\mathbf{u}_2\|^2 = 6 & \|\mathbf{u}_3\|^2 = \frac{33}{2} \end{array}$$

$$\begin{aligned} \mathbf{x} &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{\frac{33}{2}}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

The coordinates of $\mathbf{x}$ in the basis B are

$$[\mathbf{x}]_B = (1, -2, -2)$$

Orthogonal projections

Orthogonal projection onto a vector

Consider a vector $\mathbf{y}$ and another one $\mathbf{u}$. Let us assume we want to decompose $\mathbf{y}$ as the sum of two orthogonal vectors $\hat{\mathbf{y}}$ (along the direction of $\mathbf{u}$) and another vector $\mathbf{z}$ (orthogonal to $\mathbf{u}$):

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} = \alpha\mathbf{u} + \mathbf{z} \Rightarrow$$
$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$

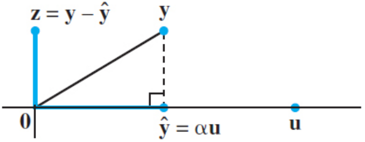


FIGURE 2

Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

We need to find α that makes $\mathbf{u}$ and $\mathbf{z}$ orthogonal.

$$0 = \mathbf{z}\cdot\mathbf{u} = (\mathbf{y} - \alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - \alpha\|\mathbf{u}\|^2 \Rightarrow \alpha = \frac{\mathbf{y}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}$$

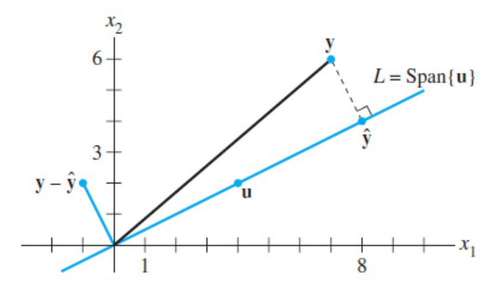
$\hat{\mathbf{y}}$ is the **orthogonal projection** of $\mathbf{y}$ onto $\mathbf{u}$.

Orthogonal projections

Example

Let $\mathbf{y} = (7,6)$ and $\mathbf{u} = (4,2)$. Then,

$$\left.\begin{array}{r} \mathbf{y}\cdot\mathbf{u} = 40 \\ \|\mathbf{u}\|^2 = 20 \end{array}\right\} \Rightarrow \qquad \begin{array}{c} \hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{40}{20}\mathbf{u} = 2\mathbf{u} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \\ \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} 7 \\ 6 \end{pmatrix} - \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\ d(\mathbf{y},\hat{\mathbf{y}}) = \|\mathbf{y}-\hat{\mathbf{y}}\| = \|\mathbf{z}\| = \sqrt{(-1)^2+2^2} = \sqrt{5} \end{array}$$



Orthonormal set

Definition 2.3 (Orthonormal set)

$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ *is an **orthonormal set** if it is an orthogonal set and all* $\mathbf{u}_i$ *vectors have unit length.*

Example

Show that the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthonormal, with

$$\mathbf{u}_1 = \frac{1}{\sqrt{11}}\begin{pmatrix}3\\1\\1\end{pmatrix} \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\1\end{pmatrix} \quad \mathbf{u}_3 = \frac{1}{\sqrt{66}}\begin{pmatrix}-1\\-4\\7\end{pmatrix}$$

Solution

Let's check that they are orthogonal:

$$\begin{aligned}
\mathbf{u}_1 \cdot \mathbf{u}_2 &= \frac{1}{\sqrt{11}}\frac{1}{\sqrt{6}}(3\cdot(-1)+1\cdot 2+1\cdot 1)=0\\
\mathbf{u}_1 \cdot \mathbf{u}_3 &= \frac{1}{\sqrt{11}}\frac{1}{\sqrt{66}}(3\cdot(-1)+1\cdot(-4)+1\cdot 7)=0\\
\mathbf{u}_2 \cdot \mathbf{u}_3 &= \frac{1}{\sqrt{6}}\frac{1}{\sqrt{66}}((-1)\cdot(-1)+(2)\cdot(-4)+(1)\cdot 7)=0
\end{aligned}$$

Orthonormal set

Example (continued)

Now, let's check that they have unit length:

$$\|\mathbf{u}_1\| = \sqrt{\left(\frac{1}{\sqrt{11}}\right)^2 (3^2 + 1^2 + 1^2)} = \sqrt{\frac{9+1+1}{11}} = 1$$

$$\|\mathbf{u}_2\| = \sqrt{\left(\frac{1}{\sqrt{6}}\right)^2 ((-1)^2 + 2^2 + 1^2)} = \sqrt{\frac{1+4+1}{6}} = 1$$

$$\|\mathbf{u}_3\| = \sqrt{\left(\frac{1}{\sqrt{66}}\right)^2 ((-1)^2 + (-4)^2 + 7^2)} = \sqrt{\frac{1+16+49}{66}} = 1$$

Theorem 2.3

If $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal set, then it is an orthonormal basis of $\mathrm{Span}\{S\}$.

Example

$\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ is an orthonormal basis of $\mathbb{R}^n$.

Orthonormal basis

Theorem 2.4

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *is an orthogonal set of vectors, then the set* $S' = \{\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_n\}$ *where*

$$\mathbf{u}'_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$$

is a orthonormal set (this operation is called ***vector normalization****).*

Proof

Let's check that the $\mathbf{u}'_i$ *vectors are orthogonal:*

$$\mathbf{u}'_i \cdot \mathbf{u}'_j = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|} \cdot \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|} = \frac{1}{\|\mathbf{u}_i\|\|\mathbf{u}_j\|}\mathbf{u}_i \cdot \mathbf{u}_j$$

But this product is obviusly 0 because the $\mathbf{u}_i$ *vectors are orthogonal. Let's check now that the* $\mathbf{u}'_i$ *vectors have unit length:*

$$\|\mathbf{u}'_i\| = \left\|\frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}\right\| = \frac{\|\mathbf{u}_i\|}{\|\mathbf{u}_i\|} = 1$$

Orthonormal matrix

Theorem 2.5

Let $U \in \mathcal{M}_{m\times n}$ be a square matrix. The columns of U form an orthonormal set iff

$$U^T U = I_n$$

It is said that U is an ***orthonormal matrix****.*

Proof

Let's consider the columns of U

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}$$

Let's calculate now $U^T U$

$$U^T U = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \ldots \\ \mathbf{u}_n^T \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \ldots & \mathbf{u}_1^T\mathbf{u}_n \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \ldots & \mathbf{u}_2^T\mathbf{u}_n \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{u}_n^T\mathbf{u}_1 & \mathbf{u}_n^T\mathbf{u}_2 & \ldots & \mathbf{u}_n^T\mathbf{u}_n \end{pmatrix}$$

The condition $U^T U = I_n$ simply states $\left\{ \begin{array}{ll} \mathbf{u}_i^T\mathbf{u}_j = 0 & i \neq j \\ \mathbf{u}_i^T\mathbf{u}_j = 1 & i = j \end{array} \right.$*, which is the definition of an orthonormal set.*

Orthonormal matrix

Theorem 2.6

Let $U \in \mathcal{M}_{n\times n}$ *be an orthonormal matrix and* $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *then*

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0 \Leftrightarrow \mathbf{x} \cdot \mathbf{y} = 0$

Example

Let $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix}$.

U is an orthonormal matrix because

$$U^T U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Orthonormal matrix

Example (continued)

Let's calculate now $U\mathbf{x}$

$$U\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$$

Let's check now that $\|U\mathbf{x}\| = \|\mathbf{x}\|$

$$\|U\mathbf{x}\| = \|(3,-1,1)\| = \sqrt{3^2 + (-1)^2 + 1^2} = \sqrt{11}$$
$$\|\mathbf{x}\| = \left\|(\sqrt{2},3)\right\| = \sqrt{(\sqrt{2})^2 + 3^2} = \sqrt{11}$$

Theorem 2.7

Let U be an orthonormal and square matrix. Then,

1. $U^{-1} = U^T$
2. U^T *is also an orthonormal matrix (i.e., the rows of U also form an orthonormal set of vectors).*

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 2:

- 6.2.1
- 6.2.10
- 6.2.15
- 6.2.25
- 6.2.26
- 6.2.29
- 6.2.35 (computer)

Outline

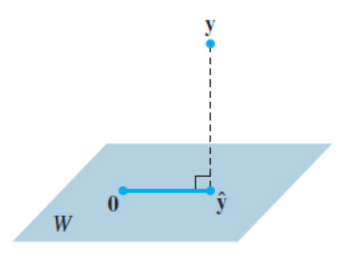
7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Orthogonal projections

Definition 3.1 (Orthogonal projection)

The orthogonal projection of a point $\mathbf{y}$ *onto a vector subspace* W *is a point* $\hat{\mathbf{y}} \in W$ *such that*

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$
$$\mathbf{z} \perp W$$



Orthogonal projections

Example

Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_5\}$ be an orthogonal basis of $\mathbb{R}^5$. Consider the subspace $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Given any vector $\mathbf{y} \in \mathbb{R}^5$, we can decompose it as the sum of a vector in W and a vector perpendicular to W

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

<u>*Solution*</u>

If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_5\}$ is a basis of $\mathbb{R}^5$, then any vector $\mathbf{y} \in \mathbb{R}^5$ can be written as

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_5\mathbf{u}_5$$

We may decompose this sum as

$$\hat{\mathbf{y}} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$
$$\mathbf{z} = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5$$

Orthogonal projections

Example (continued)

It is obvious that $\hat{\mathbf{y}} \in W$. Now we need to show that $\mathbf{z} \in W^{\perp}$. For doing so, we will show that

$$\begin{aligned} \mathbf{z} \cdot \mathbf{u}_1 &= 0 \\ \mathbf{z} \cdot \mathbf{u}_2 &= 0 \end{aligned}$$

To show the first equation we note that

$$\begin{aligned} \mathbf{z} \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3(\mathbf{u}_3 \cdot \mathbf{u}_1) + c_4(\mathbf{u}_4 \cdot \mathbf{u}_1) + c_5(\mathbf{u}_5 \cdot \mathbf{u}_1) \\ &= c_3 \cdot 0 + c_4 \cdot 0 + c_5 \cdot 0 \\ &= 0 \end{aligned}$$

We would proceed analogously for $\mathbf{z} \cdot \mathbf{u}_2 = 0$.

Orthogonal projections

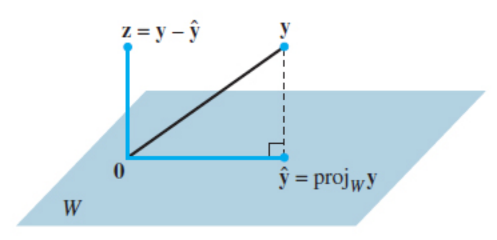
Theorem 3.1 (Orthogonal Decomposition Theorem)

Let W be a vector subspace of a vector space V. Then, any vector $\mathbf{y} \in V$ can be written uniquely as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

with $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact, if $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ is an orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 + ... + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\|\mathbf{u}_p\|^2}\mathbf{u}_p$$



Orthogonal projections

Proof

$\hat{\mathbf{y}}$ is obviously in W since it has been written as a linear combination of vectors in a basis of W. $\mathbf{z}$ is perpendicular to W because

$$\begin{aligned}
\mathbf{z}\cdot\mathbf{u}_1 &= \left(\mathbf{y} - \left(\frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 + \ldots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\|\mathbf{u}_p\|^2}\mathbf{u}_p\right)\right)\cdot\mathbf{u}_1 \\
&= \mathbf{y}\cdot\mathbf{u}_1 - \frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}(\mathbf{u}_1\cdot\mathbf{u}_1) - \frac{\mathbf{y}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}(\mathbf{u}_2\cdot\mathbf{u}_1) - \ldots - \frac{\mathbf{y}\cdot\mathbf{u}_p}{\|\mathbf{u}_p\|^2}(\mathbf{u}_p\cdot\mathbf{u}_1) \\
&\quad [\{\mathbf{u}_i\} \text{ is an orthogonal set}] \\
&= \mathbf{y}\cdot\mathbf{u}_1 - \frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}(\mathbf{u}_1\cdot\mathbf{u}_1) \\
&= \mathbf{y}\cdot\mathbf{u}_1 - \frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\|\mathbf{u}_1\|^2 \\
&= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 \\
&= 0
\end{aligned}$$

We could proceed analogously for all elements in the basis of W.

Orthogonal projections

We need to show now that the decomposition is unique. Let us assume that it is not unique. Consequently, there exist different vectors such that

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$
$$\mathbf{y} = \hat{\mathbf{y}}' + \mathbf{z}'$$

We subtract both equations

$$\mathbf{0} = (\hat{\mathbf{y}} - \hat{\mathbf{y}}') + (\mathbf{z} - \mathbf{z}') \Rightarrow \hat{\mathbf{y}} - \hat{\mathbf{y}}' = \mathbf{z}' - \mathbf{z}$$

Let $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}'$. It is obvious that $\mathbf{v} \in W$ because it is written as a linear combination of vectors in W. On the other side, $\mathbf{v} = \mathbf{z}' - \mathbf{z}$, i.e., it is a linear combination of vectors in $W^\perp$, so $\mathbf{v} \in W^\perp$. The only vector that belongs to W and $W^\perp$ at the same time is

$$\mathbf{v} = \mathbf{0} \Rightarrow \begin{cases} \hat{\mathbf{y}} = \hat{\mathbf{y}}' \\ \mathbf{z} = \mathbf{z}' \end{cases}.$$

and consequently, the orthogonal decomposition is unique. Additionally, the uniqueness of the decomposition depends only on W and not on the particular basis chosen for W.

Orthogonal projections

Example

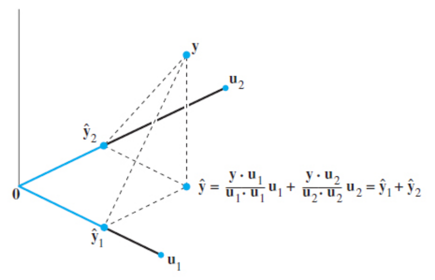
Let $\mathbf{u}_1 = (2, 5, -1)$ and $\mathbf{u}_2 = (-2, 1, 1)$. Let W be the subspace spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Let $\mathbf{y} = (1, 2, 3) \in \mathbb{R}^3$. The orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned}
\hat{\mathbf{y}} &= \tfrac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \tfrac{\mathbf{y}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 \\
&= \tfrac{1\cdot 2+2\cdot 5+3\cdot(-1)}{2^2+5^2+(-1)^2}\begin{pmatrix}2\\5\\-1\end{pmatrix} + \tfrac{1\cdot(-2)+2\cdot 1+3\cdot 1}{(-2)^2+1^2+1^2}\begin{pmatrix}-2\\1\\1\end{pmatrix} \\
&= \tfrac{9}{30}\begin{pmatrix}2\\5\\-1\end{pmatrix} + \tfrac{15}{30}\begin{pmatrix}-2\\1\\1\end{pmatrix} = \begin{pmatrix}-\frac{2}{5}\\2\\\frac{1}{5}\end{pmatrix} \\
\mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix}1\\2\\3\end{pmatrix} - \begin{pmatrix}-\frac{2}{5}\\2\\\frac{1}{5}\end{pmatrix} = \begin{pmatrix}\frac{7}{5}\\0\\\frac{14}{5}\end{pmatrix}
\end{aligned}$$

Orthogonal projections

Geometrical interpretation

$\hat{\mathbf{y}}$ can be understood as the sum of the orthogonal projection of $\mathbf{y}$ onto each one of the elements of the basis of W.



Theorem 3.2

If $\mathbf{y}$ *belongs to* W*, then the orthogonal projection of* $\mathbf{y}$ *onto* W *is itself:*

$$\hat{\mathbf{y}} = \mathbf{y}$$

Properties of orthogonal projections

Theorem 3.3 (Best approximation theorem)

The orthogonal projection of $\mathbf{y}$ *onto* W *is the point in* W *with minimum distance to* $\mathbf{y}$, *i.e.,*

$$\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{v}\|$$

for all $\mathbf{v} \in W, \mathbf{v} \neq \hat{\mathbf{y}}$.

Proof

We know that $\mathbf{y} - \hat{\mathbf{y}}$ *is orthogonal to* W. *For any vector* $\mathbf{v} \in W, \mathbf{v} \neq \hat{\mathbf{y}}$, *we have that* $\hat{\mathbf{y}} - \mathbf{v}$ *is in* W. *Now consider the orthogonal decomposition of the vector* $\mathbf{y} - \mathbf{v}$

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

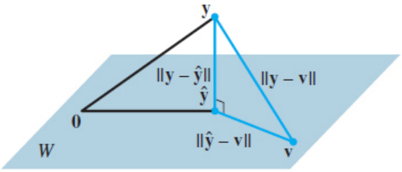


FIGURE 4 The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

Properties of orthogonal projections

Due to the orthogonal decomposition theorem (Theorem 3.1), this decomposition is unique and due to the Pythagorean theorem (Theorem 1.5) we have

$$\|\mathbf{y}-\mathbf{v}\|^2 = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}-\mathbf{v}\|^2$$

Since $\mathbf{v} \neq \hat{\mathbf{y}}$ we have $\|\hat{\mathbf{y}}-\mathbf{v}\|^2 > 0$ and consequently

$$\|\mathbf{y}-\mathbf{v}\|^2 > \|\mathbf{y}-\hat{\mathbf{y}}\|^2$$

Properties of orthogonal projections

Theorem 3.4

If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ *is an orthonormal basis of* W*, then the orthogonal projection of* $\mathbf{y}$ *onto* W *is*

$$\hat{\mathbf{y}} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + ... + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

If we construct the orthonormal matrix $U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & ... & \mathbf{u}_p \end{pmatrix}$*, then*

$$\hat{\mathbf{y}} = UU^T\mathbf{y}$$

Proof
By Theorem 3.1 we know that for all orthogonal bases it is verified

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 + ... + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\|\mathbf{u}_p\|^2}\mathbf{u}_p$$

Since the basis is in this case orthonormal, then $\|\mathbf{u}\| = 1$ *and consequently*

$$\hat{\mathbf{y}} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + ... + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

Properties of orthogonal projections

On the other side we have

$$U^T\mathbf{y} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \ldots \\ \mathbf{u}_p^T \end{pmatrix} \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{y} \\ \mathbf{u}_2^T\mathbf{y} \\ \ldots \\ \mathbf{u}_p^T\mathbf{y} \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{y} \rangle \\ \langle \mathbf{u}_2, \mathbf{y} \rangle \\ \ldots \\ \langle \mathbf{u}_p, \mathbf{y} \rangle \end{pmatrix}$$

Then,

$$UU^T\mathbf{y} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_p \end{pmatrix} \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{y} \rangle \\ \langle \mathbf{u}_2, \mathbf{y} \rangle \\ \ldots \\ \langle \mathbf{u}_p, \mathbf{y} \rangle \end{pmatrix} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \ldots + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

(q.e.d.)

Properties of orthogonal projections

Corollary

Let $U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_p \end{pmatrix}$ be a $n \times p$ matrix with orthonormal columns and $W = \mathrm{Col}\{U\}$ its column space. Then,

$$\begin{array}{lll} \forall \mathbf{x} \in \mathbb{R}^p & U^T U \mathbf{x} = \mathbf{x} & \text{No effect} \\ \forall \mathbf{y} \in \mathbb{R}^n & U U^T \mathbf{y} = \hat{\mathbf{y}} & \text{Orthogonal projection of } \mathbf{y} \text{ onto } W \end{array}$$

If U is a $n \times n$, then $W = \mathbb{R}^n$ and the projection has no effect

$$\forall \mathbf{y} \in \mathbb{R}^n \quad U U^T \mathbf{y} = \hat{\mathbf{y}} = \mathbf{y} \quad \text{No effect}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 3:

- 6.3.1
- 6.3.7
- 6.3.15
- 6.3.23
- 6.3.24
- 6.3.25 (computer)

Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Gram-Schmidt orthogonalization

Gram-Schmidt orthogonalization is a procedure aimed at producing an orthogonal basis of any subspace W.

Example

Let $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with $\mathbf{x}_1 = (3,6,0)$ and $\mathbf{x}_2 = (1,2,2)$. Let's look for an orthogonal basis of W.

Solution

We may keep the first vector for the basis

$$\mathbf{v}_1 = \mathbf{x}_1 = (3,6,0)$$

For the second vector in the basis, we need to keep the component of $\mathbf{x}_2$ that is orthogonal to $\mathbf{x}_1$. For doing so we calculate the projection of $\mathbf{x}_2$ onto $\mathbf{x}_1$ ($\mathbf{p}$), and we decompose $\mathbf{x}_2$ as

$$\mathbf{x}_2 = \mathbf{p} + (\mathbf{x}_2 - \mathbf{p}) = (1,2,0) + (0,0,2)$$

We, then, keep the orthogonal part of $\mathbf{x}_2$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = (0,0,2)$$

Gram-Schmidt orthogonalization

Example (continued)

The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis of W.

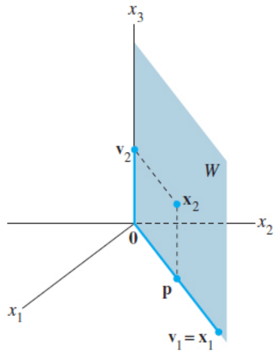


FIGURE 1
Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

Gram-Schmidt orthogonalization

Example

Let $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with $\mathbf{x}_1 = (1, 1, 1, 1)$, $\mathbf{x}_2 = (0, 1, 1, 1)$ and $\mathbf{x}_3 = (0, 0, 1, 1)$. Let's look for an orthogonal basis of W.

Solution

We may keep the first vector for the basis. Then we construct a subspace (W_1) with a single element in its basis

$$\mathbf{v}_1 = \mathbf{x}_1 = (1, 1, 1, 1) \quad W_1 = \mathrm{Span}\{\mathbf{v}_1\}$$

For the second vector in the basis, we need to keep the component of $\mathbf{x}_2$ that is orthogonal to W_1. With the already computed basis vectors, we construct a new subspace (W_2) with two elements in its basis

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{W_1}(\mathbf{x}_2) = (-\tfrac{3}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) \quad W_2 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

For the third vector in the basis, we repeat the same procedure

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{Proj}_{W_2}(\mathbf{x}_3) = (0, -\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{1}{3}) \quad W_3 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

Gram-Schmidt orthogonalization

Theorem 4.1 (Gram-Schmidt orthogonalization)

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p\}$ *for a vector subspace* W. *Define*

$$\begin{array}{rl} \mathbf{v}_1 = \mathbf{x}_1 & W_1 = \mathrm{Span}\{\mathbf{v}_1\} \\ \mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{W_1}(\mathbf{x}_2) & W_2 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\} \\ \ldots & \\ \mathbf{v}_p = \mathbf{x}_p - \mathrm{Proj}_{W_{p-1}}(\mathbf{x}_p) & W_p = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\} = W \end{array}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\}$ *is an orthogonal basis of* W.

Gram-Schmidt orthogonalization

Proof

Consider $W_k = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ and let us assume that $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is a basis of W_k. Now we construct

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \mathrm{Proj}_{W_k}(\mathbf{x}_{k+1}) \quad W_{k+1} = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{k+1}\}$$

By the orthogonal decomposition theorem (Theorem 3.1), we know that $\mathbf{v}_{k+1}$ is orthogonal to W_k. Because $\mathbf{x}_{k+1}$ is an element of a basis, we know that $\mathbf{x}_{k+1} \notin W_k$. Therefore, $\mathbf{v}_{k+1}$ is not null and $\mathbf{x}_{k+1} \in W_{k+1}$. Finally, the set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{k+1}\}$ is a set of orthogonal, non-null vectors in the $(k+1)$-dimensional space W_{k+1}. Consequently, by Theorem 9.4 in Chapter 5, it must be a basis of W_{k+1}. This process can be iterated till $k = p$.

Gram-Schmidt orthogonalization

Orthonormal basis

Once we have an orthogonal basis, we simply have to normalize each vector to have an orthonormal basis.

Example

Let $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with $\mathbf{x}_1 = (3, 6, 0)$ and $\mathbf{x}_2 = (1, 2, 2)$. Let's look for an orthonormal basis of W.

Solution

In Slide 52 we learned that an orthogonal basis was given by

$$\begin{aligned} \mathbf{v}_1 &= (3, 6, 0) \\ \mathbf{v}_2 &= (0, 0, 2) \end{aligned}$$

Now, we normalize these two vectors to obtain an orthonormal basis

$$\begin{aligned} \mathbf{v}'_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{45}}(3, 6, 0) = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, 0\right) \\ \mathbf{v}'_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{2}(0, 0, 2) = (0, 0, 1) \end{aligned}$$

QR factorization of matrices

If we apply the Gram-Schmidt factorization to the columns of a matrix, we have the following factorization scheme. This factorization is used in practice to find eigenvalues and eigenvectors as well as to solve linear equation systems.

Theorem 4.2 (QR factorization)

Let $A \in \mathcal{M}_{m\times n}$ with linearly independent columns. Then, A can be factored as

$$A = QR$$

where $Q \in \mathcal{M}_{m\times n}$ is a matrix whose columns form an orthonormal basis of $\mathrm{Col}\{A\}$ and $R \in \mathcal{M}_{n\times n}$ is an upper triangular invertible matrix with positive entries on its diagonal.

Proof

Let's orthogonalize the columns of A following the Gram-Schmidt procedure and construct the orthonormal basis of $\mathrm{Col}\{A\}$. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be such a basis. Let us construct the matrix

$$Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{pmatrix}$$

QR factorization of matrices

Let us call $\mathbf{a}_i$ $(i = 1, 2, ..., n)$ to the columns of A. By the Gram-Schmidt orthogonalization, we know that for any k between 1 and n we have

$$\mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k\} = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$$

Consequently, we can express each column of A in the orthonormal basis:

$$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + r_{2k}\mathbf{u}_2 + ... + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + ... + 0 \cdot \mathbf{u}_n$$

If r_{kk} is negative, we can multiply both r_{kk} and $\mathbf{u}_k$ by -1. We now collect all these coefficients in a vector $\mathbf{r}_k = (r_{1k}, r_{2k}, ..., r_{kk}, 0, 0, ..., 0)$ to have

$$\mathbf{a}_k = Q\mathbf{r}_k$$

By gathering all these vectors in a matrix, we have the triangular matrix R

$$R = \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & ... & \mathbf{r}_n \end{pmatrix}$$

R is invertible because the columns of A are linearly independent.

QR factorization of matrices

Example

Let's calculate the QR factorization of $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. From Slide 54 we know that the vectors

$$\begin{aligned} \mathbf{v}_1 &= (1, 1, 1, 1) \\ \mathbf{v}_2 &= (-\tfrac{3}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) \\ \mathbf{v}_3 &= (0, -\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{1}{3}) \end{aligned}$$

Is an orthogonal basis of the column space of A. We now normalize these vectors to obtain the orthonormal basis in Q

$$Q = \begin{pmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

Example (continued)

To find R we multiply on both sides of the factorization by Q^T

$$A = QR \Rightarrow Q^T A = Q^T QR = R$$

$$\begin{aligned} R = Q^T A \quad &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{2}{\sqrt{12}} \\ 0 & 0 & \frac{1}{\sqrt{6}} \end{pmatrix} \end{aligned}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 4:

- 6.4.7
- 6.4.13
- 6.4.19
- 6.4.22
- 6.4.24

Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Let's assume we want to solve the equation system $A\mathbf{x} = \mathbf{b}$, but, due to noise, there is no solution. We may still look for a solution such that $A\mathbf{x} \approx \mathbf{b}$. In fact the goal will be to minimize $d(A\mathbf{x}, \mathbf{b})$.

Definition 5.1 (Least squares solution)

*Let A be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. $\mathbf{x} \in \hat{} \mathbb{R}^n$ is a **least squares solution** of the equation system $A\mathbf{x} = \mathbf{b}$ iff*

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

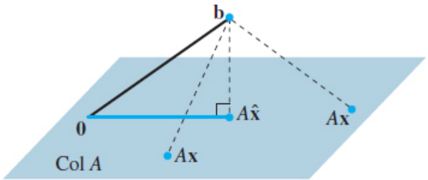


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Least squares

Solution of the general least squares problem

Applying the Best Approximation Theorem (Theorem 3.3), we may project **b** onto the column space of A

$$\hat{\mathbf{b}} = \mathrm{Proj}_{\mathrm{Col}\{A\}}\{\mathbf{b}\}$$

Then, we solve the equation system

$$A\mathbf{x} = \hat{\mathbf{b}}$$

that has at least one solution.

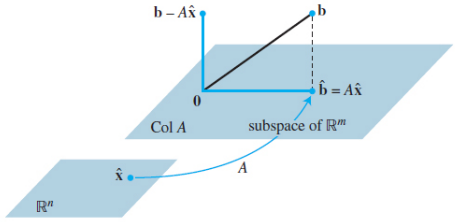


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Least squares

Theorem 5.1

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is the same as the set of solutions of the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

Proof: least-squares solutions $\subset$ normal equations solutions

Let us assume that $\hat{\mathbf{x}}$ is a least-squares solution. Then, $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathrm{Col}\{A\}$, and in particular, to each one of the columns of A ($\mathbf{a}_i$, $i = 1, 2, ..., n$):

$$\begin{gathered}
\mathbf{a}_i \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \forall i \in \{1, 2, ..., n\} \Rightarrow \\
\mathbf{a}_i^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \forall i \in \{1, 2, ..., n\} \Rightarrow \\
A^T (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \Rightarrow \\
A^T\mathbf{b} = A^T A\hat{\mathbf{x}}
\end{gathered}$$

That is, every least-squares solution is also a solution of the normal equations.

Proof: least-squares solutions $\supset$ normal equations solutions

Let us assume that $\hat{\mathbf{x}}$ is solution of the normal equations. Then,

$$A^T\mathbf{b} = A^T A\hat{\mathbf{x}} \Rightarrow$$
$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \Rightarrow$$
$$\mathbf{a}_i^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \forall i \in \{1, 2, ..., n\}$$

That is, $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the columns of A and, consequently, to $\mathrm{Col}\{A\}$. Hence, the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is the orthogonal decomposition of $\mathbf{b}$ as a vector in $\mathrm{Col}\{A\}$ and a vector orthogonal to $\mathrm{Col}\{A\}$. By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}\{A\}$ so that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}$$

and, therefore, $\hat{\mathbf{x}}$ is a least-squares solution.

Least squares

Example

Find a least-squares solution to $A\mathbf{x} = \mathbf{b}$ with $A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$.

Solution

Let's solve the normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$

$$A^T A = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \quad A^T\mathbf{b} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$\begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \hat{\mathbf{x}} = \begin{pmatrix} 19 \\ 11 \end{pmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 19 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Let's check that $\hat{\mathbf{x}}$ is not a solution of the original equation system but a least-squares solution

$$A\hat{\mathbf{x}} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix} = \hat{\mathbf{b}} \neq \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$$

Least squares

Definition 5.2 (Least-squares error)

The least-squares error is defined as

$$\sigma_\epsilon^2 \triangleq \|A\hat{\mathbf{x}} - \mathbf{b}\|^2 = \|\hat{\mathbf{b}} - \mathbf{b}\|^2$$

Example (continued)

In this case:

$$\sigma_\epsilon^2 = \|(4,4,3) - (2,0,11)\| = \|(2,4,-8)\| \approx 9.165$$

Example

Unfortunately, the least-squares solution may not be unique as shown in the next example (arising in ANOVA). Find a least-squares solution to $A\mathbf{x} = \mathbf{b}$ with

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix}.$$

Solution

$$A^T A = \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} \quad A^T\mathbf{b} = \begin{pmatrix} 4 \\ -4 \\ 2 \\ 6 \end{pmatrix}$$

Example (continued)

The augmented matrix is

$$\left(\begin{array}{cccc|c} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Any point of the form

$$\hat{\mathbf{x}} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \forall x_4 \in \mathbb{R}$$

is a least-squares solution of the problem.

Least squares

Theorem 5.2

The matrix $A^T A$ is invertible iff the columns of A are linearly independent. In this case, the equation system $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = A^+\mathbf{b}$$

where A^+ is the ***Moore-Penrose pseudoinverse***

$$A^+ = (A^T A)^{-1} A^T$$

Least squares and QR decomposition

Sometimes $A^T A$ is ill-conditioned, this means that small perturbations in $\mathbf{b}$ translate into large perturbations in $\hat{\mathbf{x}}$. The QR decomposition offers a numerically more stable way of finding the least-squares solution.

Theorem 5.3

Let there be $A \in \mathcal{M}_{m \times n}$ with linearly independent columns. Consider its QR decomposition ($A = QR$). Then, for each $\mathbf{b} \in \mathbb{R}^m$ there is a unique least-squares solution of $A\mathbf{x} = \mathbf{b}$ given by

$$\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$$

Proof

If we substitute $\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$ into $A\mathbf{x}$ we have

$$A\hat{\mathbf{x}} = AR^{-1} Q^T \mathbf{b} = QRR^{-1} Q^T \mathbf{b} = QQ^T \mathbf{b}.$$

But Q is an orthonormal basis of $\mathrm{Col}\{A\}$ (Theorem 4.2 and Corollary in Slide 49) and consequently $QQ^T \mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}\{A\}$, that is, $\hat{\mathbf{b}}$. So, $\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. Additionally, since the columns of A are linearly independent, by Theorem 5.2, this solution is unique.

Remind that numerically it is easier to solve $R\hat{\mathbf{x}} = Q^T\mathbf{b}$ than $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$

L

et $A = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}$. Its QR decomposition is

$$A = QR = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

$$Q^T\mathbf{b} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \hat{\mathbf{x}} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{pmatrix} 10 \\ -6 \\ 2 \end{pmatrix}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 5:

- 6.5.1
- 6.5.19
- 6.5.20
- 6.5.21
- 6.5.24

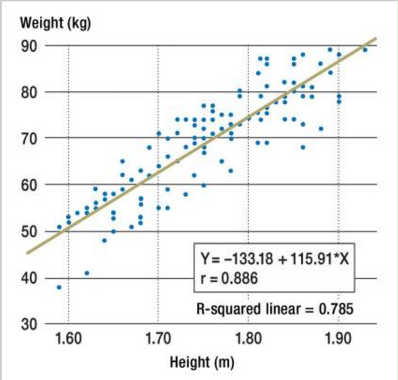
Outline

Least-squares linear regression

Example

In many scientific and engineering problems, it is needed to explain some observations **y** as a linear function of an independent variable **x**. For instance, we may try to explain the weight of a person as a linear function of its height

$$Weight = \beta_0 + \beta_1 Height$$



A. Schneider, G. Hommel, M. Blettner. *Linear Regression Analysis*. Dtsch Arztebl Int. **2010** November; 107(44): 776–782.

Least-squares linear regression

Example (continued)

For each observation we have an equation

Height (m.)	**Weight** (kg.)
1.70	57
1.53	43
1.90	94
...	...

$$57 = \beta_0 + 1.70\beta_1$$
$$43 = \beta_0 + 1.53\beta_1$$
$$94 = \beta_0 + 1.90\beta_1$$
$$\dots$$

$$\begin{pmatrix} 1 & 1.70 \\ 1 & 1.53 \\ 1 & 1.90 \\ \dots & \dots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 57 \\ 43 \\ 94 \\ \dots \end{pmatrix}$$

which is of the form

$$X\beta = \mathbf{y}$$

Least-squares linear regression

Least-squares regression

Each one of the observed **data points** (x_j, y_j) gives an equation. All together provide an equation system

$$X\beta = \mathbf{y}$$

that is an overdetermined, linear equation system of the form $A\mathbf{x} = \mathbf{b}$. The matrix X is called the **system matrix** and it is related to the **independent** (**predictor**) variables (the height in this case). The vector $\mathbf{y}$ is called the **observation vector** and collects the values of the **dependent** (**predicted**) variable (the weight in this case). The model

$$y = \beta_0 + \beta_1 x + \epsilon$$

is called the **linear regression of** y **on** x. β_0 and β_1 are called the **regression coefficients**. The difference between the predicted value and the observed value for a particular observation (ϵ) is called the **residual** of that observation.

Least-squares linear regression

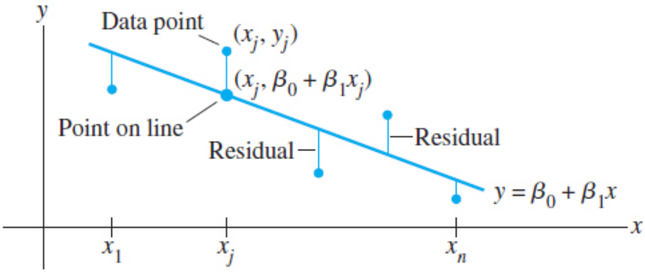


FIGURE 1 Fitting a line to experimental data.

The residual of the j-th observation is defined as

$$\epsilon_j = y_j - (\beta_0 + \beta_1 x_j)$$

Least-squares linear regression

The goal of least-squares regression is to minimize

$$\sum_{j=1}^{n} \epsilon_j^2 = \|\mathbf{y} - X\beta\|^2$$

Let's analyze this term

$$X\beta = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_2 x_2 \\ \dots \\ \beta_0 + \beta_n x_n \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_n \end{pmatrix}$$

Then

$$\|\mathbf{y} - X\beta\|^2 = \left\| \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \dots \\ y_n - \hat{y}_n \end{pmatrix} \right\|^2 = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} \epsilon_j^2$$

Least-squares linear regression

Example

Suppose we have observed the following values of height and weight (1.70,57), (1.53,43), (1.90,94). We construct the system matrix $X = \begin{pmatrix} 1 & 1.70 \\ 1 & 1.53 \\ 1 & 1.90 \end{pmatrix}$ and the observation vector $\mathbf{y} = \begin{pmatrix} 57 \\ 43 \\ 94 \end{pmatrix}$. Now we look the normal equations

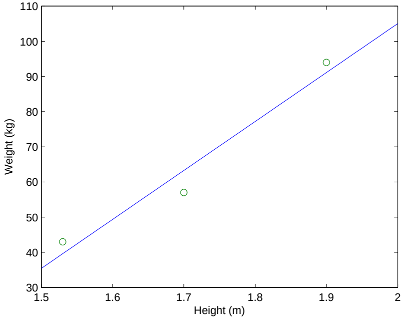
$$X\beta = \mathbf{y} \Rightarrow X^T X\beta = X^T\mathbf{y}$$

$$X^T X = \begin{pmatrix} 3.00 & 5.13 \\ 5.13 & 8.84 \end{pmatrix} \quad X^T\mathbf{y} = \begin{pmatrix} 194.00 \\ 341.29 \end{pmatrix} \quad \hat{\beta} = (X^T X)^{-1} X^T\mathbf{y} = \begin{pmatrix} -173.14 \\ 137.90 \end{pmatrix}$$

$$Weight = -173.39 + 139.21 Height$$

Least-squares linear regression

Example



MATLAB:

```
 X=[1 1.70; 1 1.53; 1 1.90];
y=[57; 43; 94];
beta=inv(X'*X)*X'*y
x=1.5:0.01:2.00;
yp=beta(1)+beta(2)*x;
plot(x,yp,X(:,1),y,'o')
xlabel('Height (m)')
ylabel('Weight (kg)')
```

Least-squares linear regression

The general linear model

The linear model is not restricted to straight lines. We can use it to fit any kind of curves:

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \beta_2 f_2(x) + \ldots$$

Fitting a parabola

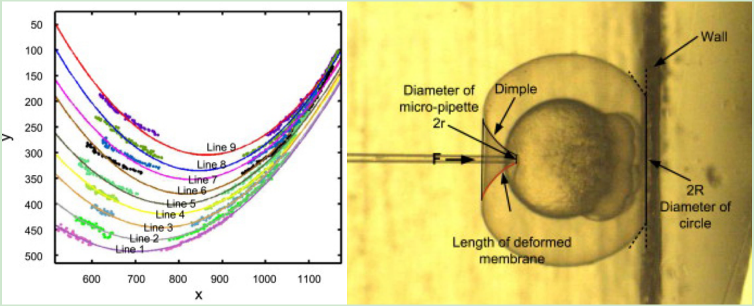
$$\begin{array}{c} f_0(x) = 1 \\ f_1(x) = x \\ f_2(x) = x^2 \end{array} \Rightarrow \quad \begin{array}{l} y_1 = f_0(x_1) + \beta_1 f_1(x_1) + \beta_2 f_2(x_1) \\ y_2 = f_0(x_2) + \beta_1 f_1(x_2) + \beta_2 f_2(x_2) \\ \ldots \\ y_n = f_0(x_n) + \beta_1 f_1(x_n) + \beta_2 f_2(x_n) \end{array}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \ldots \\ \epsilon_n \end{pmatrix} \quad \Rightarrow \mathbf{y} = X\beta + \epsilon$$

Least-squares linear regression

Fitting a parabola

In this example they model the deformation of the wall of the zebra fish embryo as a function of strain.



Z. Lua, P. C.Y. Chen, H. Luo, J. Nam, R. Ge, W. Lin. *Models of maximum stress and strain of zebrafish embryos under indentation*. J. Biomechanics 42 (5): 620–625 (**2009**)

Least-squares linear regression

Multivariate linear regression

The linear model is not restricted to one variable. By fitting several variables we may fit surfaces and hypersurfaces

$$y = \beta_0 f_0(x_1, x_2) + \beta_1 f_1(x_1, x_2) + \beta_2 f_2(x_1, x_2) + \ldots$$

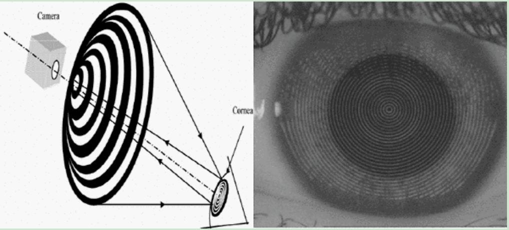
Fitting a parabolic surface

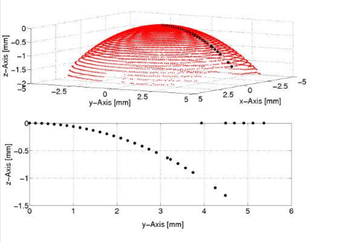
$$\begin{array}{l} f_0(x_1, x_2) = 1 \\ f_1(x_1, x_2) = x_1 \\ f_2(x_1, x_2) = x_2 \\ f_3(x_1, x_2) = x_1^2 \\ f_4(x_1, x_2) = x_2^2 \\ f_5(x_1, x_2) = x_1 x_2 \end{array} \quad \Rightarrow X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{pmatrix}$$

Least-squares linear regression

Fitting a parabolic surface

In this example they model the shape of cornea using videokeratoscopic images.





Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 6:

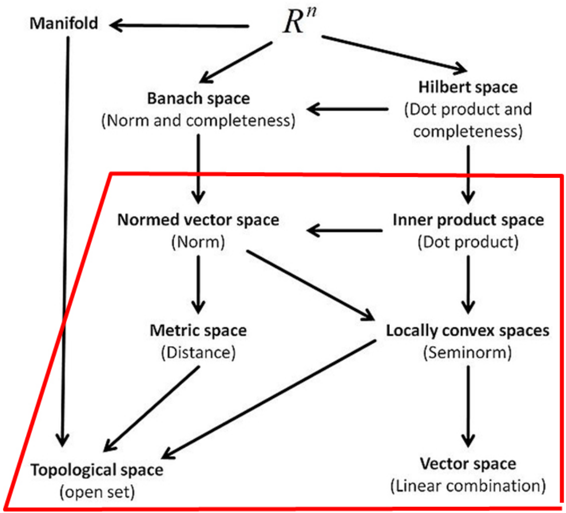
- 6.6.1
- 6.6.5
- 6.6.9
- 6.6.12 (computer)

Outline

7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

Inner product spaces



Inner product spaces

Definition 7.1 (Inner product)

An inner product in a vector space V is a function that assigns a real number to every pair of vectors $\mathbf{u}$ and $\mathbf{v}$, $\langle \mathbf{u}, \mathbf{v} \rangle$ and that satisfies the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ *and* $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ *iff* $\mathbf{u} = \mathbf{0}$.

Example

For instance in **Weighted Least Squares** (WLS) we may use an inner product in $\mathbb{R}^2$ defined as:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2$$

In this way we give less weight to distances in the first component with respect to distances in the second component.

Inner product spaces

Now we have to prove that this function is effectively an inner product:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

$$\begin{array}{rcll} \langle \mathbf{u}, \mathbf{v} \rangle & = & 4u_1v_1 + 5u_2v_2 & \text{[by definition]} \\ & = & 4v_1u_1 + 5v_2u_2 & \text{[commutativity of scalar multiplication]} \\ & = & \langle \mathbf{v}, \mathbf{u} \rangle & \text{[by definition]} \end{array}$$

2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

$$\begin{array}{rcll} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle & = & 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 & \text{[by definition]} \\ & = & 4u_1w_1 + 4v_1w_1 + 5u_2w_2 + 5v_2w_2 & \text{[distributivity of scalar]} \\ & & & \text{[multiplication/addition]} \\ & = & 4u_1w_1 + 5u_2w_2 + 4v_1w_1 + 5v_2w_2 & \text{[commutativity]} \\ & & & \text{[of scalar addition]} \\ & = & \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle & \text{[by definition]} \end{array}$$

Inner product spaces

3. $\langle c\mathbf{u}, \mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$

$$\begin{aligned}
\langle c\mathbf{u}, \mathbf{v}\rangle &= 4cu_1v_1 + 5cu_2v_2 && \text{[by definition]}\\
&= c4v_1u_1 + c5v_2u_2 && \text{[commutativity of scalar multiplication]}\\
&= c(4v_1u_1 + 5v_2u_2) && \text{[distributivity of scalar multiplication]}\\
&= c\langle \mathbf{u}, \mathbf{v}\rangle && \text{[by definition]}
\end{aligned}$$

4. $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ iff $\mathbf{u} = \mathbf{0}$.
   1. $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$

      $$\langle \mathbf{u}, \mathbf{u}\rangle = 4u_1^2 + 5u_2^2 \qquad \text{[by definition]}$$

      which is obviously larger than 0.
   2. $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ iff $\mathbf{u} = \mathbf{0}$.

      $$\langle \mathbf{u}, \mathbf{u}\rangle = 0 \Leftrightarrow 4u_1^2 + 5u_2^2 = 0 \Leftrightarrow u_1 = u_2 = 0$$

Inner product spaces

Example

Consider two vectors p and q the vector space of polynomials of degree n ($\mathbb{P}_n$). Let $t_0, t_1, ..., t_n$ be n distinct real numbers and K any scalar. The inner product between p and q is defined as

$$\langle p, q\rangle = K\left(p(t_0)q(t_0) + p(t_1)q(t_1) + ... + p(t_n)q(t_n)\right)$$

Axioms 1-3 are easy to check. Let's prove Axiom 4

4. $\langle p, p\rangle \geq 0$ and $\langle p, p\rangle = 0$ iff $p = 0$.
    1. $\langle p, p\rangle \geq 0$

       $$\langle p, p\rangle = K\left(p^2(t_0) + p^2(t_1) + ... + p^2(t_n)\right) \quad \text{[by definition]}$$

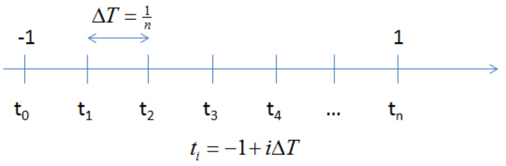
       which is obviously larger than 0.
    2. $\langle p, p\rangle = 0$ iff $p = 0$.

       $$\langle p, p\rangle = 0 \Leftrightarrow K\left(p^2(t_0) + p^2(t_1) + ... + p^2(t_n)\right) \Leftrightarrow$$
       $$p(t_0) = p(t_1) = ... = p(t_n) = 0$$

       But p is a polynomial of degree n so, at most, it can have n zeros. However, the previous condition requires the polynomial to vanish at $n+1$ points. This is impossible unless $p = 0$.

Example

Consider two vectors p and q the vector space of polynomials of degree n ($\mathbb{P}_n$). Assume that we regularly space the $n+1$ points in the interval $[-1, 1]$



and set $K = \Delta T$, then the inner product between the two polynomials becomes

$$\langle p, q \rangle = (p(t_0)q(t_0) + p(t_1)q(t_1) + ... + p(t_n)q(t_n))\,\Delta T = \sum_{i=0}^{n} p(t_i)q(t_i)\Delta T$$

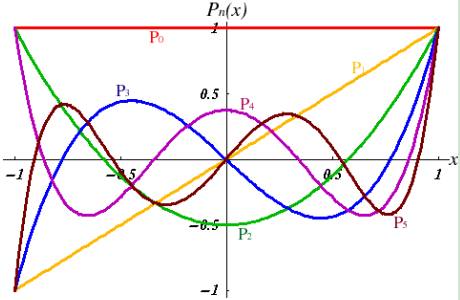
Making ΔT tend to 0, the inner product becomes

$$\langle p, q \rangle = \int_{-1}^{1} p(t)q(t)dt$$

Inner product spaces

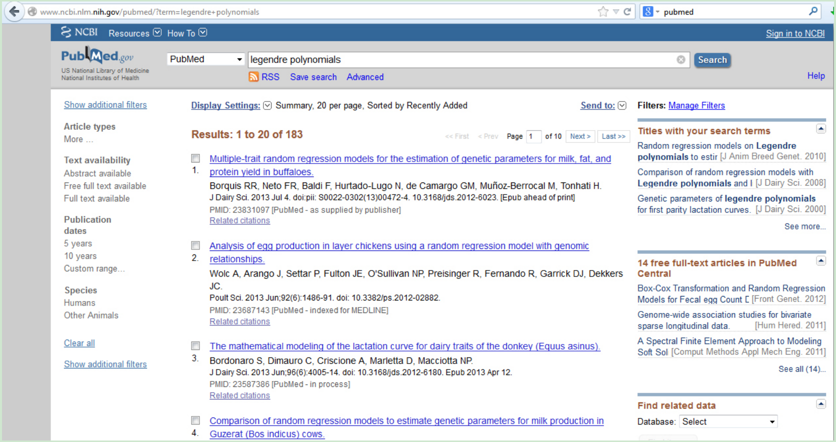
Legendre polynomials are orthogonal polynomials in the interval $[-1, 1]$

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$P_2(x) = \tfrac{1}{2}(3x^2 - 1)$$
$$P_3(x) = \tfrac{1}{2}(5x^3 - 3x)$$
$$P_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3)$$
$$P_5(x) = \tfrac{1}{8}(63x^5 - 70x^3 + 15x)$$
$$P_6(x) = \tfrac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$$
$$P_7(x) = \tfrac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$$



Legendre polynomials are very useful for regression of high-order polynomials as shown in next slide.

Inner product spaces



Length, distance and orthogonality

Length, distance and orthogonality

The **length** of a vector $\mathbf{u}$ in an inner product space is defined in the standard way

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

Similarly, the **distance between two vectors $\mathbf{u}$** and $\mathbf{v}$ is defined as

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Finally, two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be **orthogonal** iff

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0$$

Length, distance and orthogonality

Example

In the vector space of polynomials in the interval $[0,1]$, $\mathbb{P}[0,1]$, let's define the inner product

$$\langle p, q\rangle = \int_0^1 p(t)q(t)dt$$

What is the length of the vector $p(t) = 3t^2$?

Solution

$$\begin{aligned}\|p\| &= \sqrt{\langle p, p\rangle} = \sqrt{\int_0^1 p^2(t)dt} = \sqrt{\int_0^1 (3t^2)^2 dt} = \sqrt{\int_0^1 9t^4 dt} \\ &= \sqrt{9\frac{t^5}{5}\Big|_0^1} = \sqrt{9\left(\frac{1}{5}-0\right)} = \frac{3}{\sqrt{5}}\end{aligned}$$

Gram-Schmidt orthogonalization

Example

Gram-Schmidt is applied in the standard way. For instance, find an orthogonal basis of $\mathbb{P}_2[-1,1]$. A basis that spans this space is

$$\{1, t, t^2\}$$

Let's orthogonalize it

$$\begin{aligned}
p_0(t) &= 1 \\
p_1(t) &= t - \frac{\langle t, p_0(t)\rangle}{\|p_0\|^2} p_0(t) = t - \frac{\int_{-1}^{1} t dt}{\int_{-1}^{1} dt} 1 = t - \frac{0}{2} 1 = t \\
p_2(t) &= t^2 - \frac{\left\langle t^2, p_0(t)\right\rangle}{\|p_0\|^2} p_0(t) - \frac{\left\langle t^2, p_1(t)\right\rangle}{\|p_1\|^2} p_1(t) \\
&= t^2 - \frac{\int_{-1}^{1} t^2 dt}{\int_{-1}^{1} dt} - \frac{\int_{-1}^{1} t^2 t dt}{\int_{-1}^{1} t^2 dt} t = t^2 - \frac{\frac{2}{3}}{2} = t^2 - \frac{1}{3}
\end{aligned}$$

In Slide 97 we proposed the Legendre polynomial of degree 2 to be $P_2(t) = \frac{1}{2}(3t^2 - 1)$, we can easily show that $P_2(t) = \frac{3}{2} p_2(t)$. Consequently, if $p_2(t)$ is orthogonal to $p_0(t)$ and $p_1(t)$ so is $P_2(t)$.

Best approximation

Example

What is the best approximation in $\mathbb{P}_2[-1,1]$ of $p(t) = t^3$?

<u>*Solution*</u>

We know the answer is the orthogonal projection of $p(t)$ onto $\mathbb{P}_2[-1,1]$. An orthogonal basis of $\mathbb{P}_2[-1,1]$ is $\{1, t, t^2 - \frac{1}{3}\}$. Therefore, this projection can be calculated as

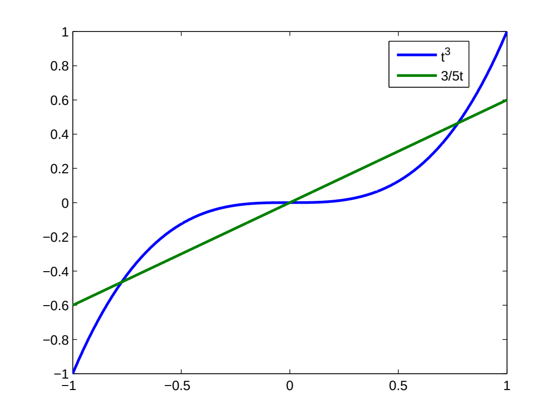
$$\hat{p}(t) = \mathrm{Proj}_{\mathbb{P}_2[-1,1]}\{p(t)\} = \frac{\langle p, p_0\rangle}{\|p_0\|^2} p_0(t) + \frac{\langle p, p_1\rangle}{\|p_1\|^2} p_1(t) + \frac{\langle p, p_2\rangle}{\|p_2\|^2} p_2(t)$$

Let's perform these calculations:

$$\begin{array}{l|l}
\langle p, p_0(t)\rangle = \int_{-1}^{1} t^3 dt = 0 & \|p_0\|^2 = \int_{-1}^{1} dt = 2 \\
\langle p, p_1(t)\rangle = \int_{-1}^{1} t^3 t dt = \frac{2}{5} & \|p_1\|^2 = \int_{-1}^{1} t^2 dt = \frac{2}{3} \\
\langle p, p_2(t)\rangle = \int_{-1}^{1} t^3 (t^2 - \frac{1}{3}) dt = 0 & \|p_2\|^2 = \int_{-1}^{1} (t^2 - \frac{1}{3})^2 dt = \frac{8}{45}
\end{array}$$

$$\hat{p}(t) = \frac{0}{2} + \frac{\frac{2}{5}}{\frac{2}{3}} t + \frac{0}{\frac{8}{45}}(t^2 - \frac{1}{3}) = \frac{3}{5} t$$

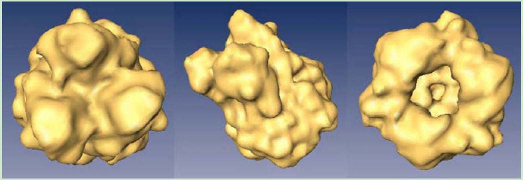
Best approximation



Best approximation

Example

In this example we exploited the best approximation property of orthogonal wavelets to speed-up and make more robust angular alignments of projections in 3D Electron Microscopy.



C.O.S.Sorzano, S. Jonic, C. El-Bez, J.M. Carazo, S. De Carlo, P. Thévenaz, M. Unser. *A multiresolution approach to orientation assignment in 3-D electron microscopy of single particles.* Journal of Structural Biology 146(3): 381-392 (**2004**, cover article)

Theorem 7.1 (Pythagorean theorem)

Given any vector $\mathbf{v}$ *in an inner product space* V *and a subspace of it* $W \subseteq V$ *we have*

$$\|\mathbf{v}\|^2 = \|\mathrm{Proj}_W\{\mathbf{v}\}\|^2 + \|\mathbf{v} - \mathrm{Proj}_W\{\mathbf{v}\}\|^2$$

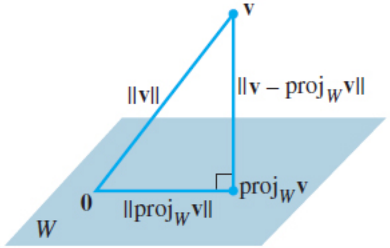


FIGURE 2

The hypotenuse is the longest side.

The Cauchy-Schwarz inequality

Theorem 7.2 (The Cauchy-Schwarz inequality)

For all $\mathbf{u}, \mathbf{v} \in V$ *it is verified*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Proof
If $\mathbf{u} = \mathbf{0}$*, then*

$$|\langle \mathbf{0}, \mathbf{v} \rangle| = 0 \quad \textit{and} \quad \|\mathbf{0}\|\|\mathbf{v}\| = 0\|\mathbf{v}\| = 0$$

So the inequality becomes an equality.
If $\mathbf{u} \neq \mathbf{0}$*, then consider* $W = \mathrm{Span}\{\mathbf{u}\}$ *and*

$$\|\mathrm{Proj}_W\{\mathbf{v}\}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|^2} \|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|}$$

But by the Pythagorean Theorem (Theorem 7.1) we have $\|\mathrm{Proj}_W\{\mathbf{v}\}\| \leq \|\mathbf{v}\|$.
Consequently,

$$\frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|} \leq \|\mathbf{v}\| \Rightarrow |\langle \mathbf{v}, \mathbf{u} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\| \textit{ (q.e.d.)}$$

The Triangle inequality

Theorem 7.3 (The Triangle inequality)

For all $\mathbf{u}, \mathbf{v} \in V$ *it is verified*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof

$$\begin{array}{rcll}
\|\mathbf{u} + \mathbf{v}\|^2 & = & \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle & \textit{[By definition]} \\
 & = & \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle & \textit{[Properties of inner product]} \\
 & \leq & \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| & \langle \mathbf{u}, \mathbf{v} \rangle \leq |\langle \mathbf{u}, \mathbf{v} \rangle| \\
 & \leq & \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| & \textit{Cauchy-Schwarz} \\
 & = & (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 & \\
 & \Rightarrow & & \\
\|\mathbf{u} + \mathbf{v}\| & \leq & \|\mathbf{u}\| + \|\mathbf{v}\| & \textit{[Taking square root]}
\end{array}$$

(q.e.d.)

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 7:

- 6.7.1
- 6.7.13
- 6.7.16
- 6.7.18

Outline

7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

Weighted Least Squares

Weighted Least Squares

Let us assume we have a table of collected data and we want to fit a least squares model. However, we want to give more importance to some observations because we are more confident about them or they are more important. We encode the importance as a weight value (the larger the weight, the more importance the observation has)

X	**Y**	**W**
x_1	y_1	w_1
x_2	y_2	w_2
x_3	y_3	w_3
...	...	...

Let us call $\hat{y}_j$ the prediction of the model for the j-th observation and ϵ_j the committed error

$$y_j = \hat{y}_j + \epsilon_j$$

Weighted Least Squares

The goal is now to minimize the weighted sum of square errors

$$\sum_{j=1}^{n}(w_j\epsilon_j)^2=\sum_{j=1}^{n}(w_j(y_j-\hat{y}_j))^2=\sum_{j=1}^{n}(w_jy_j-w_j\hat{y}_j)^2$$

Let us collect all observed values into a vector $\mathbf{y}$ and do analogously with the predictions $\hat{\mathbf{y}}$. Let us define the diagonal matrix

$$W=\begin{pmatrix} w_1 & 0 & 0 & \ldots & 0 \\ 0 & w_2 & 0 & \ldots & 0 \\ 0 & 0 & w_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & w_n \end{pmatrix}$$

Then, the previous objective function becomes

$$\sum_{j=1}^{n}(w_jy_j-w_j\hat{y}_j)^2=\|W\mathbf{y}-W\hat{\mathbf{y}}\|^2$$

Weighted Least Squares

Now, suppose that $\hat{\mathbf{y}}$ is calculated from the columns of a matrix A, that is, $\hat{\mathbf{y}} = A\mathbf{x}$. The objective function becomes

$$\sum_{j=1}^{n} (w_j y_j - w_j \hat{y}_j)^2 = \|W\mathbf{y} - WA\mathbf{x}\|^2$$

The minimum of this objective function is attained for $\hat{\mathbf{x}}$ that is the least-squares solution of the equation system

$$WA\mathbf{x} = W\mathbf{y}$$

The normal equations of the problem are

$$(WA)^T WA\mathbf{x} = (WA)^T W\mathbf{y}$$

Weighted Least Squares

Example

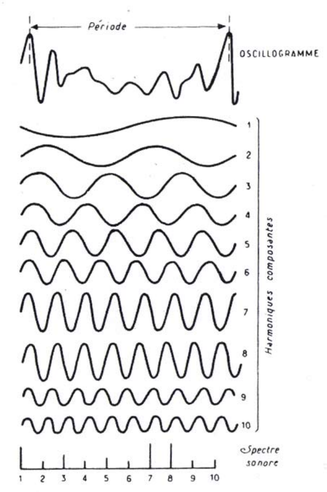
In this work they used Weighted Least Squares to calibrate a digital system to measure maximum respiratory pressures.

J.L. Ferreira, F.H. Vasconcelos, C.J. Tierra-Criollo. *A Case Study of Applying Weighted Least Squares to Calibrate a Digital Maximum Respiratory Pressures Measuring System*. Applied Biomedical Engineering, Chapter 18 (**2011**)

Fourier Series

Example

Fourier tools are, maybe, the most widespread tool to analyze signals and its frequency components. Fourier decomposition states that any signal can be obtained by summing sine waves of different amplitude, phase and frequency.



Fourier Series

Theorem 8.1

Consider the vector space of continuous functions in the interval $[0, 2\pi]$, $C[0, 2\pi]$. *The set*

$$S = \{1, \cos(t), \sin(t), \cos(2t), \sin(2t), ..., \cos(Nt), \sin(Nt)\}$$

is orthogonal with respect to the inner product defined as

$$\langle f(t), g(t)\rangle = \int_0^{2\pi} f(t)g(t)dt$$

Proof

$$\begin{array}{rcl} \langle\cos(nt), \cos(mt)\rangle & = & \int_0^{2\pi} \cos(nt)\cos(mt)dt \\ & = & \int_0^{2\pi} \frac{1}{2}(\cos((n+m)t) + \cos((n-m)t))dt \\ & = & \frac{1}{2}\left(\frac{\sin((n+m)t)}{n+m} + \frac{\sin((n-m)t)}{n-m}\right)\Big|_0^{2*\pi} \\ & = & 0 \end{array}$$

where we have used $\cos(A)\cos(B) = \frac{1}{2}(\cos(A+B) + \cos(A-B))$.

Fourier Series

Analogously we could prove that

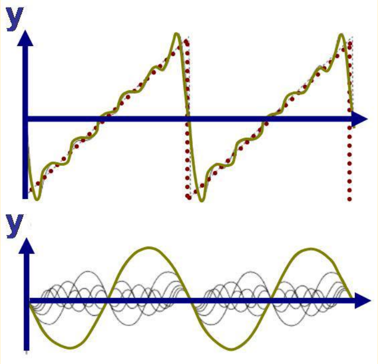
$$\begin{aligned}
\langle \cos(nt), \sin(mt) \rangle &= 0 \\
\langle \cos(nt), 1 \rangle &= 0 \\
\langle \sin(nt), 1 \rangle &= 0 \\
\| \cos(nt) \|^2 &= \pi \\
\| \sin(nt) \|^2 &= \pi \\
\| 1 \|^2 &= 2\pi
\end{aligned}$$

Fourier Series

Theorem 8.2 (Fourier series)

Given any function $f(t) \in C[0, 2\pi]$, $f(t)$ can be approximated as closely as desired by a sum of the form simply by orthogonally projecting it onto $W = \mathrm{Span}\{S\}$

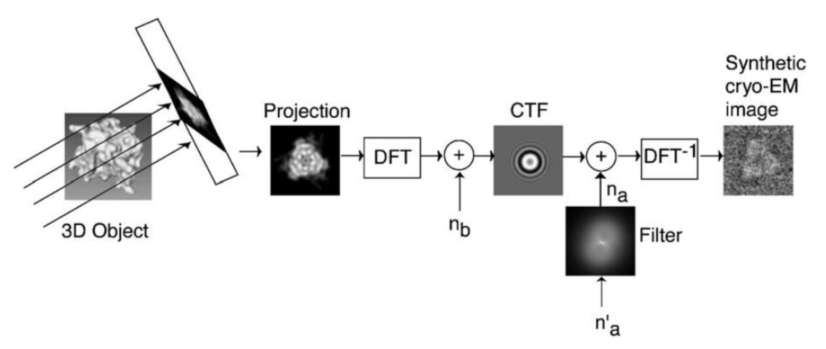
$$f(t) \approx \mathrm{Proj}_W\{f(t)\} = \frac{\langle f(t), 1\rangle}{\|1\|^2} + \sum_{n=1}^{N} \left(\frac{\langle f(t), \cos(nt)\rangle}{\|\cos(nt)\|^2} \cos(nt) + \frac{\langle f(t), \sin(nt)\rangle}{\|\sin(nt)\|^2} \sin(nt) \right)$$



Fourier Series

Example

In this work we used Fourier space to simulate and to align electron microscopy images



S. Jonic, C.O.S.Sorzano, P. Thévenaz, C. El-Bez, S. De Carlo, M. Unser. *Spline-Based image-to-volume registration for three-dimensional electron microscopy*. Ultramicroscopy, 103: 303-317 (**2005**)

Exercises

Exercises

From Lay (3rd ed.), Chapter 6, Section 8:

- 6.8.1
- 6.8.6
- 6.8.8
- 6.8.11

Outline

7. Orthogonality and least squares
   - Inner product, length and orthogonality (a)
   - Orthogonal sets, bases and matrices (a)
   - Orthogonal projections (b)
   - Gram-Schmidt orthogonalization (b)
   - Least squares (c)
   - Least-squares linear regression (c)
   - Inner product spaces (d)
   - Applications of inner product spaces (d)

Chapter 8. Symmetric matrices and quadratic forms

C.O.S. Sorzano

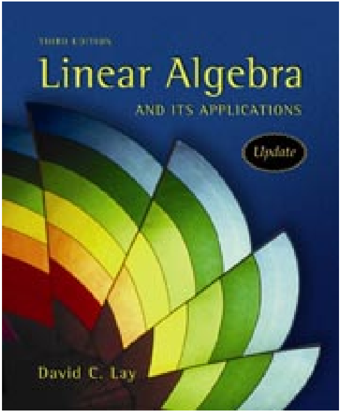
Biomedical Engineering

December 3, 2013

Outline

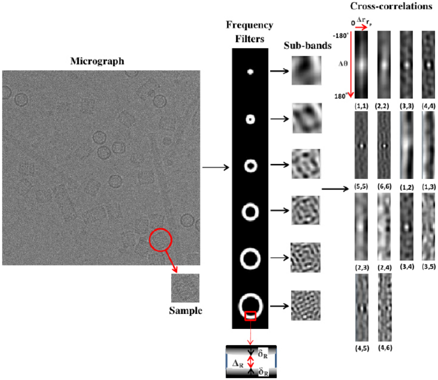
8 Symmetric matrices and quadratic forms
- Diagonalization of symmetric matrices (a)
- Quadratic forms (b)
- Constrained optimization (b)
- Singular Value Decomposition (SVD) (c)

References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 7.

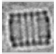
In this example of particle picking in Single Particles, one of the features we analyze is the autocorrelation function at different subbands. The autocorrelation is a symmetric matrix.



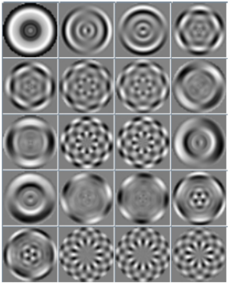
V. Abrishami, A. Zaldívar-Peraza, J.M. de la Rosa-Trevín, J. Vargas, J. Otón, R. Marabini, Y. Shkolnisky, J.M. Carazo, C.O.S. Sorzano. *A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs*. Bioinformatics (**2013**)

Applications

In one of the steps, we construct a basis that spans the set of rotations of the particle template. For doing so, perform a Principal Component Analysis that diagonalizes the covariance matrix (which is again a symmetric matrix).



(a)



V. Abrishami, A. Zaldívar-Peraza, J.M. de la Rosa-Trevín, J. Vargas, J. Otón, R. Marabini, Y. Shkolnisky, J.M. Carazo, C.O.S. Sorzano. *A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs*. Bioinformatics (**2013**)

Outline

8. Symmetric matrices and quadratic forms
   - Diagonalization of symmetric matrices (a)
   - Quadratic forms (b)
   - Constrained optimization (b)
   - Singular Value Decomposition (SVD) (c)

Diagonalization of symmetric matrices

Definition 1.1 (Symmetric matrix)

*$A \in \mathcal{M}_{n\times n}$ is a **symmetric matrix** iff $A = A^T$.*

Example

The following two matrices are symmetric

$$\begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix} \quad \begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{pmatrix}$$

Example

Let's diagonalize the matrix $A = \begin{pmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{pmatrix}$ The characteristic equation is

$$|A - \lambda I| = 0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

Diagonalization of symmetric matrices

The associated eigenvectors are

$$\begin{array}{ll} \lambda = 8 & \mathbf{v}_1 = (-1, 1, 0) \rightarrow \mathbf{u}_1 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0) \\ \lambda = 6 & \mathbf{v}_2 = (-1, -1, 2) \rightarrow \mathbf{u}_2 = (-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}) \\ \lambda = 3 & \mathbf{v}_3 = (1, 1, 1) \rightarrow \mathbf{u}_3 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}) \end{array}$$

The $\mathbf{v}$ vectors constitute an orthogonal basis of $\mathbb{R}^3$ and after normalizing them ($\mathbf{u}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$), we have an orthonormal basis Thus, we can factorize A as $A = PDP^{-1}$ with

$$P = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \quad D = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Exploiting the fact that P is orthonormal, then $P^{-1} = P^T$ and $A = PDP^T$.

Diagonalization of symmetric matrices

Theorem 1.1

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two eigenvectors from two different eigenvalues λ_1 and λ_2. Let's show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$

$$\begin{aligned}
\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 && \text{[By definition]} \\
&= (A\mathbf{v}_1)^T \mathbf{v}_2 && \text{[Definition of eigenvector]} \\
&= \mathbf{v}_1^T A^T \mathbf{v}_2 && \text{[Transpose of product]} \\
&= \mathbf{v}_1^T (A\mathbf{v}_2) && \text{[A is symmetric]} \\
&= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) && \text{[Definition of eigenvector]} \\
&= \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2) && \text{[By definition]}
\end{aligned}$$

Hence $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$ but $\lambda_1 - \lambda_2 \neq 0$ because the two eigenvalues are different. Consequently, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ (q.e.d.)

Diagonalization of symmetric matrices

Definition 1.2 (Orthogonal diagonalization)

*A is **orthogonally diagonalizable** iff $A = PDP^T$ being P an orthogonal (i.e., $P^{-1} = P^T$).*

Theorem 1.2

A is orthogonally diagonalizable iff A is symmetric.

Proof orthogonally diagonalizable $\Rightarrow$ symmetric

Let us assume that $A = PDP^T$, then

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PD^T P^T = PDP^T = A$$

Proof orthogonally diagonalizable $\Leftarrow$ symmetric

We omit this proof since it is more difficult.

Example

Let's orthogonally diagonalize $A = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$.

<u>*Solution*</u>
The characteristic equation is

$$|A - \lambda I| = 0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

Its associated eigenvectors are

$$\begin{array}{ll} \lambda = 7 & \mathbf{v}_1 = (1, 0, 1) \rightarrow \mathbf{u}_1 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \\ & \mathbf{v}_2 = (-\frac{1}{2}, 1, 2) \rightarrow \mathbf{u}_2 = (-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, 0) \\ \lambda = -2 & \mathbf{v}_3 = (-1, -\frac{1}{2}, 1) \rightarrow \mathbf{u}_3 = (-\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}) \end{array}$$

$\mathbf{u}_1$ and $\mathbf{u}_2$ are unitary and linearly independent, but they are not orthogonal. $\mathbf{u}_3$ is orthogonal to the other two vectors because it belongs to a different eigenspace (see Theorem 1.1).

Diagonalization of symmetric matrices

We can orthogonalize $\mathbf{u}_1$ and $\mathbf{u}_2$ following the Gram-Schmidt procedure:

$$\mathbf{w}_1 = \mathbf{v}_1 = (\tfrac{1}{\sqrt{2}}, 0, \tfrac{1}{\sqrt{2}})$$
$$\mathbf{w}_2' = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{w}_1 \rangle \mathbf{w}_1 = (-\tfrac{1}{\sqrt{5}}, \tfrac{2}{\sqrt{5}}, 0) - \left(-\tfrac{1}{\sqrt{10}}\right) (\tfrac{1}{\sqrt{2}}, 0, \tfrac{1}{\sqrt{2}}) = (-\tfrac{1}{2\sqrt{5}}, \tfrac{2}{\sqrt{5}}, \tfrac{1}{2\sqrt{5}})$$
$$\mathbf{w}_2 = \tfrac{\mathbf{w}_2'}{\|\mathbf{w}_2'\|} = (-\tfrac{1}{3\sqrt{2}}, \tfrac{2\sqrt{2}}{3}, \tfrac{1}{3\sqrt{2}})$$
$$\mathbf{w}_3 = \mathbf{v}_3 = (-\tfrac{2}{3}, -\tfrac{1}{3}, \tfrac{2}{3})$$

So $A = PDP^T$ with

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \end{pmatrix} \quad D = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Diagonalization of symmetric matrices

Definition 1.3 (Spectrum of a matrix)

*The set of eigenvalues of a matrix is called the **spectrum** of that matrix.*

Theorem 1.3 (Spectral theorem for symmetric matrices)

An $n \times n$ symmetric matrix has the following properties:

1. *A has n real eigenvalues (including multiplicities).*
2. *The dimension of each eigenspace is the multiplicity of the corresponding eigenvalue as root of the characteristic equation.*
3. *Eigenspaces corresponding to distinct eigenvalues are mutually orthogonal.*
4. *A is orthogonally diagonalizable.*

Diagonalization of symmetric matrices

Definition 1.4 (Spectral decomposition of symmetric matrices)

Let $A = PDP^T$ *with* $P = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{pmatrix}$. *Then*

$$\begin{aligned}
A &= \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \dots \\ \mathbf{u}_n^T \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 \mathbf{u}_1 & \lambda_2 \mathbf{u}_2 & \dots & \lambda_n \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \dots \\ \mathbf{u}_n^T \end{pmatrix} \\
&= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \dots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T
\end{aligned}$$

The latest equation is the ***spectral decomposition*** *of A. Each one of the terms* $\lambda_i \mathbf{u}_i \mathbf{u}_i^T$ *is an* $n \times n$ *matrix of rank 1 (since all the columns are multiples of* $\mathbf{u}_i$. *Additionally,* $\mathbf{u}_i \mathbf{u}_i^T \mathbf{x}$ *is the orthogonal projection of any vector onto the subspace generated by* $\mathbf{u}_i$.

Diagonalization of symmetric matrices

Example

Write the spectral decomposition of

$$A = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

Solution

Consider $\mathbf{u}_1 = (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$ be the first column of P and $\mathbf{u}_2 = (-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$. Then

$$\mathbf{u}_1\mathbf{u}_1^T = \begin{pmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} \quad \mathbf{u}_2\mathbf{u}_2^T = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{pmatrix}$$

The spectral decomposition is therefore

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T = 8\begin{pmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{pmatrix}$$

Exercises

Exercises

From Lay (3rd ed.), Chapter 7, Section 1:

- 7.1.6
- 7.1.7
- 7.1.13
- 7.1.23
- 7.1.27
- 7.1.29
- 7.1.35

Outline

8. Symmetric matrices and quadratic forms
   - Diagonalization of symmetric matrices (a)
   - Quadratic forms (b)
   - Constrained optimization (b)
   - Singular Value Decomposition (SVD) (c)

Quadratic forms

Introduction

Most expressions appearing so far are linear: $A\mathbf{x}$, $\langle \mathbf{w}, \mathbf{x} \rangle$, $\mathbf{x}^T$, that is, if we construct an operator $T(\mathbf{x})$ with them (e.g., $T(\mathbf{x}) = A\mathbf{x}$, $T(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$, $T(\mathbf{x}) = \mathbf{x}^T$), it meets

$$T(a\mathbf{x}_1 + b\mathbf{x}_2) = aT(\mathbf{x}_1) + bT(\mathbf{x}_2)$$

However, there are nonlinear expressions like $\mathbf{x}^T\mathbf{x}$. Particularly, this one is said to be quadratic and they normally appear in applications of linear algebra to engineering (like optimization) and signal processing (like signal power). They also arise in physics (as potential and kinetic energy), differential geometry (as the normal curvature of surfaces) and statistics (as confidence ellipsoids).

Quadratic forms

Definition 2.1 (Quadratic forms)

A quadratic form in $\mathbb{R}^n$ *is a function* $Q(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ *that can be computed as*

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

being $A \in \mathcal{M}_{n\times n}$ *a symmetric matrix.*

Example

1. $Q(\mathbf{x}) = \mathbf{x}^T I \mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2$
2. $Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \mathbf{x} = 4x_1^2 + 3x_2^2$
3. $Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 3 & -2 \\ -2 & 7 \end{pmatrix} \mathbf{x} = 3x_1^2 + 7x_2^2 - 4x_1x_2$
4. $Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 5 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 3 & 4 \\ 0 & 4 & 2 \end{pmatrix} \mathbf{x} = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$

Change of variables in quadratic forms

Change of variables

A change of variables is an equation of the form $\mathbf{x} = P\mathbf{y}$ or equivalently $P^{-1}\mathbf{x} = \mathbf{y}$, where P is an invertible matrix. Exploiting the fact that, in a quadratic form, A is symmetric, then we have $A = PDP^T$. We perform the change of variables

$$\mathbf{x} = P\mathbf{y}$$

to obtain

$$Q(\mathbf{x}) = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T AP\mathbf{y} = Q(\mathbf{y})$$

But we know

$$A = PDP^T \Rightarrow D = P^T AP$$

Consequently

$$Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y}$$

That is, there is a basis, in which the matrix of the quadratic form is diagonal.

Change of variables in quadratic forms

Example

Consider $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ with

$$A = \begin{pmatrix} 1 & -4 \\ -4 & -5 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -7 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

That is

$$Q(\mathbf{x}) = x_1^2 - 5x_2^2 - 8x_1x_2$$

If we make the change of variable

$$\mathbf{y} = P^T\mathbf{x} = \begin{pmatrix} \frac{2}{\sqrt{5}}x_1 - \frac{1}{\sqrt{5}}x_2 \\ \frac{1}{\sqrt{5}}x_1 + \frac{2}{\sqrt{5}}x_2 \end{pmatrix}$$

then

$$Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y} = 3y_1^2 - 7y_2^2$$

Change of variables in quadratic forms

Let's check that effectively both ways of calculating the quadratic form are equivalent. For doing so, we'll calculate the value of $Q(\mathbf{x})$ for $\mathbf{x} = (2, -2)$:

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = 2^2 - 5 \cdot (-2)^2 - 8 \cdot 2 \cdot (-2) = 4 - 20 + 32 = 16$$

If we make the change of variable

$$\mathbf{y} = \begin{pmatrix} \frac{2}{\sqrt{5}}2 - \frac{1}{\sqrt{5}}(-2) \\ \frac{1}{\sqrt{5}}2 + \frac{2}{\sqrt{5}}(-2) \end{pmatrix} = \begin{pmatrix} \frac{6}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}$$

then

$$Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y} = 3\left(\frac{6}{\sqrt{5}}\right)^2 - 7\left(-\frac{2}{\sqrt{5}}\right)^2 = 3\frac{36}{5} - 7\frac{4}{5} = \frac{80}{5} = 16$$

Change of variables in quadratic forms

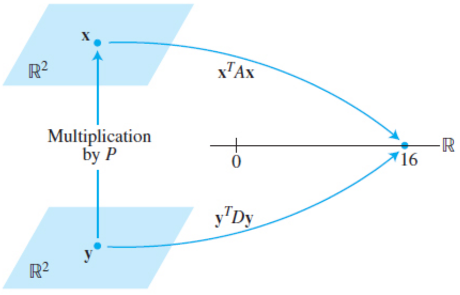


FIGURE 1 Change of variable in $\mathbf{x}^T A\mathbf{x}$.

Theorem 2.1 (Principal axes theorem)

Let $A \in \mathcal{M}_{n\times n}$ be a symmetric matrix. Then, there exists a change of variable $\mathbf{x} = P\mathbf{y}$ such that the quadratic form $\mathbf{x}^T A\mathbf{x}$ becomes $\mathbf{y}^T D\mathbf{y}$ with D an $n \times n$ diagonal matrix. The columns of P are the principal axes.

A geometric view of the principal axes

Consider the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ with $\mathbf{x} \in \mathbb{R}^2$ and the isocurve $Q(\mathbf{x}) = c$. The isocurve is either an ellipse, a circle, a hyperbola, two intersecting lines, a point, or contains no points at all. If A is diagonal, then

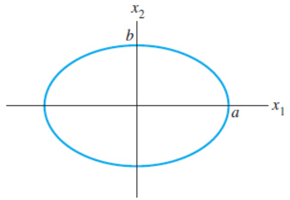
$$Q(\mathbf{x}) = a_{11}x_1^2 + a_{22}x_2^2 = c$$

The equation of an ellipse is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

with $a, b > 0$. Therefore

$$a = \sqrt{\frac{c}{a_{11}}} \quad b = \sqrt{\frac{c}{a_{22}}}$$



$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, \; a > b > 0$

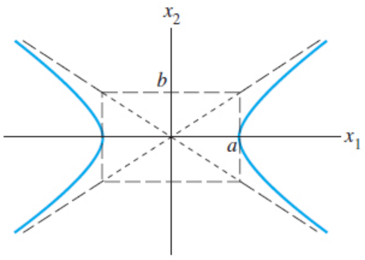
ellipse

Principal axes

The equation of a hyperbola is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$$

with $a, b > 0$

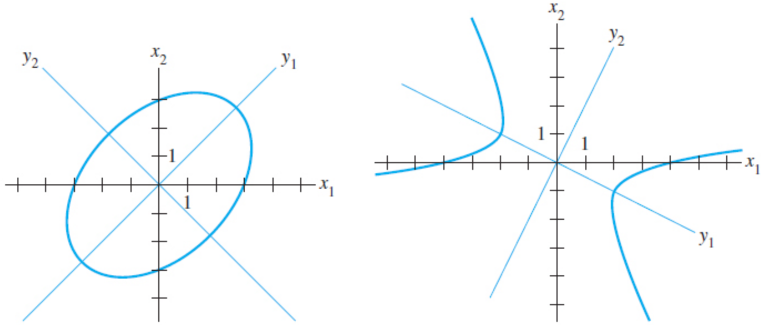


$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1, \; a > b > 0$$

hyperbola

Principal axes

If A is not diagonal, then the ellipse or the hyperbola are rotated



(a) $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$ (b) $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

Principal axes

Example

Let's analyze the rotated ellipse

$$5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$$

The corresponding matrix is

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

So,

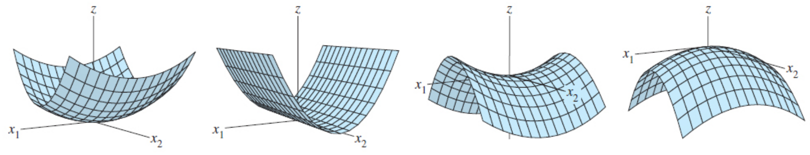
$$a = \sqrt{\frac{c}{a_{11}}} = \sqrt{\frac{48}{3}} = 3 \quad b = \sqrt{\frac{c}{a_{22}}} = \sqrt{\frac{48}{7}} \approx 2.65$$

The change of variable $\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \mathbf{y}$ diagonalizes the quadratic form (see the new axes in the previous slide).

Classification of quadratic forms

Example

Look at the following surfaces defined as $z = Q(\mathbf{x})$



(a) $z = 3x_1^2 + 7x_2^2$ (b) $z = 3x_1^2$ (c) $z = 3x_1^2 - 7x_2^2$ (d) $z = -3x_1^2 - 7x_2^2$

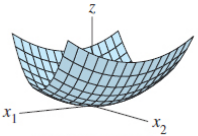
The curves seen in $\mathbb{R}^2$ are the cut of these surfaces with the plane $z = c$. It is obvious that some of the surfaces are always above $z = 0$ (a and b), others are always below $z = 0$ (d), and still other are sometimes below and sometimes above $z = 0$ (c).

Classification of quadratic forms

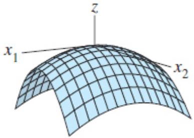
Definition 2.2 (Classification of quadratic forms)

We say $Q(\mathbf{x})$ is

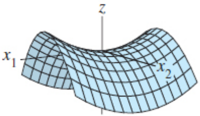
- ***positive definite*** *if*
  $Q(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$
- ***negative definite*** *if*
  $Q(\mathbf{x}) < 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$
- ***indefinite*** *if $Q(\mathbf{x})$ assumes both positive and negative values*
- ***positive semidefinite*** *if*
  $Q(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$
- ***negative semidefinite*** *if*
  $Q(\mathbf{x}) \leq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$



Positive definite



Negative definite



Indefinite

Classification of quadratic forms

Theorem 2.2 (Classification of quadratic forms and quadratic forms)

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ *with* $A \in \mathcal{M}_{n\times n}$ *and symmetric. Let* λ_i *be the eigenvalues of* A. $Q(\mathbf{x})$ *is*

- ***positive definite*** *iff* $\lambda_i > 0 \quad \forall i$
- ***negative definite*** *iff* $\lambda_i < 0 \quad \forall i$
- ***indefinite*** *iff there are positive and negative eigenvalues*
- ***positive semidefinite*** *iff* $\lambda_i \geq 0 \quad \forall i$
- ***negative semidefinite*** *iff* $\lambda_i \leq 0 \quad \forall i$

Proof
By the Theorem of Principal Axes (Theorem 2.1), there is a change of variable such that

$$Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_n y_n^2$$

where λ_i *is the i-th eigenvalue. The values of* Q *depend on* λ_i *in the way that the theorem states (e.g.,* $\forall \mathbf{y} \neq \mathbf{0} \quad Q(\mathbf{y}) > 0$ *iff* $\lambda_i > 0 \quad \forall i$*, etc.)*

Classification of quadratic forms

Examples

- $Q(\mathbf{x}) = 3x_1^2 + 7x_2^2$ is positive definite because its eigenvalues are 3 and 7 (both larger than 0).
- $Q(\mathbf{x}) = 3x_1^2$ is positive semidefinite because its eigenvalues are 3 and 0 (both larger or equal than 0).
- $Q(\mathbf{x}) = 3x_1^2 - 7x_2^2$ is indefinite because its eigenvalues are 3 and -7 (one positive and another negative).
- $Q(\mathbf{x}) = -3x_1^2 - 7x_2^2$ is negative definite because its eigenvalues are -3 and -7 (both smaller than 0).

Definition 2.3 (Classification of symmetric matrices)

*A symmetric **matrix is positive definite** if its corresponding quadratic form is positive definite. Analogously for the rest of the classification.*

Classification of quadratic forms

Cholesky factorization

Cholesky factorization factorizes a symmetric matrix A as

$$A = R^T R$$

being R an upper triangular matrix. A is positive definite if all entries in the diagonal of R are positive.

Exercises

Exercises

From Lay (3rd ed.), Chapter 7, Section 2:

- 7.2.1
- 7.2.3
- 7.2.5
- 7.2.7
- 7.2.19
- 7.2.23
- 7.2.24
- 7.2.26
- 7.2.27

Outline

8. Symmetric matrices and quadratic forms
   - Diagonalization of symmetric matrices (a)
   - Quadratic forms (b)
   - Constrained optimization (b)
   - Singular Value Decomposition (SVD) (c)

Constrained optimization

Introduction

Many problems in engineering or physics are of the form

$$\begin{array}{ll} \min & Q(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|^2 = 1 \end{array} \quad \text{or} \quad \begin{array}{ll} \max & Q(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|^2 = 1 \end{array}$$

Example

Calculate the minimum and maximum of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to $\|\mathbf{x}\|^2 = 1$. *Solution*
By taking the minimum and maximum coefficient in $Q(\mathbf{x})$ we have

$$\begin{array}{rcl} 3x_1^2 + 3x_2^2 + 3x_3^2 \leq & Q(\mathbf{x}) & \leq 9x_1^2 + 9x_2^2 + 9x_3^2 \\ 3(x_1^2 + x_2^2 + x_3^2) \leq & Q(\mathbf{x}) & \leq 9(x_1^2 + x_2^2 + x_3^2) \\ 3 \leq & Q(\mathbf{x}) & \leq 9 \end{array}$$

The minimum value $Q(\mathbf{x}) = 3$ is attained for $\mathbf{x} = (0, 0, 1)$, while the maximum value $Q(\mathbf{x}) = 9$ is attained for $\mathbf{x} = (1, 0, 0)$. In fact the minimum and maximum values that the constrained quadratic form can take are λ_{min} and λ_{max}.

Constrained optimization

Example

Calculate the minimum and maximum of $Q(\mathbf{x}) = 3x_1^2 + 7x_2^2$ subject to $\|\mathbf{x}\|^2 = 1$.

Solution

$\|\mathbf{x}\|^2 = 1$ is a cylinder in $\mathbb{R}^3$ while $z = Q(\mathbf{x})$ is a parabolic surface. The minimum and maximum of the constrained problem are attained among those points belonging to the curve that is the intersection of both surfaces.

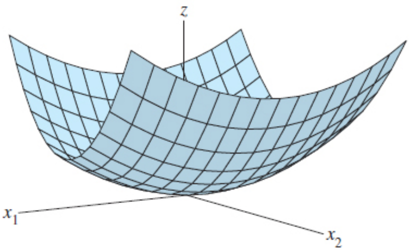


FIGURE 1 $z = 3x_1^2 + 7x_2^2$.

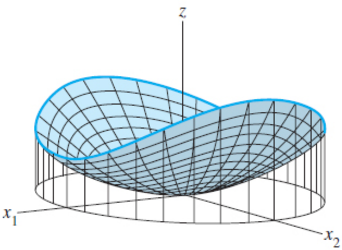


FIGURE 2 The intersection of $z = 3x_1^2 + 7x_2^2$ and the cylinder $x_1^2 + x_2^2 = 1$.

Constrained optimization

Theorem 3.1

Let A be a symmetric matrix and let

$$m = \min \left\{ \mathbf{x}^T A \mathbf{x} \,\middle|\, \|\mathbf{x}\|^2 = 1 \right\}$$
$$M = \max \left\{ \mathbf{x}^T A \mathbf{x} \,\middle|\, \|\mathbf{x}\|^2 = 1 \right\}$$

Then, $M = \lambda_{max}$ and $m = \lambda_{min}$. M is attained for $\mathbf{x} = \mathbf{u}_{max}$ (the eigenvector associated to λ_{max}) and m is attained for $\mathbf{x} = \mathbf{u}_{min}$ (the eigenvector associated to λ_{min}).

Proof

Let's orthogonally diagonalize A as $A = PDP^T$ and we make the change variables $\mathbf{y} = P^T\mathbf{x}$. We already know that

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y}$$

Additionally $\|\mathbf{y}\|^2 = \|\mathbf{x}\|^2$ because

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = (P^T\mathbf{x})^T(P^T\mathbf{x}) = \mathbf{x}^T PP^T \mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$$

In particular $\|\mathbf{y}\| = 1 \Leftrightarrow \|\mathbf{x}\| = 1$.

Constrained optimization

Then,

$$m = \min \left\{ \mathbf{y}^T D \mathbf{y} \,\middle|\, \|\mathbf{y}\|^2 = 1 \right\}$$
$$M = \max \left\{ \mathbf{y}^T D \mathbf{y} \,\middle|\, \|\mathbf{y}\|^2 = 1 \right\}$$

Since D is diagonal we have

$$\mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2$$

Let's look for the maximum of these values subject to $\|\mathbf{y}\| = 1$. Consider the maximum eigenvalue, λ_{max}, then

$$\begin{aligned} \mathbf{y}^T D \mathbf{y} &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2 \\ &\leq \lambda_{max} y_1^2 + \lambda_{max} y_2^2 + \ldots + \lambda_{max} y_n^2 \\ &= \lambda_{max} (y_1^2 + y_2^2 + \ldots + y_n^2) \\ &= \lambda_{max} \|\mathbf{y}\| = \lambda_{max} \end{aligned}$$

Constrained optimization

In fact the value λ_{max} is attained for $\mathbf{y}_{max} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}$, where the 1 is at the location corresponding to λ_{max}. The corresponding $\mathbf{x}$ is

$$\mathbf{x} = P\mathbf{y} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_{max-1} & \mathbf{u}_{max} & \mathbf{u}_{max+1} & \dots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} = \mathbf{u}_{max}$$

We could reason analogously for the minimum.

Constrained optimization

Example

Let $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}$. Solve the following optimization problem

$$\begin{array}{cc} \max & Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|^2 = 1 \end{array}$$

Solution

The characteristic equation is

$$|A - \lambda I| = 0 = -(\lambda - 6)(\lambda - 3)(\lambda - 1)$$

The maximum eigenvalue is $\lambda = 6$ and its corresponding eigenvector is $\mathbf{u} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$. Therefore, the maximum of $Q(\mathbf{x})$ is 6 that is attained for $\mathbf{x} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$.

Constrained optimization

Theorem 3.2

Let A, λ_{max} and $\mathbf{u}_{max}$ be defined as in the previous theorem. Then the solution of

$$\begin{array}{cc} \max & Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} \\ \textit{subject to} & \|\mathbf{x}\|^2 = 1 \\ & \mathbf{x} \cdot \mathbf{u}_{max} = 0 \end{array}$$

is given by the second largest eigenvalue λ_{max-1} that is attained for its associated eigenvector ($\mathbf{u}_{max-1}$).

Exercises

Exercises

From Lay (3rd ed.), Chapter 7, Section 3:

- 7.3.1
- 7.3.3
- 7.3.13

Outline

8 Symmetric matrices and quadratic forms

- Diagonalization of symmetric matrices (a)
- Quadratic forms (b)
- Constrained optimization (b)
- Singular Value Decomposition (SVD) (c)

Singular Value Decomposition (SVD)

Introduction

Unfortunately, not all matrices can be diagonalized and factorized as

$$A = PDP^{-1}$$

However, all of them (even rectangular matrices) can be factorized as

$$A = QDP^{-1}$$

This is called the **Singular Value Decomposition**. It imitates the property of stretching/shrinking of eigenvalues and eigenvectors. For instance, assume $\mathbf{u}$ is an eigenvector, then

$$A\mathbf{u} = \lambda\mathbf{u} \Rightarrow \|A\mathbf{u}\| = |\lambda|\|\mathbf{u}\|$$

If $|\lambda| > 1$, then the transformed vector $A\mathbf{u}$ is stretched with respect to $\mathbf{u}$. On the contrary, if $|\lambda| < 1$, then the transformed vector $A\mathbf{u}$ is shrinked with respect to $\mathbf{u}$.

Singular Value Decomposition (SVD)

Example

Consider $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$ and the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. It transforms the unit sphere in $\mathbb{R}^3$ onto an ellipse of $\mathbb{R}^2$

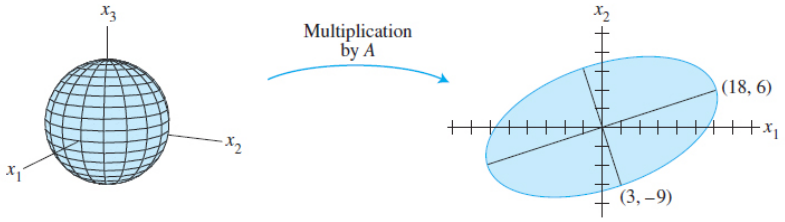


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Look for the direction that maximizes $\|A\mathbf{x}\|$ subject to $\|\mathbf{x}\| = 1$.

Singular Value Decomposition (SVD)

Solution

We may maximize $\|A\mathbf{x}\|^2$ because $\|A\mathbf{x}\|$ is maximum iff $\|A\mathbf{x}\|^2$ is maximum.

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x}$$

which is a quadratic form since $A^T A$ is symmetric:

$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$$

By Theorem 3.1, the maximum eigenvalue is $\max \|A\mathbf{x}\|^2 = \lambda_{max} = 360$ and its associated eigenvector $\mathbf{u}_{max} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$. Consequently $\max \|A\mathbf{x}\| = \sqrt{360} = 6\sqrt{10}$ that is attained for

$$A\mathbf{u}_{max} = \begin{pmatrix} 18 \\ 6 \end{pmatrix}$$

Definition 4.1

Singular Values of a matrix Let $A \in \mathcal{M}_{m\times n}$. A^TA can always be orthogonally diagonalized. Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ a base of $\mathbb{R}^n$ formed by the eigenvectors of A^TA and let $\lambda_1, \lambda_2, ..., \lambda_n$ be its corresponding eigenvalues. Then

$$\|A\mathbf{v}_i\|^2 = (A\mathbf{v}_i)^T(A\mathbf{v}_i) = \mathbf{v}_i^T A^T A\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2$$

If we take the square root

$$\|A\mathbf{v}_i\| = \sqrt{\lambda_i}\|\mathbf{v}_i\|$$

*That is, $\sqrt{\lambda_i}$ reflects the amount by which $\mathbf{v}_i$ is stretched or shrinked. $\sqrt{\lambda_i}$ is called a **singular value** and it is denoted as σ_i.*

Singular Value Decomposition (SVD)

Example (continued)

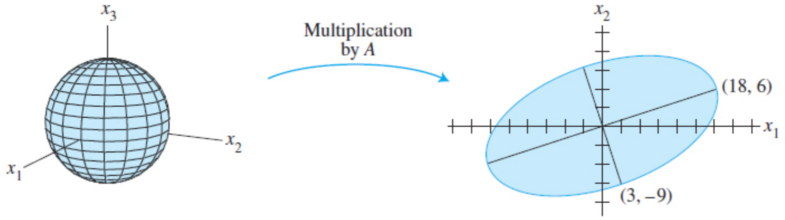


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

In the example of Slide 45, the singular values are the lengths of the ellipse in $\mathbb{R}^2$ and they are $6\sqrt{10}$, $3\sqrt{10}$ and 0. From the singular values we learn that the unit sphere in $\mathbb{R}^3$ (there are 3 singular values) is collapsed in 2D (one of the singular values is 0) onto an ellipse (the remaining two singular values are different from each other).

Singular Value Decomposition (SVD)

Theorem 4.1

Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ *a basis of* $\mathbb{R}^n$ *formed by the eigenvectors of* $A^T A$ *sorted in descending order and let* $\lambda_1, \lambda_2, ..., \lambda_n$ *be its corresponding eigenvalues. Let us assume that* A *has* r *non-null singular values. Then*

$$S = \{A\mathbf{v}_1, A\mathbf{v}_2, ..., A\mathbf{v}_r\}$$

is a basis of $\mathrm{Col}\{A\}$ *and*

$$\mathrm{Rank}\{A\} = r$$

Proof
By Theorem 1.1, any two eigenvectors are orthogonal to each other if they correspond to different eigenvalues, that is, $\mathbf{v}_i \cdot \mathbf{v}_j = 0$*. Then,*

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_i^T A^T A\mathbf{v}_j = \mathbf{v}_i^T(\lambda_j \mathbf{v}_j) = \lambda_j(\mathbf{v}_i^T \mathbf{v}_j) = \lambda_j(\mathbf{v}_i \cdot \mathbf{v}_j) = 0$$

That is $A\mathbf{v}_i$ *and* $A\mathbf{v}_j$ *are also orthogonal.*

Singular Value Decomposition (SVD)

Additionally, if the eigenvectors $\mathbf{v}_i$ are unitary, then (see Definition 4.1)

$$\sigma_i = \|A\mathbf{v}_i\|$$

Since there are r non-null singular values, $A\mathbf{v}_i \neq \mathbf{0}$ only for $i = 1, 2, ..., r$. So the set S is a set of non-null, orthogonal vectors. To show it is a basis of $\mathrm{Col}\{A\}$ we still need to show that any vector in $\mathrm{Col}\{A\}$ can be expressed as a linear combination of the vectors in S. We know that the eigenvalues of $A^T A$ is a basis of $\mathbb{R}^n$. Then for any vector $\mathbf{x} \in \mathbb{R}^n$ there exist coefficients $c_1, c_2, ..., c_n$ not all of them zero such that

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n$$

If we transform this vector

$$\begin{array}{rcll} A\mathbf{x} & = & A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n) & \text{[Linear transformation]} \\ & = & c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + ... + c_n A\mathbf{v}_n & \text{[non-null singular values]} \\ & = & c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + ... + c_r A\mathbf{v}_r & \end{array}$$

Singular Value Decomposition (SVD)

That is any transformed vector $A\mathbf{x}$ can be expressed as a linear combination of the elements in S. Consequently, S is a basis of $\mathrm{Col}\{A\}$.
Finally, $\mathrm{Rank}\{A\}$ is nothing more than the dimension of $\mathrm{Col}\{A\}$. Since A is a basis of $\mathrm{Col}\{A\}$ and it has r vectors, then $\mathrm{Rank}\{A\} = r$.

Singular Value Decomposition (SVD)

Theorem 4.2 (The Singular Value Decomposition)

Let $A \in \mathcal{M}_{m\times n}$ be a matrix with rank r. Then, there exists a matrix $\Sigma \in \mathcal{M}_{m\times n}$ whose diagonal entries are the first r singular values of A sorted in descending order ($\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$) and there exist orthogonal matrices $U \in \mathcal{M}_{m\times m}$ and $V \in \mathcal{M}_{n\times n}$ such that

$$A = U\Sigma V^T$$

Σ is unique but U and V are not. The columns of U are called the left singular vectors, and the columns of V are the right singular vectors.

Example

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} & v_{31} & v_{41} \\ v_{12} & v_{22} & v_{32} & v_{42} \\ v_{13} & v_{23} & v_{33} & v_{43} \\ v_{14} & v_{24} & v_{34} & v_{44} \end{pmatrix}$$

Singular Value Decomposition (SVD)

Proof

Let λ_i and $\mathbf{v}_i$ $(i = 1, 2, ..., n)$ be the eigenvalues and eigenvectors of $A^T A$. By Theorem 4.1 we know that $S = \{A\mathbf{v}_1, A\mathbf{v}_2, ..., A\mathbf{v}_r\}$ is an orthogonal basis of $\mathrm{Col}\{A\}$. Let's normalize these vectors

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i} \quad i = 1, 2, ..., r$$

and we extend the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r\}$ to be an orthogonal basis of $\mathbb{R}^m$. Let us construct the matrices

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m \end{pmatrix}$$
$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{pmatrix}$$

By construction U and V are orthogonal, and

$$\begin{array}{rcl} AV & = & \begin{pmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \ldots & A\mathbf{v}_r & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix} \\ & = & \begin{pmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \ldots & \sigma_r\mathbf{u}_r & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix} \end{array}$$

Singular Value Decomposition (SVD)

Proof (continued)

On the other side, let

$$D = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ .. & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_r \end{pmatrix} \quad \Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

Then,

$$U\Sigma = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \dots & \sigma_r\mathbf{u}_r & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}$$

Therefore,

$$U\Sigma = AV \Rightarrow A = U\Sigma V^T$$

since V is orthogonal.

Singular Value Decomposition (SVD)

Theorem 4.3 (Properties of the SVD decomposition)

In a SVD decomposition

- *The left singular vectors of A are eigenvectors of AA^T.*
- *The right singular vectors of A are eigenvectors of A^TA.*
- *The singular values are the square root of the eigenvalues of both AA^T and A^TA.*
- *The singular values are the length of the semiaxes of the mapping of the unit hypersphere in $\mathbb{R}^n$ onto $\mathbb{R}^m$.*
- *The columns of U form an orthogonal basis of $\mathbb{R}^m$.*
- *The columns of V form an orthogonal basis of $\mathbb{R}^n$.*

Singular Value Decomposition (SVD)

Example

Let's calculate the SVD decomposition of $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$.

Step 1: Orthogonally diagonalize $A^T A$

$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$$

Its eigenvalues and eigenvectors are

$$\begin{aligned} \lambda_1 = 360 \quad & \mathbf{v}_1 = \left(\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{2}{3}\right) \\ \lambda_2 = 90 \quad & \mathbf{v}_2 = \left(-\tfrac{2}{3}, -\tfrac{1}{3}, \tfrac{2}{3}\right) \\ \lambda_3 = 0 \quad & \mathbf{v}_3 = \left(\tfrac{2}{3}, -\tfrac{2}{3}, \tfrac{1}{3}\right) \end{aligned}$$

Singular Value Decomposition (SVD)

Step 2: Construct V and Σ

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \end{pmatrix} = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}$$

Step 3: Construct U

$$\mathbf{u}_1 = \frac{A\mathbf{v}_1}{\sigma_1} = \left(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$$
$$\mathbf{u}_2 = \frac{A\mathbf{v}_2}{\sigma_2} = \left(\frac{1}{\sqrt{10}}, -\frac{3}{\sqrt{10}}\right)$$

The set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is already a basis of $\mathbb{R}^2$, so there is no need to extend it.

Singular Value Decomposition (SVD)

Finally we have

$$A = U\Sigma V^T$$

$$\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

MATLAB: `[U,S,V]=svd([4 11 14; 8 7 -2])`

Singular Value Decomposition (SVD)

Example

Let's calculate the SVD decomposition of $A = \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}$.

Step 1: Orthogonally diagonalize $A^T A$

$$A^T A = \begin{pmatrix} 9 & -9 \\ -9 & 9 \end{pmatrix}$$

Its eigenvalues and eigenvectors are

$$\begin{aligned} \lambda_1 = 18 \quad & \mathbf{v}_1 = \left(\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right) \\ \lambda_2 = 0 \quad & \mathbf{v}_2 = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right) \end{aligned}$$

Singular Value Decomposition (SVD)

Step 2: Construct V and Σ

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Step 3: Construct U

$$\mathbf{u}_1 = \frac{A\mathbf{v}_1}{\sigma_1} = \left(\tfrac{1}{3}, -\tfrac{2}{3}, \tfrac{2}{3}\right)$$

The set $\{\mathbf{u}_1\}$ is not yet a basis of $\mathbb{R}^3$, so we need to extend it with orthogonal vectors. All vectors orthogonal to $\mathbf{u}_1$ fulfill

$$\mathbf{u}_1 \cdot \mathbf{u} = 0 = \tfrac{1}{3}x_1 - \tfrac{2}{3}x_2 + \tfrac{2}{3}x_3 \Rightarrow x_1 = 2x_2 - 2x_3$$

Singular Value Decomposition (SVD)

Step 3: Construct U (continued)
A basis of this space is $\mathbf{w}_2 = (2, 1, 0)$ and $\mathbf{w}_3 = (-2, 0, 1)$. But this basis is not orthogonal. Let's make it orthogonal following Gram-Schmidt procedure

$$\begin{array}{l} \mathbf{u}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0) \\ \mathbf{w}_3' = \mathbf{w}_3 - < \mathbf{w}_3, \mathbf{v}_2 > \mathbf{v}_2 = (-\frac{2}{5}, \frac{4}{5}, 1) \\ \mathbf{u}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3} = (-\frac{2}{3\sqrt{5}}, \frac{4}{3\sqrt{5}}, \frac{\sqrt{5}}{3}) \end{array}$$

In fact, SVD does not require the $\mathbf{u}$ vectors to be unitary, but it is simply convenient. We can make $\mathbf{u}_2$ and $\mathbf{u}_3$ unitary because they are "free" (we are constructing them simply to extend the set of $\mathbf{u}$ vectors to be a basis of $\mathbb{R}^3$), but not $\mathbf{u}_1$ because it is "bound" to the singular value.

Singular Value Decomposition (SVD)

Finally we have

$$A = U\Sigma V^T$$

$$\begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & -\frac{2}{3\sqrt{5}} \\ -\frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \end{pmatrix} \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Algebraic applications of SVD

Matrix condition number

Let σ_1 and σ_r be the largest and smallest singular values of a matrix A. The condition number of the matrix is defined as

$$\kappa(A) = \frac{\sigma_1}{\sigma_r}$$

If this condition number is very large, the equations system $A\mathbf{x} = \mathbf{b}$ is ill-posed and small perturbations in $\mathbf{b}$ translate into large perturbations in $\mathbf{x}$. As a rule of thumb, if $\kappa(A) = 10^k$, then you may lose up to k digits of accuracy.

Algebraic applications of SVD

Bases for fundamental spaces

The U and V matrices provide bases for $\mathrm{Row}\{A\}$, $\mathrm{Col}\{A\} = \mathrm{Row}\{A^T\}$, $\mathrm{Nul}\{A\}$ and $\mathrm{Nul}\{A^T\}$

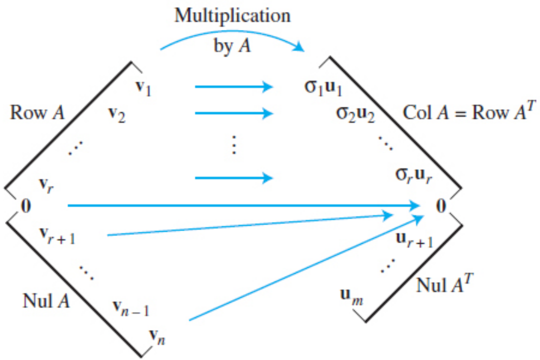


FIGURE 4 The four fundamental subspaces and the action of A.

Algebraic applications of SVD

Theorem 4.4 (The Invertible Matrix Theorem (continued))

The Invertible Matrix Theorem has been developed in Theorems 5.1 and 11.5 of Chapter 3, Theorem 10.5 of Chapter 5, Theorem 2.1 of Chapter 6. Here, we give an extension if A is invertible, then the following statements are equivalent to the previous statements:

xxvii. $(\mathrm{Col}\{A\})^{\perp} = \{\mathbf{0}\}$.

xxviii. $(\mathrm{Nul}\{A\})^{\perp} = \mathbb{R}^n$.

xxix. $(\mathrm{Row}\{A\}) = \mathbb{R}^n$.

xxx. *A has n non-null singular values.*

Algebraic applications of SVD

Reduced SVD and pseudoinverse of A

If within U and V we distinguish two submatrices, each one with r columns we have

$$U = \left(U_r U_{m-r}\right) \text{ and } V = \left(V_r V_{n-r}\right)$$

Then,

$$A = U\Sigma V^T = \left(U_r U_{m-r}\right) \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_r^T \\ V_{n-r}^T \end{pmatrix} = U_r D V_r^T$$

Despite the fact that we may have removed many columns of U and V, we have not lost any information and the recovery of A is exact. The Moore-Penrose pseudoinverse is defined as

$$A^+ = V_r D^{-1} U_r^T$$

that is a $n \times m$ matrix such that

$$A^+AA^+ = A^+ \quad AA^+A = A$$

Algebraic applications of SVD

Pseudoinverse of A and Least Squares

It can be shown that the least-squares solution of the equation system $A\mathbf{x} = \mathbf{b}$ is given by

$$\hat{\mathbf{x}} = A^{+}\mathbf{b}$$

Matrix approximation

If instead of taking r components in the split of U and V (see previous slide) we take only k (assuming singular values have been ordered in descending order), and we reconstruct A_k

$$A_k = U_k D_k V_k^T$$

This matrix is the matrix of rank k that minimizes the Frobenius norm of the difference

$$A_k = \min_{\mathrm{Rank}\{B\}=k} \|A - B\|_F^2 = \min_{\mathrm{Rank}\{B\}=k} \sum_{i,j=1}^{n} (a_{ij} - b_{ij})^2$$

Exercises

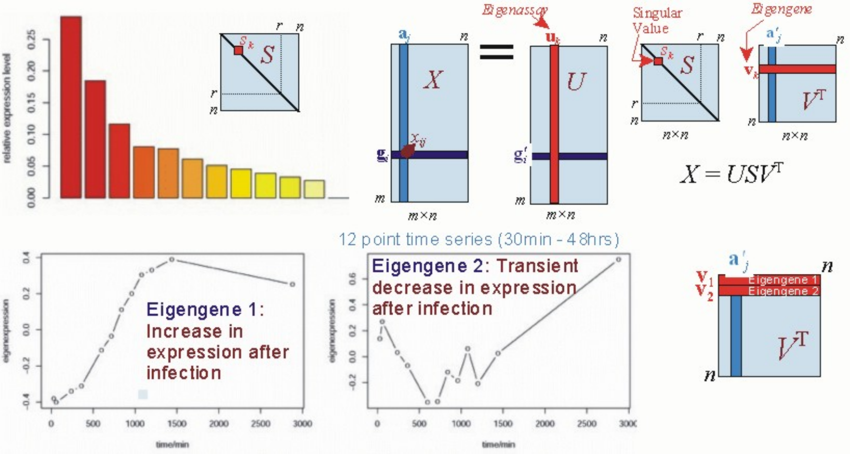
Exercises

From Lay (3rd ed.), Chapter 7, Section 4:

- 7.4.3
- 7.4.11
- 7.4.15
- 7.4.17
- 7.4.18
- 7.4.19
- 7.4.20
- 7.4.23
- 7.4.24

Eigengenes and eigenassays

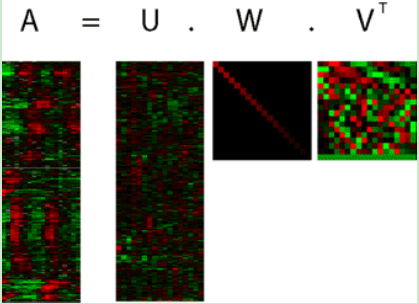
SVD is very much used to analyze the response of different genes to different assays or conditions.



Applications of SVD

Eigengenes and eigenassays

SVD is very much used to analyze the response of different genes to different assays or conditions.



Alter, O., Brown, P. O. and Botstein, D. (2000) Proc. Natl. Acad. Sci. USA 97, 10101

Applications of SVD

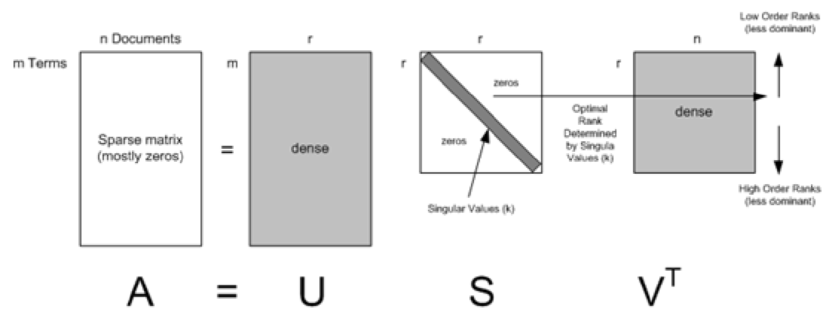
Eigenfaces

In this example we see the effect of matrix approximation by the reduced SVD.

Applications of SVD

Eigenfaces

We can also use SVD to automatically analyze documents.



P. Marksberry, D. Parsley. Managing the IE (Industrial Engineering) *Mindset: A quantitative investigation of Toyota's practical thinking shared among employees*. J. Industrial Engineering and Manegement, 4: 771-799 (**2011**)

Outline

8. Symmetric matrices and quadratic forms
   - Diagonalization of symmetric matrices (a)
   - Quadratic forms (b)
   - Constrained optimization (b)
   - Singular Value Decomposition (SVD) (c)

Chapter 9. Linear algebra applications in geometry

C.O.S. Sorzano

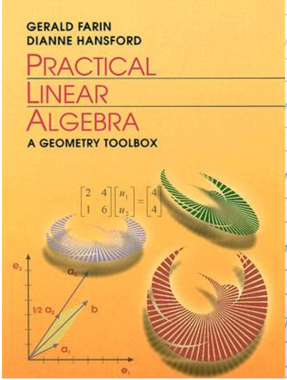
Biomedical Engineering

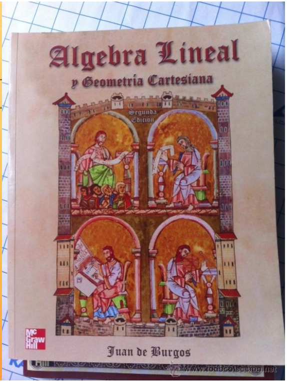
August 25, 2013

Outline

9. Linear algebra applications in geometry
   - Local and global coordinates
   - Points and vectors
   - Lines in 2D
   - Affine maps in 2D
   - Conic sections in 2D
   - 3D Geometry
   - Quadrics in 3D

References





G. Farin, D. Hansford. Practical Linear Algebra: a geometry toolbox. A.K. Peters (2005).
J. de Burgos. Álgebra lineal y geometría cartesiana. McGraw Hill 2ª Ed. (2000)

Outline

9. Linear algebra applications in geometry
   - Local and global coordinates
   - Points and vectors
   - Lines in 2D
   - Affine maps in 2D
   - Conic sections in 2D
   - 3D Geometry
   - Quadrics in 3D

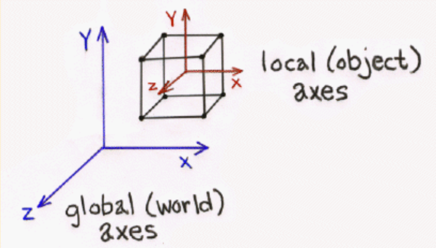
Local and global coordinates

Reference

Farin and Hansford, Chapter 1

Local and global coordinates

In real applications we may need to distinguish between local and global coordinates.



And we need some way of transforming one into the other. This is nothing more than a change of basis.

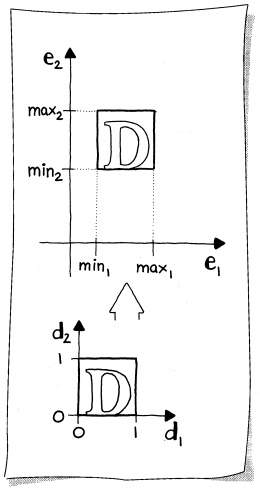
Local and global coordinates

Shift and scale

In Vector Graphics it is common to design objects in a local coordinate system ($\mathbf{d}$) and, then, place, rotate and scale the object in the global coordinate system ($\mathbf{e}$). We need some transformation to go from one space to the other.

For the first component, d_1, we note that we go from a local interval $[0,1]$ to a global interval $[min_1, max_1]$. We may easily perform the transformation as

$$\frac{d_1 - 0}{1 - 0} = \frac{e_1 - min_1}{max_1 - min_1} \Rightarrow$$
$$e_1 = min_1 + (max_1 - min_1) d_1$$



Local and global coordinates

Shift and scale

The more general transformation maps the local interval $[min_{d1}, max_{d1}]$ to the global interval $[min_{e1}, max_{e1}]$. This is achieved with transformation

$$e_1 = min_{e1} + \frac{max_{e1} - min_{e1}}{max_{d1} - min_{d1}} d_1$$

The same kind of transformation is applied to the second component ($d_2 \to e_2$). Putting everything in matrix notation we have

$$\mathbf{e} = \begin{pmatrix} min_{e1} \\ min_{e2} \end{pmatrix} + \begin{pmatrix} \frac{max_{e1} - min_{e1}}{max_{d1} - min_{d1}} & 0 \\ 0 & \frac{max_{e2} - min_{e2}}{max_{d2} - min_{d2}} \end{pmatrix} \mathbf{d}$$

This transformation is of the form

$$\mathbf{e} = T(\mathbf{d}) = \mathbf{e}_{min} + A\mathbf{d}$$

that is not a linear transformation because of the shift (e.g., show that $T(\mathbf{d}_1 + \mathbf{d}_2) \neq T(\mathbf{d}_1) + T(\mathbf{d}_2)$).

Outline

9 Linear algebra applications in geometry

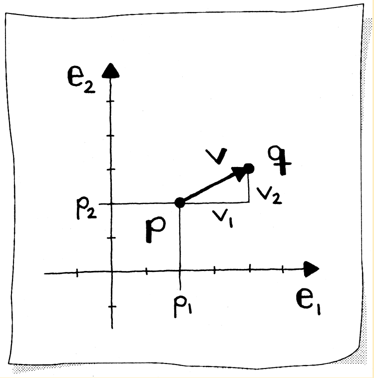
Points and vectors

Reference

Farin and Hansford, Chapter 2

Points and vectors

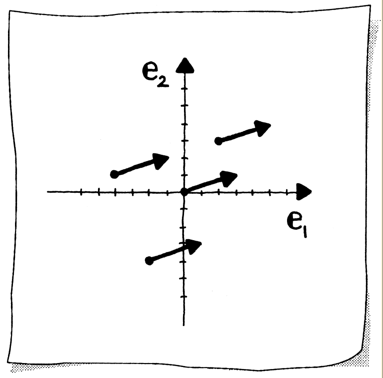
We also need to distinguish between points and vectors. Both are represented as a list of coordinates. Informally, a point indicates a location in space, while a vector indicates a direction (orientation+sense) in space. In this example, we have two points, **p** and **q**, and a vector **v** that goes from **p** to **q**. We may talk about the length of a vector, but not of a point.



Points and vectors

Points and vectors

In this example we have multiple copies of the same vector (since they all have the same direction and magnitude). In Physics, forces are vectors that are applied to objects that are located at points. In this figure we would see the same force applied to different objects.



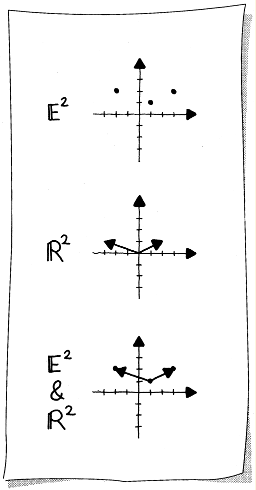
Points and vectors

Points and vectors

More formally, points belong to an Euclidean space while vectors belong to a vector space.

$$\mathbf{p}, \mathbf{q} \in \mathbb{E}^2$$
$$\mathbf{v} \in \mathbb{R}^2$$

Although we may represent both spaces in the same figure and we may define operations using both kinds of spaces. The goal of distinguishing between points and vectors is to distinguish between operations that depend on the coordinate system and operations that do not.



Operations on points and vectors

Coordinate independent operations

$$
\begin{array}{rlll}
- : & \mathbb{E}^2 \times \mathbb{E}^2 \to \mathbb{R}^2 & & \mathbf{v} = \mathbf{q} - \mathbf{p} \\
+ : & \mathbb{E}^2 \times \mathbb{R}^2 \to \mathbb{E}^2 & & \mathbf{p} = \mathbf{q} + \mathbf{v} \\
+ : & \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2 & & \mathbf{v} = \mathbf{u} + \mathbf{w} \\
\cdot : & \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2 & & \mathbf{v} = r\mathbf{u}
\end{array}
$$

Coordinate dependent operations

$$
\begin{array}{rlll}
+ : & \mathbb{E}^2 \times \mathbb{E}^2 \to \mathbb{E}^2 & & \mathbf{t} = \mathbf{p} + \mathbf{q} \\
\cdot : & \mathbb{R} \times \mathbb{E}^2 \to \mathbb{E}^2 & & \mathbf{q} = r\mathbf{p}
\end{array}
$$

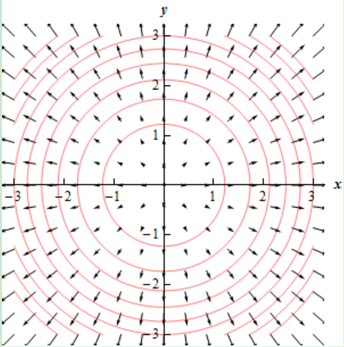
Vector fields

Vector fields

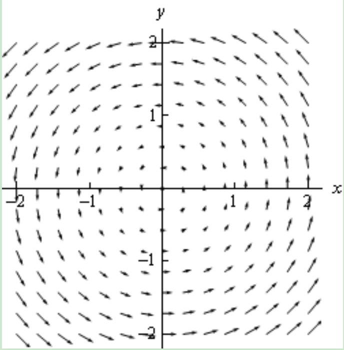
Any function that assigns a vector to a point $f: \quad \mathbb{E}^2 \to \mathbb{R}^2 \quad \mathbf{v} = f(\mathbf{p})$

Example

$f(x, y) = (x, y)$



$f(x, y) = (-y, x)$



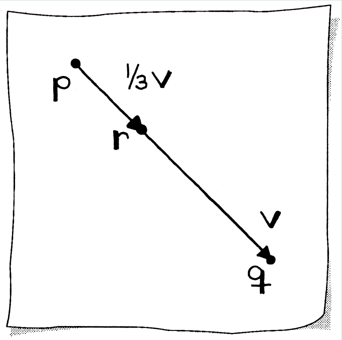
Combinations of points

Barycentric combinations

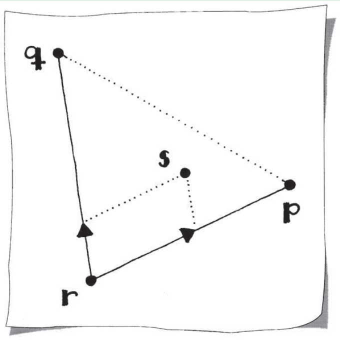
A weighted sum of points where the weights add up to 1 is called a barycentric combination

Example

$$\mathbf{r} = (1-t)\mathbf{p} + t\mathbf{q} = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$$



$$\mathbf{s} = t_1\mathbf{r} + t_2\mathbf{p} + t_3\mathbf{q}$$



Outline

9. Linear algebra applications in geometry
   - Local and global coordinates
   - Points and vectors
   - Lines in 2D
   - Affine maps in 2D
   - Conic sections in 2D
   - 3D Geometry
   - Quadrics in 3D

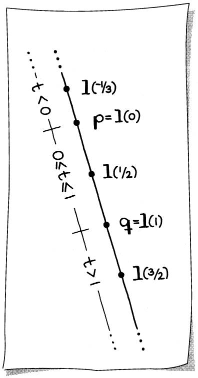
Lines in 2D

Reference

Farin and Hansford, Chapter 3

Parametric equation of a line

- Given two points:
  $\mathbf{l}(t) = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) \quad t \in \mathbb{R}$
- Given point and vector:
  $\mathbf{l}(t) = \mathbf{p} + t\mathbf{v} \quad t \in \mathbb{R}$



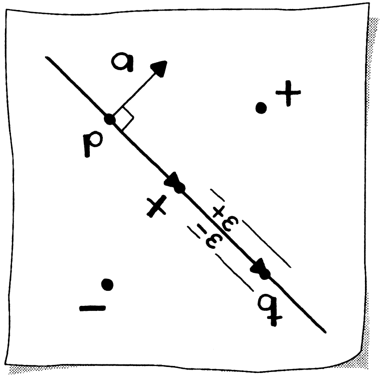
Lines in 2D

Implicit equation of a line

- Given a point and the normal direction:
  $\mathbf{a} \cdot (\mathbf{x} - \mathbf{p}) = 0$

  In 2D:

$$(a_1, a_2) \cdot (x_1 - p_1, x_2 - p_2) = 0 \Rightarrow$$
$$ax_1 + bx_2 + c = 0$$



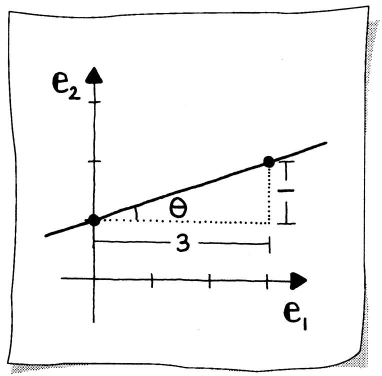
Lines in 2D

Explicit equation of a line

- Given a point and slope:
  In 2D:

$$x_2 = p_2 + m(x_1 - p_1)$$
$$x_2 = mx_1 + b$$
$$x_2 = (\tan\Theta)x_1 + b$$

But it is not a good representation for vertical lines.



Distance of a point to a line

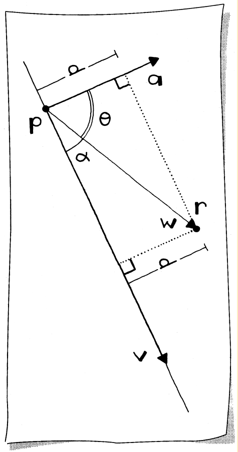
- Implicit line:
  Line: $\mathbf{a} \cdot (\mathbf{x} - \mathbf{p}) = 0$
  Point: $\mathbf{r}$

  Let $\mathbf{w} = \mathbf{r} - \mathbf{p}$ and calculate:

  $$\mathbf{a} \cdot \mathbf{w} = \|\mathbf{a}\|\|\mathbf{w}\| \cos(\theta)$$

  Analyzing the figure we note that $\cos(\theta) = \frac{d}{\|w\|}$. Then

  $$\mathbf{a} \cdot \mathbf{w} = \|\mathbf{a}\| d \Rightarrow d = \frac{\mathbf{a} \cdot \mathbf{w}}{\|\mathbf{a}\|}$$



Lines in 2D

Distance of a point to a line

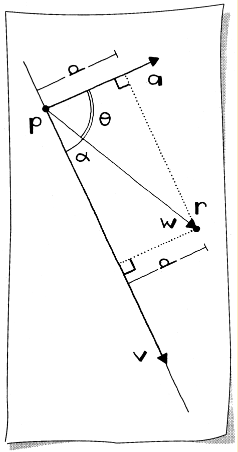
- Parametric line:
  Line: $\mathbf{l}(t) = \mathbf{p} + t\mathbf{v}$
  Point: $\mathbf{r}$

  Let $\mathbf{w} = \mathbf{r} - \mathbf{p}$ and calculate:

  $$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos(\alpha)$$

  Analyzing the figure we note that $\sin(\alpha) = \frac{d}{\|w\|} = \sqrt{1 - \cos^2(\alpha)}$. Then

  $$d = \|\mathbf{w}\|\sqrt{1 - \left(\frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}\right)^2}$$



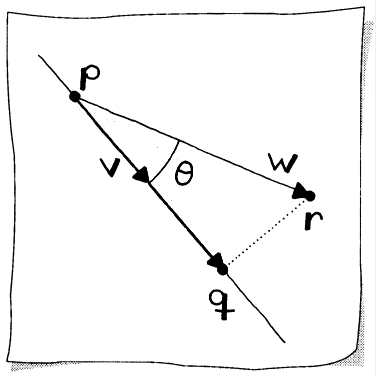
Lines in 2D

The foot of a point

- Parametric line:
  Line: $\mathbf{l}(t) = \mathbf{p} + t\mathbf{v}$
  Point: $\mathbf{r}$

  Let $\mathbf{w} = \mathbf{r} - \mathbf{p}$. The closest point within the line to $\mathbf{r}$ is

$$\mathbf{q} = \mathbf{p} + \mathrm{Proj}_{\mathbf{v}}\{\mathbf{w}\} = \mathbf{p} + \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|^2}\mathbf{v}$$



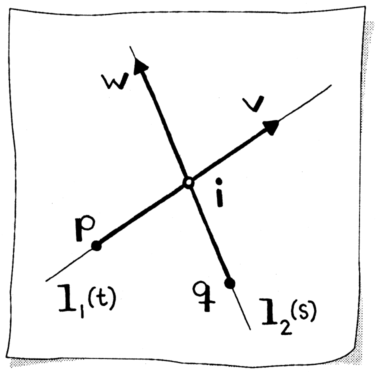
Lines in 2D

The intersection of two lines

- Parametric lines:
  Line 1: $\mathbf{l}_1(t) = \mathbf{p} + t\mathbf{v}$
  Line 2: $\mathbf{l}_2(s) = \mathbf{q} + s\mathbf{w}$

  We need to solve the equation system

$$\mathbf{l}_1(t) = \mathbf{l}_2(s)$$
$$\mathbf{p} + t\mathbf{v} = \mathbf{q} + s\mathbf{w}$$
$$\begin{pmatrix}\mathbf{v} & -\mathbf{w}\end{pmatrix}\begin{pmatrix}t \\ s\end{pmatrix} = \mathbf{q} - \mathbf{p}$$



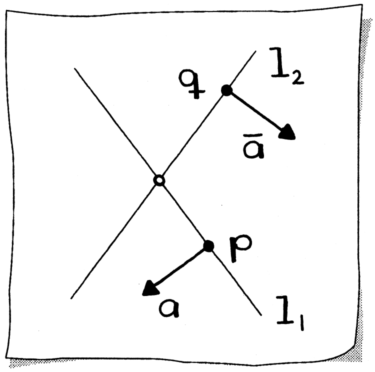
Lines in 2D

The intersection of two lines

- Implicit lines:
  Line 1: $\mathbf{a} \cdot (\mathbf{x} - \mathbf{p}) = 0$
  Line 1: $\bar{\mathbf{a}} \cdot (\mathbf{x} - \mathbf{q}) = 0$

  We need to find $\mathbf{x}$ satisfying both equations at the same time

$$\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\mathbf{p} = 0$$
$$\bar{\mathbf{a}}^T\mathbf{x} - \bar{\mathbf{a}}^T\mathbf{q} = 0$$
$$\begin{pmatrix} \mathbf{a}^T \\ \bar{\mathbf{a}}^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}^T\mathbf{p} \\ \bar{\mathbf{a}}^T\mathbf{q} \end{pmatrix}$$



Outline

9. Linear algebra applications in geometry
   - Local and global coordinates
   - Points and vectors
   - Lines in 2D
   - Affine maps in 2D
   - Conic sections in 2D
   - 3D Geometry
   - Quadrics in 3D

Affine maps in 2D

Reference

Farin and Hansford, Chapter 6

Affine change of coordinates

We transform the point $\mathbf{x}$ into point $\mathbf{x}'$. Note that the matrix multiplication is performed on vectors, not on points

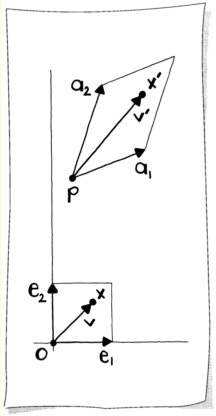
$$\mathbf{v} = \mathbf{x} - \mathbf{o}$$
$$\mathbf{v}' = A\mathbf{v}$$
$$\mathbf{x}' = \mathbf{p} + \mathbf{v}'$$

In total

$$\mathbf{x}' = \mathbf{p} + A(\mathbf{x} - \mathbf{o})$$

We may go back by

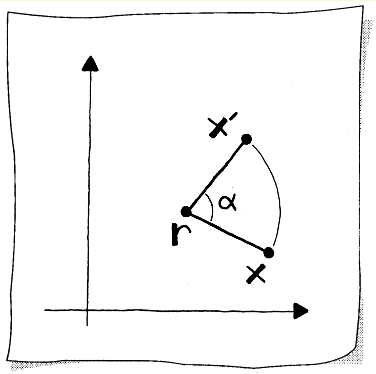
$$\mathbf{x} = \mathbf{o} + A^{-1}(\mathbf{x}' - \mathbf{p})$$



Affine maps in 2D

Translations and rotations

- Translation: $\mathbf{x}' = \mathbf{p} + (\mathbf{x} - \mathbf{o})$
- Rotation: $\mathbf{x}' - \mathbf{r} = R_\alpha(\mathbf{x} - \mathbf{r})$



Affine maps in 2D

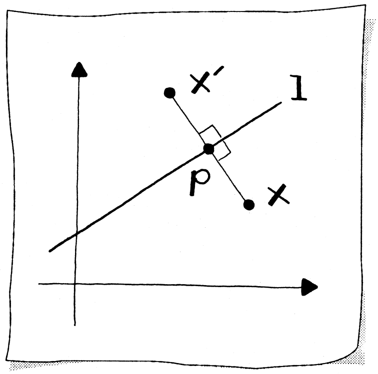
Mirrors and compositions

- Mirror:

$$\begin{aligned} \mathbf{p} &= \tfrac{1}{2}(\mathbf{x} + \mathbf{x}') \\ \mathbf{x}' &= 2\mathbf{p} - \mathbf{x} \end{aligned}$$

- Compositions:

$$\begin{array}{c} \mathbf{x}' = \mathbf{o}' + A(\mathbf{x} - \mathbf{o}) \\ \mathbf{x}'' = \mathbf{o}'' + A'(\mathbf{x}' - \mathbf{o}') \\ \hline \mathbf{x}'' = \mathbf{o}'' + A'A(\mathbf{x} - \mathbf{o}) \end{array}$$



Outline

9 Linear algebra applications in geometry

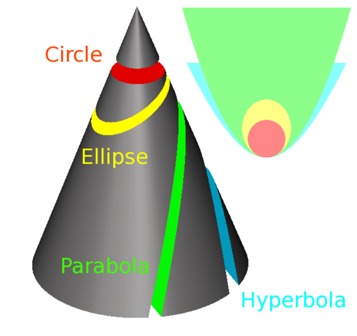
Conic sections

Reference

Juan de Buegos (2000), Capítulo 11

Conic sections

The circle, the ellipse, the parabola, and the hyperbola are all curves stemming from a section of a cone.



Conic sections

Conic sections

They are all second order curves

$$\underbrace{ax_1^2 + bx_2^2 + cx_1x_2}_{\text{2nd order}} + \underbrace{dx_1 + ex_2}_{\text{1st order}} + \underbrace{f}_{\text{0th order}} = 0$$

By renaming the coefficients, we may rewrite it as

$$a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 + 2b_1x_1 + 2b_2x_2 + c = 0$$
$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 2 \begin{pmatrix} b_1 & b_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c = 0$$
$$\mathbf{x}^T A\mathbf{x} + 2B\mathbf{x} + c = 0$$

Compare this to the more widely known equation of the parabola $y = ax^2 + bx + c$. Finally, we can write it in a very compact form

$$\tilde{\mathbf{x}}^T M\tilde{\mathbf{x}} = \begin{pmatrix} x_1 & x_2 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{12} & a_{22} & b_2 \\ b_1 & b_2 & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = 0$$

Conic sections

Definition 5.1 (Conic sections)

*A **conic section** or **conics** is the locus (lugar geométrico) of all points satisfying*

$$\tilde{\mathbf{x}}^T M \tilde{\mathbf{x}} = 0$$

Definition 5.2 (Conic equality)

*Two conics $\tilde{\mathbf{x}}^T M_1 \tilde{\mathbf{x}} = 0$ and $\tilde{\mathbf{x}}^T M_2 \tilde{\mathbf{x}} = 0$ are the **same** if*

$$M_1 = kM_2$$

for some real number k.

Definition 5.3 (Degenerate and ordinary conics)

*A conic section is **degenerate** if*

$$\det\{M\} = 0$$

*A conic section is **ordinary**, if it is not degenerate.*

Conic sections

Examples of ordinary conics

Circumphere	$\frac{x^2}{r^2} + \frac{y^2}{r^2} = 1$
Ellipse	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$
Hyperbola	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
Parabola	$y^2 = 2px$

Examples of degenerate conics

Two lines	$x^2 - y^2 = (x-y)(x+y) = 0$
Two lines	$x^2 - 4 = (x-2)(x+2) = 0$
Two lines (superposed)	$x^2 = 0$
Two complex lines	$x^2 + y^2 = (x-iy)(x+iy) = 0$

Intersection of a conics and a line

Intersection of a conics and a line

Consider the parametric equation of a line in homogeneous coordinates

$$\tilde{\mathbf{l}}(t) = \begin{pmatrix} l_1(t) \\ l_2(t) \\ 1 \end{pmatrix} = \begin{pmatrix} p_1 + tv_1 \\ p_2 + tv_2 \\ 1 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ 1 \end{pmatrix} + t \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} = \tilde{\mathbf{p}} + t\tilde{\mathbf{v}}$$

We need to find a point in the line (i.e., t) such that

$$\begin{gathered} \tilde{\mathbf{l}}(t)^T M \tilde{\mathbf{l}}(t) = 0 \\ (\tilde{\mathbf{p}} + t\tilde{\mathbf{v}})^T M (\tilde{\mathbf{p}} + t\tilde{\mathbf{v}}) = 0 \\ \tilde{\mathbf{v}}^T M \tilde{\mathbf{v}} t^2 + 2\tilde{\mathbf{v}}^T M \tilde{\mathbf{p}} t + \tilde{\mathbf{p}}^T M \tilde{\mathbf{p}} = 0 \end{gathered}$$

This is a second order equation in t. If there is no solution, then the line does not intersect the conics. If there is only 1 solution, then the line is **tangent** to the conics. If there are 2 solutions, then the line intersects the conics (the line is **secant** to the conics, *secante*).

Reduced equation of a conics

Reduced equation of a conics

Let λ_1 and λ_2 be the eigenvalues of $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$. Then, there exists a basis in which the conics can be expressed as

$\lambda_1 \neq 0, \lambda_2 \neq 0$	$\lambda_1 x^2 + \lambda_2 y^2 + \frac{\det\{M\}}{\det\{A\}} = 0$	Ellipses, hyperbolas, pairs of intersecting lines.
$\lambda_1 = 0, \lambda_2 \neq 0$ $\det\{M\} \neq 0$	$y^2 = 2\sqrt{-\frac{\det\{M\}}{\lambda_2^3}}x$	Parabolas
$\lambda_1 = 0, \lambda_2 \neq 0$ $\det\{M\} = 0$	$y^2 = k$	Pairs of parallel lines

General classification of conics

Definition 5.4 (Signature of a quadratic form)

Consider a quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ *and its diagonalization such that*

$$Q(\mathbf{y}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_n y_n^2$$

The signature of $Q(\mathbf{x})$ *is* (n_0, n_+, n_-) *where* n_0 *is the number of null* λ *coefficients,* n_+ *the number of positve* λ *coefficients, and* n_- *the number of negative* λ *coefficients.*

Theorem 5.1

The signature of a quadratic form is invariant to changes of basis, i.e., it only depends on Q*.*

Definition 5.5 (Signature of a matrix)

The signature of a symmetric matrix is the signature of its associated quadratic form.

General classification of conics

General classification of conics

A	M	**Conics**
$\det\{A\} > 0$	$\mathrm{Sig}\{M\} = (0,1,2)$ or $(0,2,1)$	(Real) Ellipse
	$\mathrm{Sig}\{M\} = (0,3)$ or $(0,3,0)$	Empty set (or imaginary ellipse)
	$\mathrm{Det}\{M\} = 0$	A point (or the intersection of two imaginary lines)
$\det\{A\} < 0$	$\det\{M\} \neq 0$	Hyperbola
	$\det\{M\} = 0$	Two secant (real) lines
$\det\{A\} = 0$	$\det\{M\} \neq 0$	Parabola
	$\det\{M\} = 0$	Two parallel (real) lines

Geometric transformations

Geometric transformations

- Shift: Shift the center to $\hat{\mathbf{c}} = (c_1, c_2, 0)$

$$(\tilde{\mathbf{x}} - \hat{\mathbf{c}})^T M_1 (\tilde{\mathbf{x}} - \hat{\mathbf{c}}) = 0$$

- Rotate: Rotate the conics with a rotation matrix R:

$$(R\tilde{\mathbf{x}})^T M_1 (R\tilde{\mathbf{x}}) = 0$$
$$\tilde{\mathbf{x}}^T (R^T M_1 R)\tilde{\mathbf{x}} = 0$$

with $R = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

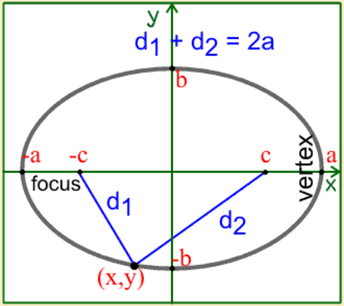
Ellipse

Ellipse

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

Parametric equation: $x = a \cos t$
$y = b \sin t$
$t \in [0, 2\pi)$

Interfocal distance: $d(F, F') = 2c$
where
$a^2 + b^2 = c^2$



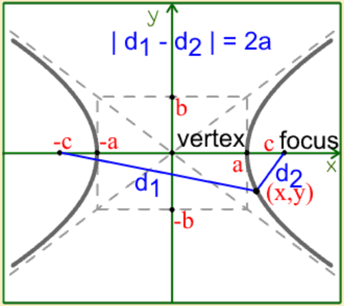
Hyperbola

Hyperbola

Reduced equation: $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

Parametric equation: $x = \pm a \cosh t$
$y = b \sinh t$
$t \in \mathbb{R}$

Interfocal distance: $d(F, F') = 2c$
where
$a^2 + b^2 = c^2$



(Calculus note)

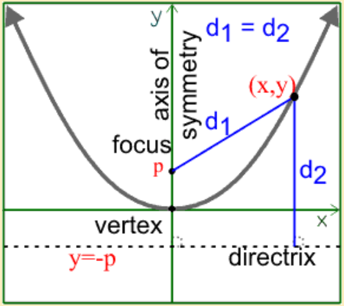
$$\cos x = \frac{e^{ix}+e^{-ix}}{2} \qquad \cosh x = \frac{e^{x}+e^{-x}}{2}$$
$$\sin x = \frac{e^{ix}-e^{-ix}}{2} \qquad \sinh x = \frac{e^{x}-e^{-x}}{2}$$
$$\cos^2 x + \sin^2 x = 1 \qquad \cosh^2 x - \sinh^2 x = 1$$

Parabola

Parabola

Reduced equation: $y^2 = 2px$

Parametric equation: $\begin{cases} x = \frac{t^2}{2p} \\ y = t \\ t \in \mathbb{R} \end{cases}$



Outline

9. Linear algebra applications in geometry
   - Local and global coordinates
   - Points and vectors
   - Lines in 2D
   - Affine maps in 2D
   - Conic sections in 2D
   - 3D Geometry
   - Quadrics in 3D

Cross product

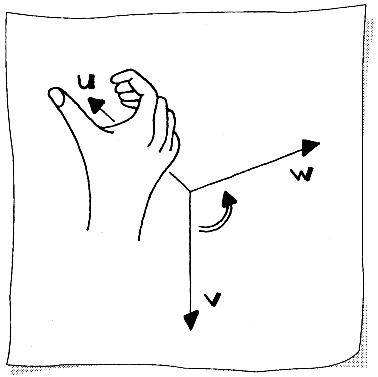
Reference

Farin and Hansford, Chapter 10

Cross product

The cross product is defined for 3D vectors as

$$\mathbf{u} = \mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$



Properties:

$$\mathbf{u} \perp \mathbf{v} \text{ and } \mathbf{u} \perp \mathbf{w}$$
$$\|\mathbf{v} \times \mathbf{w}\|^2 = \|\mathbf{v}\| \|\mathbf{w}\| - (\mathbf{v} \cdot \mathbf{w})^2$$
$$\mathbf{v} \times (c\mathbf{v}) = \mathbf{0}$$
$$\mathbf{v} \times (c\mathbf{w}) = (c\mathbf{v}) \times \mathbf{w} = c(\mathbf{v} \times \mathbf{w})$$
$$\mathbf{w} \times \mathbf{v} = -\mathbf{v} \times \mathbf{w}$$
$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$$
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) \neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$$

Cross product

Example

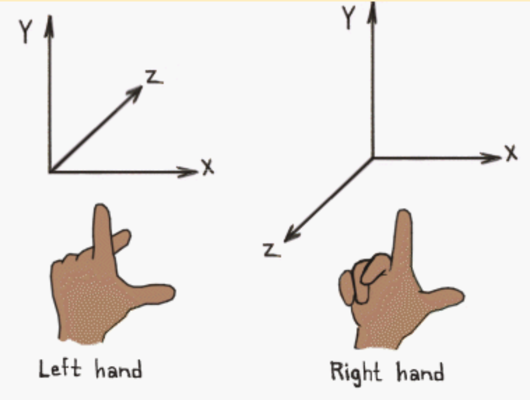
$$\mathbf{u} = \mathbf{e}_1 \times \mathbf{e}_2 = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{e}_3$$

$$\mathbf{u} = \mathbf{e}_2 \times \mathbf{e}_1 = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} = -\mathbf{e}_3$$

Cross product

Coordinate systems

- Right-handed:
  $\mathbf{x} \times \mathbf{y} = \mathbf{z}$
  $\mathbf{y} \times \mathbf{z} = \mathbf{x}$
  $\mathbf{z} \times \mathbf{x} = \mathbf{y}$
- Left-handed:
  $\mathbf{x} \times \mathbf{y} = -\mathbf{z}$
  $\mathbf{y} \times \mathbf{z} = \mathbf{x}$
  $\mathbf{z} \times \mathbf{x} = -\mathbf{y}$

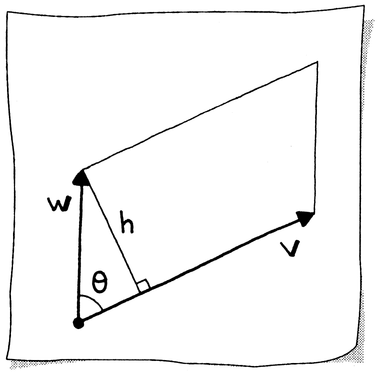


Cross product

Area of parallelogram

The norm of $\mathbf{v} \times \mathbf{w}$ is the area of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$ and is equal to:

$$A = \|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\|\|\mathbf{w}\| \sin(\theta)$$



Lines

Parametric equation of a line

A line is defined in 3D (and nD) by two points or a point and a vector

- Given two points:
  $\mathbf{l}(t) = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) \quad t \in \mathbb{R}$
- Given point and vector:
  $\mathbf{l}(t) = \mathbf{p} + t\mathbf{v} \quad t \in \mathbb{R}$

Giving a point and a perpendicular vector does no longer work.

Planes

Implicit equation of a planes

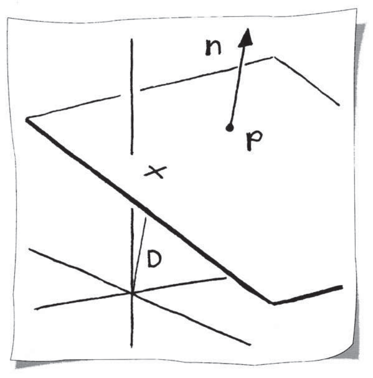
A plane is defined in 3D by a point and a perpendicular vector

- Given a point and the normal direction:
  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$

  In 3D:

  $$(n_1, n_2, n_3) \cdot (x_1 - p_1, x_2 - p_2, x_3 - p_3) = 0 \Rightarrow$$
  $$Ax_1 + Bx_2 + Cx_3 + D = 0$$

  The absolute value of D in the implicit equation is the distance of the plane to the coordinate system origin.



Hyperplanes

Hyperplanes

A hyperplane of $\mathbb{R}^n$ is an affine space of a dimension $n-1$. For instance

$\mathbb{R}^n$	Dimension	Dimension of hyperplane	Hyperplane name
$\mathbb{R}^2$	2D	1	Line
$\mathbb{R}^3$	3D	2	Plane
$\mathbb{R}^n$	nD	n-1	Hyperplane

All hyperplanes are defined by a point ($\mathbf{p}$) and a normal vector ($\mathbf{n}$)

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$

Distance of a point to a plane (hyperplane)

The distance between a point $\mathbf{r}$ and a plane (or hyperplane) is given by

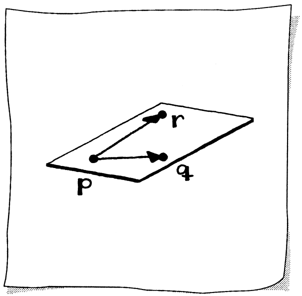
$$d = \frac{\mathbf{n} \cdot (\mathbf{r} - \mathbf{p})}{\|\mathbf{n}\|}$$

Planes

Parametric equation of a plane

A plane can also be defined in 3D (and nD) by a point and two in-plane vectors

- Given a point and two in-plane vectors:
  $P(s,t) = \mathbf{p} + s\mathbf{v} + t\mathbf{w} \quad \forall s,t \in \mathbb{R}$
- Given three points:
  $P(s,t) = \mathbf{p} + s(\mathbf{q} - \mathbf{p}) + t(\mathbf{r} - \mathbf{p}) \quad \forall s,t \in \mathbb{R}$



Scalar triple product

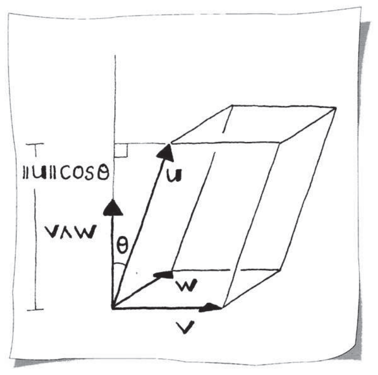
Scalar triple product

The volume of a parallelepiped can be measured with the scalar triple product

$$V = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$$

Properties:

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{u} \times \mathbf{w}) = \mathbf{w} \cdot (\mathbf{v} \times \mathbf{u})$$



Distance between two lines

Given two lines in parametric form

$$\mathbf{l}_1(s_c) = \mathbf{p}_0 + s_c\mathbf{u} \ \mathbf{l}_2(t_c) = \mathbf{q}_0 + t_c\mathbf{v}$$

The distance between the two lines is the length of the vector $\mathbf{w}_c$ that is perpendicular to both lines. $\mathbf{w}_c$ is defined by two points: one in line 1 ($\mathbf{x}_1$) and another one in line 2 ($\mathbf{x}_2$):

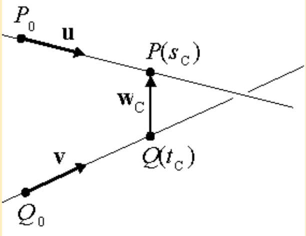
$$\mathbf{w}_c = \mathbf{x}_2 - \mathbf{x}_1 = \mathbf{q}_0 + t_c\mathbf{v} - (\mathbf{p}_0 + s_c\mathbf{u})$$

The conditions on $\mathbf{w}_c$ are:

$$\mathbf{w}_c \cdot \mathbf{u} = 0 \text{ and } \mathbf{w}_c \cdot \mathbf{v} = 0$$

After reorganizing the terms

$$\begin{pmatrix} \|\mathbf{u}\|^2 & -\mathbf{u}\cdot\mathbf{v} \\ \mathbf{u}\cdot\mathbf{v} & \|\mathbf{v}\|^2 \end{pmatrix} \begin{pmatrix} s_c \\ t_c \end{pmatrix} = \begin{pmatrix} (\mathbf{p}_0 - \mathbf{q}_0)\cdot\mathbf{u} \\ (\mathbf{p}_0 - \mathbf{q}_0)\cdot\mathbf{v} \end{pmatrix}$$



Intersection of two lines

Intersection of two lines

The two lines in the previous slide intersect if $\mathbf{x}_1 = \mathbf{x}_2$. We also note that the two lines intersect if $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{p}_0 - \mathbf{q}_0$ are in the same plane, or what is the same they are linearly dependent

$$\left| \begin{pmatrix} \mathbf{u} & \mathbf{v} & \mathbf{p}_0 - \mathbf{q}_0 \end{pmatrix} \right| = 0$$

Intersection of a line and a plane

Intersection of a line and a plane

- Parametric line, implicit plane:

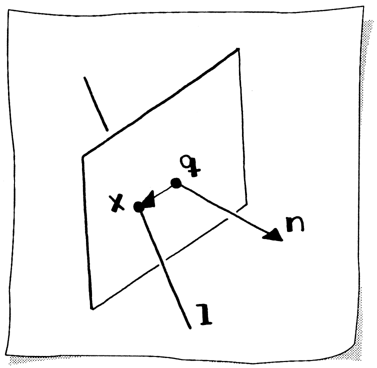
$$\mathbf{l}(t) = \mathbf{p} + t\mathbf{v}$$
$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{q}) = 0$$

For the intersection we need to find t such that

$$\mathbf{n} \cdot (\mathbf{p} + t\mathbf{v} - \mathbf{q}) = 0$$

whose solution is

$$t = \frac{\mathbf{n}\cdot(\mathbf{q}-\mathbf{p})}{\mathbf{n}\cdot\mathbf{v}}$$
$$\mathbf{x} = \mathbf{p} + \frac{\mathbf{n}\cdot(\mathbf{q}-\mathbf{p})}{\mathbf{n}\cdot\mathbf{v}}\mathbf{v}$$



Intersection of a line and a plane

- Parametric line, parametric plane:

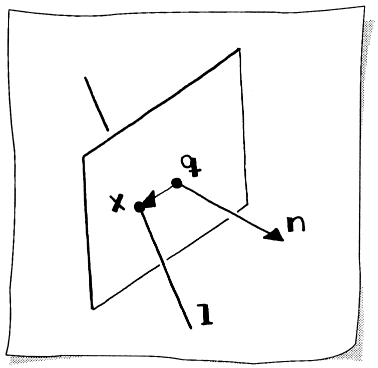
$$\mathbf{l}(t) = \mathbf{p} + t\mathbf{v}$$
$$\mathbf{P}(t_1, t_2) = \mathbf{q} + t_1\mathbf{u} + t_2\mathbf{w}$$

We need to find t, t_1 and t_2 such that

$$\mathbf{p} + t\mathbf{v} = \mathbf{q} + t_1\mathbf{u} + t_2\mathbf{w}$$

Reorganizing the terms:

$$\begin{pmatrix}\mathbf{u} & \mathbf{w} & -\mathbf{v}\end{pmatrix}\begin{pmatrix}t_1\\ t_2\\ t\end{pmatrix} = \mathbf{p} - \mathbf{q}$$



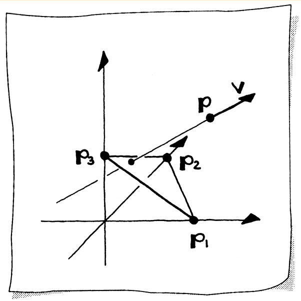
Intersection of a line and a triangle

- Parametric line, 3 points of a triangle:

$$\mathbf{l}(t) = \mathbf{p} + t\mathbf{v}$$

$$\mathbf{P}(t_1, t_2) \quad = \quad \mathbf{p}_1 + t_1(\mathbf{p}_2 - \mathbf{p}_1) + t_2(\mathbf{p}_3 - \mathbf{p}_1)$$

$$t_1, t_2 \in [0, 1], t_1 + t_2 \leq 1$$



We need to find t, t_1 and t_2 such that

$$\mathbf{p} + t\mathbf{v} = \mathbf{p}_1 + t_1(\mathbf{p}_2 - \mathbf{p}_1) + t_2(\mathbf{p}_3 - \mathbf{p}_1)$$

Reorganizing the terms:

$$\begin{pmatrix} \mathbf{p}_2 - \mathbf{p}_1 & \mathbf{p}_3 - \mathbf{p}_1 & -\mathbf{v} \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t \end{pmatrix} = \mathbf{p} - \mathbf{p}_1$$

The intersection point is within the triangle if $t_1, t_2 \in [0, 1], t_1 + t_2 \leq 1$.

Reflection

- Reflection:
  This situation is encountered, for instance, in reflected light rays. By inspecting the figure we note that

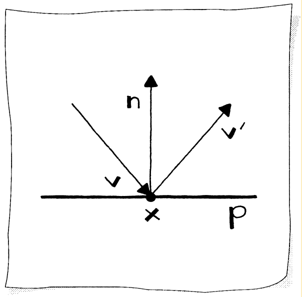
$$\mathbf{n} \cdot \mathbf{v} = -\mathbf{n} \cdot \mathbf{v}'$$

On the other side, it must also be

$$c\mathbf{n} = \mathbf{v}' - \mathbf{v}$$

We have two unknowns c and $\mathbf{v}$ and two equations. After some manipulation we reach

$$\mathbf{v}' = \mathbf{v} - 2(\mathbf{n} \cdot \mathbf{n}^T)\mathbf{v}$$



Intersection of three planes

Intersection of three planes

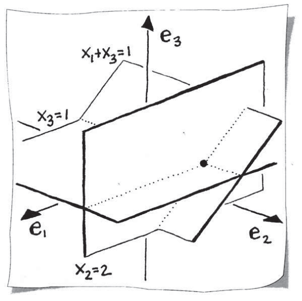
- Implicit equations:
  For each of the planes, we have

$$\mathbf{n}_1 \cdot (\mathbf{x} - \mathbf{p}_1) = 0 \Rightarrow \mathbf{n}_1^T \mathbf{x} = \mathbf{n}_1^T \mathbf{p}_1$$
$$\mathbf{n}_2 \cdot (\mathbf{x} - \mathbf{p}_2) = 0 \Rightarrow \mathbf{n}_2^T \mathbf{x} = \mathbf{n}_2^T \mathbf{p}_2$$
$$\mathbf{n}_3 \cdot (\mathbf{x} - \mathbf{p}_3) = 0 \Rightarrow \mathbf{n}_3^T \mathbf{x} = \mathbf{n}_3^T \mathbf{p}_3$$

  Gathering all together

$$\begin{pmatrix} \mathbf{n}_1^T \\ \mathbf{n}_2^T \\ \mathbf{n}_3^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{n}_1^T \mathbf{p}_1 \\ \mathbf{n}_2^T \mathbf{p}_2 \\ \mathbf{n}_3^T \mathbf{p}_3 \end{pmatrix}$$

  In non-degenerate situations, this equation system has a unique solution that is the intersection point. Otherwise, the planes may intersect in one line, two lines, three lines, or even in a plane (if the three planes are the same plane).



Intersection of two planes

Intersection of two planes

- Implicit equations:

  For each of the planes, we have

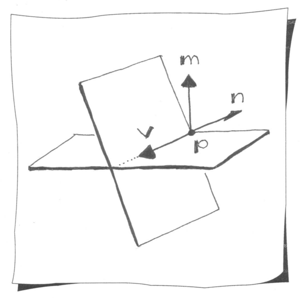
$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}_1) = 0 \Rightarrow \mathbf{n}^T\mathbf{x} = \mathbf{n}^T\mathbf{p}_1$$
$$\mathbf{m} \cdot (\mathbf{x} - \mathbf{p}_2) = 0 \Rightarrow \mathbf{m}^T\mathbf{x} = \mathbf{m}^T\mathbf{p}_2$$

  The two planes intersect in a line of the form

$$\mathbf{l}(t) = \mathbf{p} + t(\mathbf{n} \times \mathbf{m})$$

  To find $\mathbf{p}$ we solve the equation system

$$\begin{pmatrix} \mathbf{n}^T \\ \mathbf{m}^T \\ (\mathbf{n} \times \mathbf{m})^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{n}^T\mathbf{p}_1 \\ \mathbf{m}^T\mathbf{p}_2 \\ 0 \end{pmatrix}$$



Outline

9 Linear algebra applications in geometry

- Local and global coordinates
- Points and vectors
- Lines in 2D
- Affine maps in 2D
- Conic sections in 2D
- 3D Geometry
- Quadrics in 3D

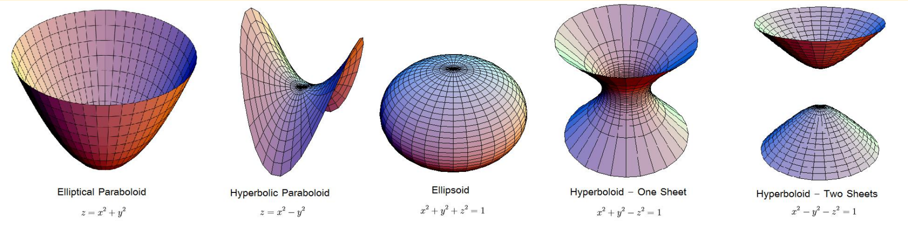
Quadrics

Reference

Juan de Buegos (2000), Capítulo 12

Quadrics

Quadrics are 3D surfaces that meet a second order equation.



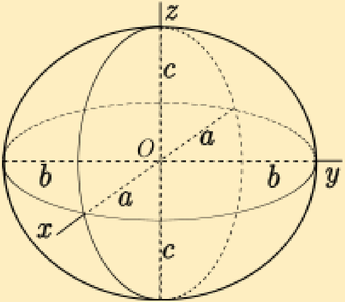
Quadrics in the Wikipedia

Ellipsoid

Ellipsoid

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

Parametric equation:
$$\begin{aligned} x &= a \cos u \sin v \\ y &= b \sin u \sin v \\ z &= c \cos v \\ u, v &\in [0, 2\pi) \end{aligned}$$



Cuts along X, Y and Z are ellipses.

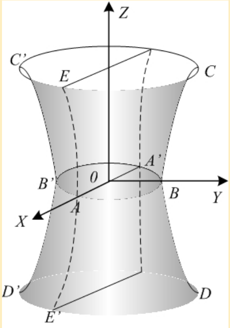
Hyperboloid of one sheet

Hyperboloid of one sheet

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$

Parametric equation:
$$x = a\sqrt{1+u^2}\cos v$$
$$y = b\sqrt{1+u^2}\sin v$$
$$z = cu$$

Parametric equation:
$$x = a\cosh u \cos v$$
$$y = b\cosh u \sin v$$
$$z = c\sinh u$$
$$v \in [0, 2\pi),\ u \in \mathbb{R}$$



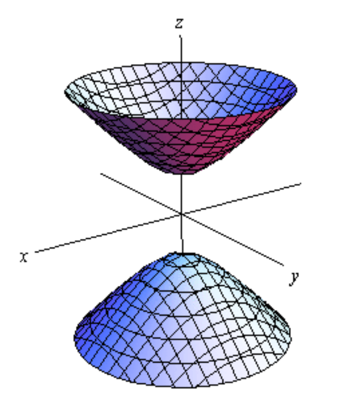
Cuts along X and Y are hyperbolas.
Cuts along Z are ellipses.

Hyperboloid of two sheets

Hyperboloid of two sheets

Reduced equation: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$

Parametric equation:
$$x = a\sinh u\cos v$$
$$y = b\sinh u\sin v$$
$$z = c\cosh u$$
$$v\in[0,2\pi),\ u\in\mathbb{R}$$



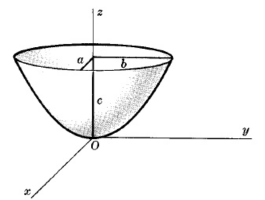
Cuts along X and Y are hyperbolas.
Cuts along Z are ellipses.

Elliptic paraboloid

Elliptic paraboloid

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z}{c} = 0$

Parametric equation:
$$\begin{aligned} x &= a\sqrt{u}\cos v \\ y &= b\sqrt{u}\sin v \\ z &= cu \end{aligned}$$
$$v \in [0, 2\pi),\ u \in [0, \infty)$$



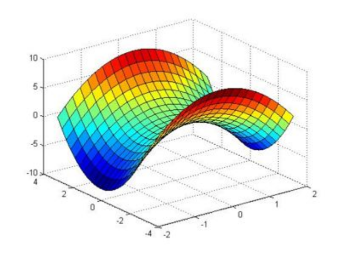
Cuts along X and Y are parabolas.
Cuts along Z are ellipses.

Hyperbolic paraboloid

Hyperbolic paraboloid

Reduced equation: $\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z}{c} = 0$

Parametric equation:
$$x = a\sqrt{u}\cosh v$$
$$y = b\sqrt{u}\sinh v$$
$$z = cu$$
$$u, v \in \mathbb{R}$$



Cuts along Y are parabolas.
Cuts along Z are hyperbolas.

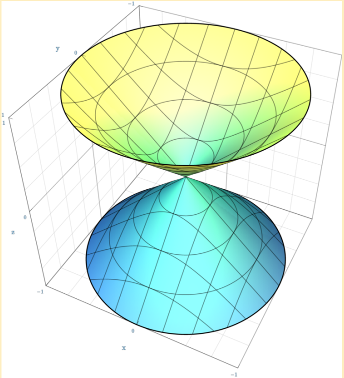
Cone

Cone

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$

Parametric equation:
$$x = au\cos v$$
$$y = bu\sin v$$
$$z = cu$$
$$v \in [0, 2\pi), u \in \mathbb{R}$$

Cuts along Y are parabolas.
Cuts along Z are ellipses.



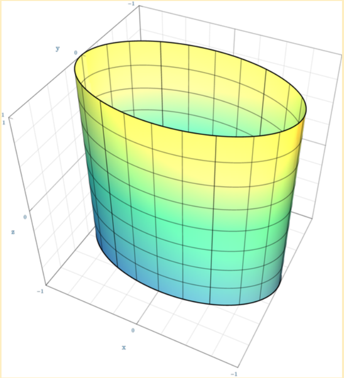
Elliptic cylinder

Elliptic cylinder

Reduced equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

Parametric equation:
$$x = a\cos v$$
$$y = b\sin v$$
$$z = u$$
$$v \in [0, 2\pi), u \in \mathbb{R}$$

Cuts along X and Y are pairs of lines.
Cuts along Z are ellipses.



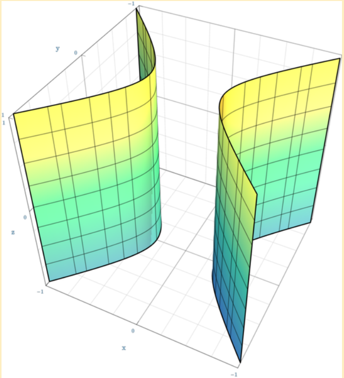
Hyperbolic cylinder

Hyperbolic cylinder

Reduced equation: $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

Parametric equation:
$$\begin{aligned} x &= a\cosh v \\ y &= b\sinh v \\ z &= u \\ u, v &\in \mathbb{R} \end{aligned}$$

Cuts along X and Y are pairs of lines.
Cuts along Z are hyperbolas.

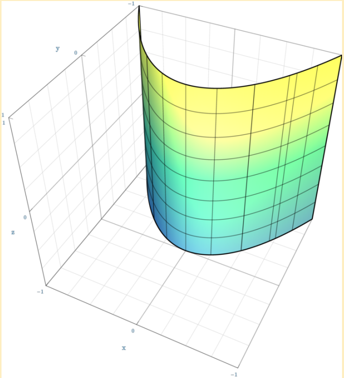


Parabolic cylinder

Parabolic cylinder

Reduced equation: $\frac{x^2}{a^2} - \frac{y}{b} = 0$

Parametric equation: $$\begin{aligned} x &= au \\ y &= bu^2 \\ z &= v \\ u, v &\in \mathbb{R} \end{aligned}$$



Cuts along X and Y are pairs of lines or single lines.
Cuts along Z are parabolas.

Quadrics

Definition 7.1

Quadrics All quadrics can be written as

$$\sum_{i,j=1}^{3} a_{ij}x_ix_j + 2\sum_{i=1}^{3} b_ix_i + c = 0$$
$$\tilde{\mathbf{x}}^T M\tilde{\mathbf{x}} = 0$$

with $a_{ij} = a_{ji}$ *and*

$$M = \begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{12} & a_{22} & a_{23} & b_2 \\ a_{13} & a_{23} & a_{33} & b_3 \\ b_1 & b_2 & b_3 & c \end{pmatrix} \text{ and } \tilde{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}$$

Quadrics

Definition 7.2 (Quadrics equality)

Two quadrics $\tilde{\mathbf{x}}^T M_1 \tilde{\mathbf{x}} = 0$ *and* $\tilde{\mathbf{x}}^T M_2 \tilde{\mathbf{x}} = 0$ *are the* ***same*** *if*

$$M_1 = kM_2$$

for some real number k*.*

Definition 7.3 (Degenerate or ordinary quadrics)

A quadrics is ***degenerate*** *if* $\det\{M\} = 0$ *(e.g., cones, cylinders and pairs of planes). It is ordinary if it is not degenerate (e.g., ellipsoids, paraboloids, hyperboloids)*

Examples of degenerate quadrics

$x^2 - y^2 = 0 = (x-y)(x+y)$	A pair of planes
$x^2 + y^2 = 0 = (x-iy)(x+iy)$	A pair of imaginary planes
$x^2 - 1 = 0 = (x-1)(x+1)$	A pair of planes
$x^2 + y^2 - 25 = 0$	Cylinder of radius 5

General classification of quadrics

Let λ_1, λ_2 and λ_3 be the eigenvalues of A. Then, there exists a basis such that the reduced equation of the quadrics is

Condition	**Quadrics**
$\lambda_1 \neq 0, \lambda_2 \neq 0, \lambda_3 \neq 0$	$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + \frac{\det\{M\}}{\det\{A\}} = 0$ Ellipsoids, hyperboloids and cones
$\lambda_1 \neq 0, \lambda_2 \neq 0, \lambda_3 = 0$ $\det\{M\} \neq 0$	$\lambda_1 x^2 + \lambda_2 y^2 = 2\sqrt{-\frac{\det\{M\}}{\lambda_1\lambda_2}}z$ Paraboloid
$\lambda_1 \neq 0, \lambda_2 \neq 0, \lambda_3 = 0$ $\det\{M\} = 0$	$\lambda_1 x^2 + \lambda_2 y^2 = k$ Elliptical cylinder
$\lambda_1 = 0, \lambda_2 \neq 0, \lambda_3 = 0$ $\mathrm{Rank}\{M\} = 3$	$y^2 = 2qx$ Parabolic cylinder
$\lambda_1 = 0, \lambda_2 \neq 0, \lambda_3 = 0$ $\mathrm{Rank}\{M\} < 3$	$y^2 = k$ Pair of planes

Outline

9. Linear algebra applications in geometry
    - Local and global coordinates
    - Points and vectors
    - Lines in 2D
    - Affine maps in 2D
    - Conic sections in 2D
    - 3D Geometry
    - Quadrics in 3D

Chapter 10. Abstract algebra

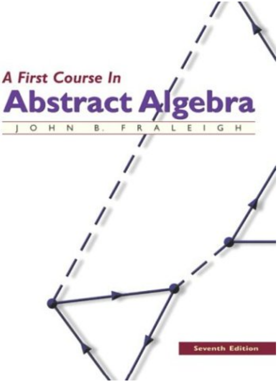
C.O.S. Sorzano

Biomedical Engineering

August 26, 2013

Outline

References



J.B. Fraleigh. A first course in Abstract Algebra. Pearson, 7th Ed. (2002)

Outline

10. Abstract algebra
    - Sets
    - Relations and functions
    - Partitions and equivalence relationships
    - Binary operations
    - Groups and subgroups
    - Homomorphisms and isomorphisms
    - Algebraic structures

Sets

Definition 1.1 (Set)

*A **set** is a well-defined collection of **elements**. We denote the different elements as $a \in S$.*

Definition 1.2 (Empty set)

*The only set without any element is the **empty set** ($\emptyset$).*

Describing sets

We may provide the elements of a set:

- Intensional definition: by giving a property they all meet (*e.g., even numbers from 1 to 10*)
- Extensional definition: by listing all the elements in the set (*e.g.,*$\{2, 4, 6, 8, 10\}$). The order in which the different elements are written has no meaning.

Sets

Definition 1.3 (Subset and proper subset)

*B is **subset** of A (denoted $B \subseteq A$ or $A \supseteq B$) if all the elements of B are also elements of A. B is a **proper subset** of A if B is a subset of A and B is different from A ($B \subset A$ or $A \supset B$).*

Properties

- A is an improper subset of A.
- $\emptyset$ is a proper subset of A.

Definition 1.4 (Power set (Partes de un conjunto))

*The set of all subsets of a set A is called the **power set** of A.*

Example

Let $A = \{1, 2, 3\}$ the power set of A is

$$P(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

Definition 1.5 (Cartesian product)

The cartesian product of the sets A and B is the set of all ordered pairs in which the first element comes from A and the second element comes from B.

$$A \times B = \{(a, b) | a \in A, b \in B\}$$

Note that because of the ordered nature of the pair $A \times B \neq B \times A$.

Example

Let $A = \{1, 2, 3\}$ and $B = \{4, 5\}$.

$$A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

Definition 1.6 (Cardinality)

The cardinality of a set is the number of elements it has.

Sets

Definition 1.7 (Disjoint sets)

Two sets are disjoint if they do not have any element in common.

Some useful sets

- Integer numbers: $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$, $|\mathbb{Z}| = \aleph_0$
- Natural numbers, positive integers: $\mathbb{N} = \mathbb{Z}^+ = \{1, 2, 3, ...\}$, $|\mathbb{N}| = \aleph_0$
- Negative integers: $\mathbb{Z}^- = \{..., -3, -2, -1\}$, $|\mathbb{Z}^-| = \aleph_0$
- Non-null integers: $\mathbb{Z}^* = \mathbb{Z} - \{0\} = \{..., -2, -1, 1, 2, ...\}$, $|\mathbb{Z}^*| = \aleph_0$
- Rational numbers: $\mathbb{Q}$, $|\mathbb{Q}| = \aleph_0$
- Real numbers: $\mathbb{R}$, $|\mathbb{R}| = \aleph_1$
- Interval: $[0, 1]$, $|[0, 1]| = \aleph_1$
- Complex numbers: $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}\}$, $|\mathbb{C}| = \aleph_1$

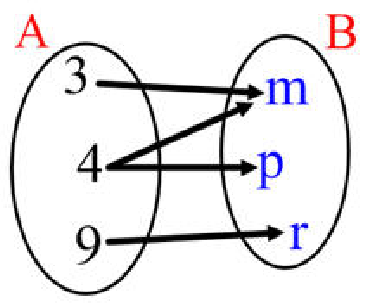
Outline

Relations

Definition 2.1 (Relation)

*A **relation** aRb is a subset of the cartesian product $A \times B$.*

Example



Functions

Definition 2.2 (Function)

*A **function** $f : X \to Y$ is a relation between X and Y in which each $x \in X$ appears at most in one of the pairs (x, y). We may write*

$$(x, y) \in f \text{ or } f(x) = y$$

*The **domain** of f is X, the **codomain** of f is Y. The **support** of f is the set of all those values in X for which there exists a pair (x, y). The **range** of f are all values in Y for which there exists at least one pair (x, y).*

Example

$$\begin{array}{c} \begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ f(x) &= x^3 \end{aligned} \\ (2,8) \in f \Leftrightarrow f(2) = 8 \\ \hline +: \mathbb{R} \times \mathbb{R} \to \mathbb{R} \\ ((2,3),5) \in + \Leftrightarrow +((2,3)) = 5 \Leftrightarrow 2+3=5 \end{array}$$

Classification of functions

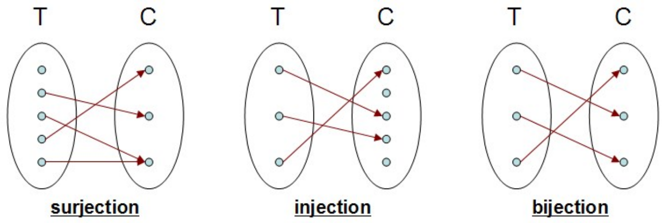
Definition 2.3

Functions can be classified as ***surjective, injective or bijective****:*

Surjective: *A function is surjective if every point of the codomain has* ***at least one*** *point of the domain that maps onto it. They are also called* ***onto*** *functions.*

Injective: *A function is injective if every point of the codomain has* ***at most one*** *point in the domain that maps onto it. They are also called* ***one-to-one*** *functions.*

Bijective: *A function is bijective if it is injective and surjective.*



Inverse function

Definition 2.4 (Inverse function)

Consider an injective function $f : X \to Y$. $f^{-1} : Y \to X$ is the ***inverse*** *of f iff*

$$(x, y) \in f \Rightarrow (y, x) \in f^{-1}$$

Example

- $f(x) = x + 3 \Rightarrow f^{-1}(y) = y - 3$
- $f(x) = x^3 \Rightarrow f^{-1}(y) = y^{\frac{1}{3}}$
- $f(x) = x^2$ is not invertible because it is not injective $(f(-2) = f(2) = 4)$

Inverse function

Theorem 2.1

- *If f is invertible, its inverse is unique.*
- *If f is bijective, so is f^{-1}.*
- *X and Y have the same cardinality if there exists a bijective function between the two.*

Example

Consider the following function $f : \mathbb{Z} \to \mathbb{N}$

$$\begin{array}{cccccccc} 0 & -1 & 1 & -2 & 2 & -3 & 3 & \dots \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & \dots \end{array}$$

$$f = \{(0,0),(-1,1),(1,2),(-2,3),(2,4),(-3,5),(3,6),\dots\}$$

f is bijective. Consequently, $\mathbb{Z}$ has the same cardinality as $\mathbb{N}$.

Outline

Partition

Definition 3.1 (Partition)

A partition of a set S is a collection of non-empty subsets such that each element of S belongs to one and only one subset (cell) of the partition. We denote as $\bar{x}$ the subset that contains the element x. All cells in a partition are disjoint to any other cell.

Examples

- We may partition the set of natural numbers into the subset of even numbers ($\{2, 4, 6, \ldots\}$) and the subset of odd numbers ($\{1, 3, 5, \ldots\}$).
- We may partition the set of integer numbers into the subset of all multiples of 3 ($\{\ldots, -6, -3, 0, 3, 6, \ldots\}$), the subset of all numbers whose remainder after dividing by 3 is 1 ($\{\ldots, -5, -2, 1, 4, 7, \ldots\}$), and the subset of all numbers whose remainder after dividing by 3 is 2 ($\{\ldots, -4, -1, 2, 5, 8, \ldots\}$).

Equivalence relation

Definition 3.2 (Equivalence relation)

*R is an **equivalence relation** in S if it verifies:*

1. *R is **reflexive**: xRx*
2. *R is **symmetric**: $xRy \Rightarrow yRx$*
3. *R is **transitive**: $xRy, yRz \Rightarrow xRz$*

Examples

1. $=$ is an equivalence relation.
2. Congruence modulo n is an equivalence relation (two numbers are related if they have the same remainder after dividing by n)
   Example: 1 and 4 have remainder 1 after dividing by 3. We write
   $$1 \equiv 4(mod 3)$$
3. $\forall n, m \in \mathbb{Z} \quad nRm \Leftrightarrow nm \geq 0$ is not an equivalence relationship because it is not transitive (e.g., $-3R0$, $0R5$ but $-3\not R5$).

Partition and equivalence relation

Theorem 3.1

Let S be a non-empty set, and R an equivalence relation defined on S. Then R partitions S with the cells

$$\bar{a} = \{x \in S | xRa\}$$

Additionally, we may define another equivalence relation $\sim$

$$a \sim b \Leftrightarrow \bar{a} = \bar{b}$$

Partition and equivalence relation

Example

Congruence modulo 3 is an equivalence relation in $\mathbb{Z}$ (two numbers are related if they have the same remainder after dividing by 3)

$$\begin{aligned}\bar{0} &= \{..., -6, -3, 0, 3, 6, ...\}\\ \bar{1} &= \{..., -5, -2, 1, 4, 7, ...\}\\ \bar{2} &= \{..., -4, -1, 2, 5, 8, ...\}\end{aligned}$$

Additionally

$$\begin{aligned}... = \bar{0} = \bar{3} = \bar{6} = ... &\Rightarrow 0 \sim 3 \sim 6 \sim ...\\ ... = \bar{1} = \bar{4} = \bar{7} = ... &\Rightarrow 1 \sim 4 \sim 7 \sim ...\\ ... = \bar{2} = \bar{5} = \bar{8} = ... &\Rightarrow 2 \sim 5 \sim 8 \sim ...\end{aligned}$$

and

$$\mathbb{Z} = \bar{0} \cup \bar{1} \cup \bar{2}$$

Partition and equivalence relation

Example

Consider the cartesian product $\mathbb{Z} \times (\mathbb{Z} - \{0\})$. Let (m_1, n_1) and (m_2, n_2) be two ordered sets of this cartesian product. Consider now the equivalence relation

$$(m_1, n_1 \sim (m_2, n_2) \Leftrightarrow m_1 n_2 - m_2 n_1 = 0$$

The set of rational numbers is formally defined $\mathbb{Q}$ as the set of equivalence classes of $\mathbb{Z} \times (\mathbb{Z} - \{0\})$ under the relation $\sim$.

Outline

10 Abstract algebra

- Sets
- Relations and functions
- Partitions and equivalence relationships
- Binary operations
- Groups and subgroups
- Homomorphisms and isomorphisms
- Algebraic structures

Binary operations

Introduction

What is addition?

Let us assume that we arrive to a classroom in Mars, and that martians are learning to add. The teacher says

Gloop, poyt

and the students reply:

Bimt.

Then, the teacher says:

Ompt, gaft

and the students reply:

Poyt.

We don't know what they do but it seems that when the teacher gives two elements, students respond with another element.

Binary operations

Introduction (continued)

What is addition?
This is what we do when we say “three plus four”, “seven”. And we may not use any two elements (“three plus apples” is not defined). We can only use elements on a given set. This is what we formally call a binary operation.

Definition 4.1 (Binary operation)

A binary operation on a set S is a function:

$$\begin{aligned} * : S \times S &\rightarrow S \\ *(a, b) &= a * b \end{aligned}$$

Binary operations

Examples

The following binary operations are all different:

$$
\begin{aligned}
&+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \\
&+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \\
&+ : \mathcal{M}_{m\times n}(\mathbb{R}) \times \mathcal{M}_{m\times n}(\mathbb{R}) \to \mathcal{M}_{m\times n}(\mathbb{R})
\end{aligned}
$$

The following is not a binary operation because it is not well defined

$$+ : \mathcal{M}(\mathbb{R}) \times \mathcal{M}(\mathbb{R}) \to \mathcal{M}(\mathbb{R})$$

we don't know how to add a 2×2 matrix with a 3×3 one.

Closed set

Definition 4.2

Let S be a set and H a subset of S. H is said to be ***closed with respect to the operation*** $*$ *defined in S iff*

$$\forall a, b \in H \quad a * b \in H$$

Then we may define the binary operation in H:

$$\begin{array}{rcl} * : H \times H & \to & H \\ *(a, b) & = & a * b \end{array}$$

which is called the ***binary operation induced*** *in H.*

Closed set

Example

Let $S = \mathbb{Z}$ and $H = \{n^2 | n \in \mathbb{Z}^+\} = \{1, 4, 9, 16, 25, 36, \ldots\}$. H is not closed with respect to addition. For example:

$$\begin{array}{c} 1 \in H \\ 4 \in H \end{array} \quad \text{but } 1 + 4 \notin H$$

Example

Let $S = \mathbb{Z}$ and $H = \{n^2 | n \in \mathbb{Z}^+\} = \{1, 4, 9, 16, 25, 36, \ldots\}$. H is closed with respect to multiplication. For example:

$$\begin{array}{c} n^2 \in H \\ m^2 \in H \end{array} \quad \text{and } n^2 \cdot m^2 = (nm)^2 \in H$$

Closed set

Example

Let S be the set of **real-valued functions with a single real argument** $S = \{\mathbb{R} \to \mathbb{R}\}$. Let us define the **addition** of functions as

$$\begin{aligned} + : (\mathbb{R} \to \mathbb{R}) \times (\mathbb{R} \to \mathbb{R}) &\to \mathbb{R} \to \mathbb{R} \\ (f+g)(x) &= f(x) + g(x) \end{aligned}$$

Similarly for the **multiplication** and **subtraction** of functions. Let us define the **composition** of functions as

$$\begin{aligned} \circ : (\mathbb{R} \to \mathbb{R}) \times (\mathbb{R} \to \mathbb{R}) &\to \mathbb{R} \to \mathbb{R} \\ (f \circ g)(x) &= f(g(x)) \end{aligned}$$

S is closed with respect to addition, subtraction, multiplication and composition.

Definition of a binary operation

Example

To define a binary operation either we give the full table (**intensional definition**) as in

$a * b$	$b = 0$	$b = 1$	$b = 2$
$a = 0$	0	1	2
$a = 1$	1	2	0
$a = 2$	2	0	1

or

$a \triangle b$	$b = 0$	$b = 1$	$b = 2$
$a = 0$	1	2	0
$a = 1$	1	1	2
$a = 2$	0	0	2

or we give a rule to compute it (**extensional definition**) as in

$$a * b = (a + b) \bmod 3$$

Properties of a binary operation

Definition 4.3 (Commutativity)

*A binary operation is **commutative** iff*

$$a * b = b * a$$

Example

$*$ is commutative because its definition table is symmetric with respect to the main diagonal, but $\triangle$ is not commutative.

Properties of a binary operation

Definition 4.4 (Associativity)

A binary operation is ***associative*** *iff*

$$(a * b) * c = a * (b * c)$$

Example

$\triangle$ is not associative because

$$(0 \triangle 0) \triangle 0 = 1 \triangle 0 = 1$$
$$0 \triangle (0 \triangle 0) = 0 \triangle 1 = 2$$

But $*$ is associative

$$(0 * 0) * 0 = 0 * 0 = 0$$
$$0 * (0 * 0) = 0 * 0 = 0$$

We would have to test all possible triples, but after a a little bit of work we could show that $*$ is associative.

Properties of a binary operation

Example

Function composition is associative although not commutative.

Proof

Function composition is not commutative

$$(f \circ g)(x) = f(g(x)) \neq g(f(x)) = (g \circ f)(x)$$

Function composition is associative

$$((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = f(g(h(x))) = f((g \circ h)(x)) = (f \circ (g \circ h))(x)$$

Properties of a binary operation

Example

A function may not be well defined. For instance,

$$\begin{array}{rcl} / : \mathbb{Q} \times \mathbb{Q}) & \rightarrow & \mathbb{Q} \\ a/b & = & \frac{a}{b} \end{array}$$

is not well defined for $b = 0 \in \mathbb{Q}$

Example

A function may not be closed in S. For instance,

$$\begin{array}{rcl} / : \mathbb{Z} \times \mathbb{Z}) & \rightarrow & \mathbb{Z} \\ a/b & = & \frac{a}{b} \end{array}$$

is not closed because $a = 1 \in \mathbb{Z}, b = 3 \in \mathbb{Z}$ but $\frac{1}{3} \notin \mathbb{Z}$.

Properties of a binary operation

Definition 4.5 (Existence of a neutral element)

A binary operation has a ***neutral element****, e, iff*

$$\forall a \in S \quad a * e = e * a = a$$

Example

0 is the neutral element of addition in $\mathbb{R}$ because

$$\forall r \in \mathbb{R} \quad r + 0 = 0 + r = r$$

1 is the neutral element of multiplication in $\mathbb{R}$ because

$$\forall r \in \mathbb{R} \quad r \cdot 1 = 1 \cdot r = r$$

Addition in $\mathbb{N}$ has no neutral element since $0 \notin \mathbb{N}$.

Properties of a binary operation

Definition 4.6 (Existence of an inverse element)

A binary operation has an ***inverse element*** *iff*

$$\forall a \in S \quad \exists b \in S | a * b = b * a = e$$

being e the neutral element of $*$.

Example

The inverse element of 2 with respect to addition in $\mathbb{R}$ is -2 because

$$2 + (-2) = (-2) + 2 = 0$$

The inverse element of 2 with respect to multiplication in $\mathbb{R}$ is $\frac{1}{2}$ because

$$2 \cdot \frac{1}{2} = \frac{1}{2} \cdot 2 = 1$$

Multiplication in $\mathbb{N}$ has no inverse element since $\forall n \in \mathbb{N} \quad \frac{1}{n} \notin \mathbb{N}$.

Outline

10 Abstract algebra

- Sets
- Relations and functions
- Partitions and equivalence relationships
- Binary operations
- **Groups and subgroups**
- Homomorphisms and isomorphisms
- Algebraic structures

Groups and subgroups

Introduction

Groups and subgroups are algebraic structures. They are the ones that allow solving equations like

$$x + x = a \Rightarrow x = \frac{a}{2}$$

and that the equation

$$x \cdot x = a$$

does not have a solution in $\mathbb{R}$ if $a < 0$.
We'll see that defining a group amounts to define the elements belonging to the group as well as the operations that can be used with them.

Groups

Definition 5.1 (Group)

Given a set S and a binary operation $*$ defined on S, the pair $(S, *)$ is a ***group*** *if G is closed under $*$ and*

G1. *$*$ is associative in S*

G2. *$*$ has a neutral element in S*

G3. *$*$ has an inverse element in S*

Definition 5.2 (Abelian group)

$(S, *)$ is an ***abelian group*** *if $(S, *)$ is a group and $*$ is commutative.*

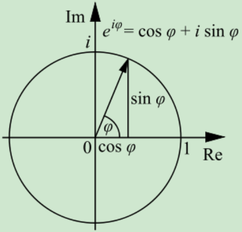
Definition 5.3 (Subgroup)

Let $(S, *)$ be a group. Let H be a subset of S, $H \subseteq S$, and $*_H$ be the $*$ induced operation in H. The pair $(H, *_H)$ is a subgroup of $(S, *)$ if it verifies the conditions to be a group.

Groups

Example

Consider $S = \{z \in \mathbb{C} | z = e^{i\varphi} \quad \forall \varphi \in \mathbb{R}\}$. $(U, \cdot)$ is a group.



Proof

G1. $\cdot$ is associative in S

$$z_1(z_2 z_3) = e^{i\varphi_1}(e^{i\varphi_2} e^{i\varphi_3}) = e^{i\varphi_1}(e^{i(\varphi_2+\varphi_3)}) = e^{i(\varphi_1+\varphi_2+\varphi_3)}$$
$$(z_1 z_2) z_3 = (e^{i\varphi_1} e^{i\varphi_2}) e^{i\varphi_3} = (e^{i\varphi_1+\varphi_2}) e^{i\varphi_3} = e^{i(\varphi_1+\varphi_2+\varphi_3)}$$

Example (continued)

Proof

G2. $\cdot$ has a neutral element in S

$1 = e^{i0} \in S$

$$z \cdot 1 = e^{i\varphi}e^{i0} = e^{i(\varphi+0)} = e^{i\varphi} = z$$
$$1 \cdot z = e^{i0}e^{i\varphi} = e^{i(0+\varphi)} = e^{i\varphi} = z$$

G3. $\cdot$ has an inverse element in S

For each $z = e^{i\varphi}$, its inverse element with respect to $\cdot$ is $z^{-1} = e^{-i\varphi}$

$$zz^{-1} = e^{i\varphi}e^{-i\varphi} = e^{i(\varphi-\varphi)} = e^{i0} = 1$$
$$z^{-1}z = e^{-i\varphi}e^{i\varphi} = e^{i(-\varphi+\varphi)} = e^{i0} = 1$$

Groups

Example

- $(\mathbb{N}, +)$ is not a group because it has no neutral element.
- $(\mathbb{N} \cup \{0\}, +)$ is not a group because it has no inverse element.
- $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$ and $(\mathbb{R}^n, +)$ are abelian groups.
- $(\mathcal{M}_{m\times n}, +)$ is an abelian group.
- $(\mathbb{R}, \cdot)$ is not a group because 0 has no inverse.
- $(\mathcal{M}_{n\times n}(\mathbb{R})), \cdot)$ is not a group because $\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$ has no inverse.
- Let $S \in \mathcal{M}_{n\times n}(\mathbb{R})$ be the set of invertible matrices of size $n \times n$. $(S, \cdot)$ is a group (although not abelian). It is called the General Linear Group of degree n $(GL(n, \mathbb{R}))$.

Example

The existence of groups is what allows us to solve equations. For instance, consider the equation

$$5 + x = 2$$

and its solution in the group $(\mathbb{Z}, +)$

$$\begin{array}{rcll} 5 + x & = & 2 & \text{[Addition of the inverse of 5 with respect to + in both} \\ -5 + (5 + x) & = & -5 + 2 & \text{[Addition is associative]} \\ (-5 + 5) + x & = & -3 & \text{[Definition of inverse]} \\ 0 + x & = & -3 & \text{[Definition of neutral element]} \\ x & = & -3 & \end{array}$$

Example

Consider the equation

$$2x = 3$$

and its solution in the group $(\mathbb{Q}, \cdot)$

$$\begin{array}{rcll}
2x & = & 3 & \text{[Multiplication by the inverse of 2 in both sides]} \\
\frac{1}{2}(2x) & = & \frac{1}{2}3 & \text{[Multiplication is associative]} \\
(\frac{1}{2}2)x & = & \frac{2}{3} & \text{[Definition of inverse]} \\
1x & = & \frac{2}{3} & \text{[Definition of neutral element]} \\
x & = & \frac{2}{3} &
\end{array}$$

Groups

Theorem 5.1 (Cancellation laws)

Given any group $(S, *)$, $\forall a, b, c \in S$ *it is verified*

- *Left cancellation:* $a * b = a * c \Rightarrow b = c$
- *Right cancellation:* $b * a = c * a \Rightarrow b = c$

Theorem 5.2 (Existance of a unique solution of linear equations)

Given any group $(S, *)$, $\forall a, b \in S$ *the linear equations*

$$a * x = b \text{ and } y * a = b$$

always have a unique solution in S.

Theorem 5.3 (Properties of the inverse)

Given any group $(S, *)$, $\forall a \in S$ *its inverse is unique and* $\forall a, b \in S$

$$(a * b)^{-1} = (b^{-1}) * (a^{-1})$$

Outline

Homomorphisms

Example

Consider the sets $S = \{a, b, c\}$ and $S' = \{A, B, C\}$ with the operations $* : S \times S \to S$ and $*' : S' \times S' \to S'$

$x * y$	$y = a$	$y = b$	$y = c$
$x = a$	a	b	c
$x = b$	b	c	a
$x = c$	c	a	b

and

$x *' y$	$y = A$	$y = B$	$y = C$
$x = A$	A	B	C
$x = B$	B	C	A
$x = C$	C	A	B

We may construct a function that "translates" elements in S into elements in S' with the "same properties".

$$\begin{array}{rcl} \phi : S & \to & S' \\ \phi(a) & = & A \\ \phi(b) & = & B \\ \phi(c) & = & C \end{array}$$

We note that

$$b * c = a \Rightarrow \phi(b) *' \phi(c) = \phi(a) \Rightarrow B *' C = A$$

Homomorphisms

Definition 6.1 (Group homomorphism)

Given two groups $(S, *)$ *and* $(S', *')$, *the function* $\phi : S \to S'$ *is a* ***group homomorphism*** *iff* $\forall a, b \in S$

$$\phi(a * b) = \phi(a) *' \phi(b)$$

Definition 6.2 (Group isomorphism)

Given two groups $(S, *)$ *and* $(S', *')$, *the function* $\phi : S \to S'$ *is a group isomorphism iff it is a group homomorphism and it is bijective.*

Homomorphisms

Example

Consider the two groups $(\mathbb{R}^n, +)$ and $(\mathbb{R}^m, +)$ and a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The application

$$\begin{array}{rcl} \phi : \mathbb{R}^n & \rightarrow & \mathbb{R}^m \\ \phi(\mathbf{x}) & = & A\mathbf{x} \end{array}$$

is a group homomorphism because

$$\phi(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \phi(\mathbf{u}) + \phi(\mathbf{v})$$

Example

Consider the two groups $(GL(n, \mathbb{R}), \cdot)$ and $(\mathbb{R}, \cdot)$. The application

$$\begin{array}{rcl} \phi : GL(n, \mathbb{R}) & \rightarrow & \mathbb{R} \\ \phi(A) & = & \det\{A\} \end{array}$$

is a group homomorphism because

$$\phi(AB) = \det\{AB\} = \det\{A\}\det\{B\} = \phi(A)\cdot\phi(B)$$

Homomorphisms

Theorem 6.1

Let $\phi : S \to S'$ be a group homomogrphism between two groups. Then,

- $\phi(e) = e'$
- $\phi(a^{-1}) = (\phi(a))^{-1}$

Definition 6.3 (Kernel of a group homomorphism)

Let $\phi : S \to S'$ be a group homomogrphism between two groups. Then, the kernel of ϕ is the set

$$\mathrm{Ker}\{\phi\} = \{x \in S | \phi(x) = e'\}$$

Example

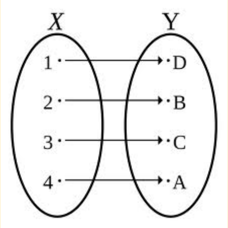
Let $\phi(\mathbf{x}) = A\mathbf{x}$. Then,

$$\mathrm{Ker}\{\phi\} = \{x \in \mathbb{R}^n | A\mathbf{x} = \mathbf{0}\} = \mathrm{Nul}\{A\}$$

Isomorphisms

Theorem 6.2 (Isomorphisms and cardinality)

If two groups $(S, *)$ and $(S', *')$ are isomorph (i.e., there exists an isomorphism between the two groups), then S and S' have the same cardinality.



Isomorphisms

Example

- $\mathbb{Q}$ and $\mathbb{R}$ cannot be isomorph because the cardinality of $\mathbb{Q}$ is $\aleph_0$ and the cardinality of $\mathbb{R}$ is $\aleph_1$.
- There are as many natural numbers as natural even numbers. In other words, the cardinality of $\mathbb{N}$ and $2\mathbb{N}$ are the same. The reason is that the function $\phi(n) = 2n$ is an isomorphism between $\mathbb{N}$ and $2\mathbb{N}$.

Example

Consider the set $\mathbb{R}_c = [0, c) \in \mathbb{R}$ and the operation $x +_c y = (x + y) \bmod c$. The pair $(\mathbb{R}_c, +_c)$ is a group. Consider now the two particular cases $(\mathbb{R}_{2\pi}, +_{2\pi})$ and $(\mathbb{R}_1, +_1)$ and the mapping

$$\begin{array}{rcl} \phi : \mathbb{R}_{2\pi} & \rightarrow & \mathbb{R}_1 \\ \phi(x) & = & \frac{x}{2\pi} \end{array}$$

ϕ is an isomorphism between $(\mathbb{R}_{2\pi}, +_{2\pi})$ and $(\mathbb{R}_1, +_1)$. In fact, all $(\mathbb{R}_c, +_c)$ groups are isomorph to any other $(\mathbb{R}_{c'}, +_{c'})$ group.

Isomorphisms

Cardinality is a *group property*. The nice things about isomorphisms is that they preserve group properties.

Theorem 6.3

If two groups $(S, *)$ and $(S', *')$ are isomorph, then

- *If $*$ is commutative, so is $*'$.*
- *If there is an order relation in S, it can be "translated" into an order relation in S'.*
- *If $\forall s \in S$ there exists a solution in S of the equation $x * x = s$, then $\forall s' \in S'$ there exists a solution in S' of the equation $x *' x = s'$.*
- *If $\forall a, b \in S$ there exists a solution in S of the equation $a * x = b$, then $\forall a', b' \in S'$ there exists a solution in S' of the equation $a' *' x = b'$.*
- *The kernel of any isomorphism ϕ between $(S, *)$ and $(S', *')$ is $\mathrm{Ker}\{\phi\} = \{e\}$ being e the neutral element of $*$ in S.*

Isomorphisms

Example

$((Z), +)$ is not isomorph to $((Q), +)$ because the equation

$$x + x = s$$

has a solution in $\mathbb{Q}$ for any $s \in \mathbb{Q}$ (that is $x = \frac{s}{2}$), but it does not have a solution in $\mathbb{Z}$ for any $s \in \mathbb{Z}$ (it only has a solution in $\mathbb{Z}$ if s is an even number).

Example

$((R), \cdot)$ is not isomorph to $((C), \cdot)$ because the equation

$$x \cdot x = z$$

has two solution in $\mathbb{C}$ for any $z \in \mathbb{C}$ (in fact there are two solutions, if $z = re^{i\theta}$, then $x = \pm re^{i\frac{\theta}{2}}$ are the two solutions) , but it does not have a solution in $\mathbb{R}$ for any $z \in \mathbb{R}$ (it only has a solution in $\mathbb{R}$ if z is a non-negative number).

Outline

Algebraic structures

Algebraic structures

Algebraic structures are tools that help us to define operate on numbers and elements within a set, solve equations, etc.

	Set ***S*** with binary operation **+**
	Operation **+** is associative
monoid	Existence of identity element of **+** in ***S***
group	Existence of inverse elements of **+** in ***S***
abelian group	Commutativity of **+**
	Associative binary operation **•**
pseudo-ring	Distributivity of **•** over **+**
ring	Existence of identity element of **•** in ***S***
commutative ring	Commutativity of **•**
field	Existence of inverse elements of **•** in ***S***

Algebraic structures

Definition 7.1 (Ring)

The tuple $(S, *, \circ)$ *is a* ***ring*** *iff*

R1. $(S, *)$ *is an abelian group.*

R2. $\circ$ *is associative.*

R3. $\circ$ *is distributive with respect to* $*$*, i.e.,* $\forall a, b, c \in S$

- *Left-distributive:* $a \circ (b * c) = (a \circ b) * (a \circ c)$
- *Right-distributive:* $(a * b) \circ c = (a \circ c) * (b \circ c)$

Example

- $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$ are rings.
- $(\mathcal{M}_{m \times n}(\mathbb{R}), +, \cdot)$ is a ring.
- $(\mathbb{R} \to \mathbb{R}, +, \cdot)$ is a ring.

Theorem 7.1 (Properties of rings)

Let $(S, *, \circ)$ be a ring and let e be the neutral element of $*$ in S. For any $a \in S$, let a' be the inverse of a with respect to the operation $*$. Then $\forall a, b \in S$

- $a \circ e = e \circ a = e$.
- $a \circ b' = a' \circ b = (a \circ b)'$
- $a' \circ b' = a \circ b$

Example

Consider the ring $(\mathbb{R}, +, \cdot)$. We are used to the properties $\forall a, b \in \mathbb{R}$

- $a \cdot 0 = 0 \cdot a = 0$.
- $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$
- $(-a) \cdot (-b) = a \cdot b$

But, as stated by the previous theorem, these are properties of all rings.

Definition 7.2 (Kinds of rings)

A ring $(S, *, \circ)$ *is*

- ***commutative*** *iff* $\circ$ *is commutative.*
- ***unitary*** *iff* $\circ$ *has a neutral element (referred as* 1*).*
- ***divisive*** *if it is unitary and*

$$\forall a \in S - \{e\} \quad \exists! a^{-1} \in S, | a \circ a^{-1} = a^{-1} \circ a = 1$$

That is each element has a multiplicative inverse.

Example

- $(\mathbb{P}, +, \cdot)$ the set of polynomials with coefficients from a ring is a ring.

Algebraic structures

Definition 7.3 (Field (*cuerpo*))

A divisive, commutative ring is called a ***field****.*

Example

- $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, and $(\mathbb{C}, +, \cdot)$ are fields.
- $(\mathbb{Z}, +, \cdot)$ is not a field because multiplication has not an inverse in $\mathbb{Z}$.

Definition 7.4 (Vector space over a field)

Consider a field $(\mathbb{K}, *, \circ)$. *A* ***vector space*** *over this field is a tuple* $(V, +, \cdot)$ *so that* V *is a set whose elements are called vectors, and* $+ : V \times V \to V$ *is a binary operation under which* V *is closed,* $\cdot : \mathbb{K} \times V \to V$ *is an operation between scalars in the field* ($\mathbb{K}$) *and vectors in the vector space* (V) *such that* $\forall a, b \in \mathbb{K}, \forall \mathbf{u}, \mathbf{v} \in V$

V1. $(V, +)$ *is an abelian group.*

V2. $(a \cdot \mathbf{u}) \in V$

V3. $a \cdot (b \cdot \mathbf{u}) = (a \circ b) \cdot \mathbf{u}$

V4. $(a * b) \cdot \mathbf{u} = a \cdot \mathbf{u} + b \cdot \mathbf{u}$

V5. $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$

V6. $1 \cdot \mathbf{u} = \mathbf{u}$

Algebraic structures

Examples

- $(\mathbb{R}^n, +, \cdot)$ and $(\mathbb{C}^n, +, \cdot)$.
- $(\mathcal{M}_{m\times n}(\mathbb{R}), +, \cdot)$: the set of matrices of a given size with coefficients in a field.
- $(\mathbb{P}, +, \cdot)$: the set of polynomials with coefficients in a field.
- $(\{X \to V\}, +, \cdot)$: the set of all functions from an arbitrary set X onto an arbitrary vector space V.
- The set of all continuous functions is a vector space.
- The set of all linear maps between two vector spaces is also a vector space.
- The set of all infinite sequences of values from a field is also a vector space.

Algebraic structures

Definition 7.5 (Algebra)

Consider a vector space $(V, +, \cdot)$ *over a field* $(\mathbb{K}, *, \circ)$ *and a binary operation* $\bullet : V \times V \to V$. $(V, +, \cdot, \bullet)$ *is an **algebra** iff* $\forall a, b \in \mathbb{K}, \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$

A1. Left distributivity: $(\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} = \mathbf{u} \bullet \mathbf{w} + \mathbf{v} \bullet \mathbf{w}$

A2. Right distributivity: $\mathbf{u} \bullet (\mathbf{v} + \mathbf{w}) = \mathbf{u} \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{w}$

A3. Compatibility with scalars: $(a \cdot \mathbf{u}) \bullet (b \cdot \mathbf{v}) = (a \circ b) \cdot (\mathbf{u} \bullet \mathbf{v})$

Examples

- Real numbers ($\mathbb{R}$) are an algebra ("1D").
- Complex numbers ($\mathbb{C}$) are an algebra ("2D").
- Quaternions are an algebra ("4D").

Outline