MAST10022 Linear Algebra: Advanced

Lawrence Reeves
School of Mathematics and Statistics
University of Melbourne

2024 semester 1

Reading mathematics is not like reading novels or history. You need to think slowly about every sentence. Usually, you will need to reread the same material later, often more than one rereading.

From "The art of proof" by Beck and Geoghegan

You have probably never had a laboratory course in mathematics. Mathematics is not considered to be an experimental science, whereas physics, chemistry, and biology are. Research for a chemist can consist of a laboratory experiment designed to validate a conjecture, to suggest a conjecture, or simply to see what happens. There is little comparable activity in mathematics.

The main business of mathematics is proving theorems.

From "A first course in abstract algebra" by J. B. Fraleigh

Lecture plan

LECTURE 1

Sets and functions

Before beginning with the linear algebra content proper we revise some important general concepts and notations. Sets and functions are fundamental to linear algebra and to modern mathematics in general.

1.1 Sets

A **set** is a collection of objects called **elements** (or **members**) of that set.[*] The notation $x \in A$ means that x is an element of the set A. The notation $x \notin A$ is used to mean that x is not a member of A.

Let A and B be sets. We say that A is **a subset of** (or **is contained in** B), written $A \subseteq B$, if every element of A is also an element of B (i.e., if $x \in A$, then $x \in B$). Two sets are **equal** if they have the same members. Thus $A = B$ exactly when both $A \subseteq B$ and $B \subseteq A$. If $A \subseteq B$ and $A \neq B$ then we say that A is a **proper subset** of B and (sometimes) write $A \subsetneq B$.

Sets are often defined either by listing their elements, as in $A = \{0, 2, 3\}$, or by giving a rule or condition which determines membership in the set, as in $A = \{x \in \mathbb{R} \mid x^3 - 5x^2 + 6x = 0\}$.

Here are some familiar (mostly mathematical) sets:

- ▷ natural numbers: $\mathbb{N} = \{1, 2, 3, 4, \dots\}$
- ▷ integers: $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- ▷ rational numbers: $\mathbb{Q} = \{\ \frac{x}{y} \mid x, y \in \mathbb{Z}, y \neq 0\}$
- ▷ real numbers:[†] $\mathbb{R}$
- ▷ complex numbers: $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$
- ▷ $(1, 3] = \{x \in \mathbb{R} \mid 1 < x \leqslant 3\}$
- ▷ Greek alphabet (lower case): $\{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta, \iota, \kappa, \lambda, \mu, \nu, \xi, o, \pi, \rho, \sigma, \tau, \upsilon, \varphi, \chi, \psi, \omega\}$

In these examples we have the following containment relations: $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ and $(1, 3] \subseteq \mathbb{R}$. Note that $(1, 3] \not\subseteq \mathbb{Q}$ because the interval $(1, 3]$ contains real numbers that are not rational. For example, $\sqrt{2} \in (1, 3]$ but $\sqrt{2} \notin \mathbb{Q}$.

As indicated above, the notation $\{\dots\}$ is used for set formation. Sets are themselves mathematical objects and so can be members of other sets. For instance, the set $\{3, 5\}$ consists of two elements, namely the numbers 3 and 5. The set $\{\{3, 5\}, \{3, 7\}, \{7\}, 3\}$ consists of 4 elements, namely the sets $\{3, 5\}$,$\{3, 7\}$, $\{7\}$ and the integer 3. Note that $7 \notin \{\{3, 5\}, \{3, 7\}, \{7\}, 3\}$. Observe that $\{7\}$ is the set whose only element is the number 7, and we have that $7 \in \{7\}$ but $7 \not\subseteq \{7\}$.

The **empty set**, denoted by $\emptyset$, is the set that has no elements, that is, $x \in \emptyset$ is never true.

[*] In fact, more care is needed in the definition of a set. In general one must place some restriction on set formation. For example, trying to form $\{x \mid x \text{ is a set}\}$ or $\{x \mid x \notin x\}$ can lead to logical paradoxes (Russell's paradox). This can be dealt with or excluded in a more formal or axiomatic treatment of set theory. We will be careful to avoid situations where this difficulty arises.

[†] It's a bit more involved to define the real numbers precisely, but one can think of them either as the points on the real line or as (infinite) decimal expansions. In this subject we will be using some standard properties of $\mathbb{R}$, but we will not give a construction.

Lemma 1.1

The empty set is a subset of every set.

Proof. Let A be a set. We need to show that the following statement is true:

$$\text{if } a \in \emptyset \text{, then } a \in A \tag{$*$}$$

Let's suppose that ($*$) were not true. Then there would be an element a such that $a \in \emptyset$ is true and $a \in A$ is false. However, since $a \in \emptyset$ is never true, no such a exists and we conclude that ($*$) must in fact be true. □

Note that $\emptyset \in \{\emptyset\}$ and $\emptyset \subseteq \{\emptyset\}$ but $\emptyset \notin \emptyset$.

Operations on sets

The **intersection** of two sets A and B is the set

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

The **union** of A and B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

Example 1.2.

1) $\{2m+5 \mid m \in \mathbb{Z}\} \cup \{2m \mid m \in \mathbb{Z}\} = \mathbb{Z}$

2) $\{2m+5 \mid m \in \mathbb{Z}\} \cap \{2m \mid m \in \mathbb{Z}\} = \emptyset$

3) $\{n \in \mathbb{N} \mid n \text{ is prime}\} \cap \{2n \mid n \in \mathbb{N}\} = \{2\}$

The **set difference** of two sets A and B is the set

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$$

If $B \subseteq A$, then $A \setminus B$ is called the **complement** of B in A. If the larger set A is clear from the context, we sometimes write B^c for the complement of B in A.

Example 1.3.

1) $\mathbb{Z} \setminus \mathbb{N} = \{\dots, -2, -1, 0\}$

2) $\mathbb{N} \setminus \mathbb{Z} = \emptyset$

3) $[0,2] \setminus \mathbb{N} = [0,1) \cup (1,2)$

4) $\mathbb{N} \setminus [0,2] = \{3, 4, \dots\}$

Proposition 1.4: De Morgan's Laws

Let $A, B \subseteq X$ be two sets. Then

1. $A \subseteq B$ iff $B^c \subseteq A^c$
2. $(A \cap B)^c = A^c \cup B^c$
3. $(A \cup B)^c = A^c \cap B^c$

Given a set A, the **power set** of A is the set containing all subsets of A. It is denoted $\mathcal{P}(A)$.

Example 1.5. For $A = \{\alpha, \beta, \gamma\}$ we have $\mathcal{P}(A) = \{\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\alpha, \beta\}, \{\alpha, \gamma\}, \{\beta, \gamma\}, \{\alpha, \beta, \gamma\}\}$.

Let A and B be two sets. We define a set, called the **Cartesian product** of A and B, by

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

Each element (a, b) of the set $A \times B$ is called an **ordered pair**.

Note.

1. $(a, b) = (a', b')$ precisely when $a = a'$ and $b = b'$.
2. If $A \neq B$, then $A \times B \neq B \times A$.
3. If $A = B$, we often write A^2 in place of $A \times A$.

1.2 Functions

The concept of a function is fundamental in mathematics. Functions on the real numbers are often described using some sort of a formula (e.g., $f(x) = \sin(x)$), but we want to define the notion of function in a way that makes sense more generally. The idea is to make a definition out of what is sometimes called the graph of a function.

Definition 1.6: function

Let A and B be sets. A **function from A to B** is a subset f of $A \times B$ such that for each $a \in A$ there is exactly one element of f whose first entry is a. We write $f(a) = b$ to mean $(a, b) \in f$. We write $f : A \to B$ to mean that f is a function from A to B. The set A is called the **domain** of the function and B is called the **codomain** of the function.

Remark.

1. Functions are often (but not always!) given by a formula such as $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. When written in this way, the subset of $A \times B$ is understood to be $\{(a, f(a)) \mid a \in A\}$.
2. The domain and codomain are part of the defining data of a function. The following two functions are *not* the same:

$$\begin{aligned} f : \mathbb{R} \to \mathbb{R}, &\quad f(x) = x^2 \\ g : [0, \infty) \to \mathbb{R}, &\quad g(x) = x^2 \end{aligned}$$

Definition 1.7: injective, surjective, bijective

Let $f : A \to B$ be a function.

1. We say that f is **injective** if for all $a_1, a_2 \in A$, if $f(a_1) = f(a_2)$ then $a_1 = a_2$.
2. We say that f is **surjective** if for all $b \in B$ there exists $a \in A$ with $f(a) = b$.
3. We say that f is **bijective** if it is both injective and surjective.

Example 1.8.

1) The function $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ is neither injective nor surjective.

2) The function $g : [0, \infty) \to \mathbb{R}, g(x) = x^2$ is injective but not surjective.

3) The function $h : \mathbb{R} \to [0, \infty), h(x) = x^2$ is surjective but not injective.

4) The function $k : [0, \infty) \to [0, \infty), k(x) = x^2$ is bijective.

Example 1.9.

1) $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}, f(m,n) = 2^m 3^n$ is injective (but not surjective).

2) $g : \mathbb{N} \to \mathbb{Z}, g(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ \frac{1-n}{2} & \text{if } n \text{ is odd} \end{cases}$ is bijective.

Let $f : A \to B$ and $g : B \to C$ be two functions. The **composition** of f and g is the function $g \circ f : A \to C$ given by $g \circ f(a) = g(f(a))$. Given a set A, the **identity function** on A is the function $\mathrm{Id}_A : A \to A$, $\mathrm{Id}_A(a) = a$. If $f : A \to B$ is a bijection, then there is a well-defined **inverse function** $f^{-1} : B \to A$ having the property that $f \circ f^{-1} = \mathrm{Id}_B$ and $f^{-1} \circ f = \mathrm{Id}_A$. Indeed, if we think of functions as sets of ordered pairs and f is a bijection, then the ordered pairs of f^{-1} are just the pairs of f in reverse order.

Example 1.10. Consider the function $f : \mathbb{N} \to \mathbb{Z}_{\geqslant 2}, f(n) = n+1$. The corresponding subset of $\mathbb{N} \times \mathbb{Z}_{\geqslant 2}$ is

$$\{(1,2),(2,3),(3,4),\dots\}$$

The function f is a bijection. Its inverse is $f^{-1} : \mathbb{Z}_{\geqslant 2} \to \mathbb{N}, f^{-1}(n) = n-1$ which as a subset of $\mathbb{Z}_{\geqslant 2} \times \mathbb{N}$ is

$$\{(2,1),(3,2),(4,3),\dots\}$$

In mathematics one often needs functions of several variables, for example the operation of addition of real numbers is a function of two variables which assigns to each pair of real numbers (x,y) their sum $x+y$. Thus addition is a function from $\mathbb{R}^2$ to $\mathbb{R}$. More generally, a **function of n variables** from A to B (or an n-ary function f from A to B) is just a function of the form $f : A^n \to B$.

1.3 Exercises

1. List five elements belonging to each of the following sets:

   (a) $\{n \in \mathbb{N} \mid n \text{ is divisible by } 5\}$

   (b) $\mathcal{P}(\{1,2,3,4,5\})$

   (c) $\{n \in \mathbb{N} \mid n+1 \text{ is a prime }\}$

   (d) $\{2^n \mid n \in \mathbb{N}\}$

   (e) $\{r \in \mathbb{Q} \mid 0 < r < 1\}$

2. List all the elements in each of the following sets:

   (a) $\{n \in \mathbb{N} \mid n^2 = 3\}$

   (b) $\{n \in \mathbb{Z} \mid 3 < |n| < 7\}$

   (c) $\{x \in \mathbb{R} \mid x < 1 \text{ and } x \geqslant 2\}$

   (d) $\{3n+1 \mid n \in \mathbb{N} \text{ and } n \leqslant 6\}$

   (e) $\{n \in \mathbb{N} \mid n \text{ is a prime and } n \leqslant 15\}$

3. Consider the sets

   $A = \{n \in \mathbb{N} \mid n \text{ is odd }\}$

   $B = \{n \in \mathbb{N} \mid n \text{ is a prime }\}$

   $C = \{4n+3 \mid n \in \mathbb{N}\}$

   $D = \{x \in \mathbb{R} \mid x^2 - 8x + 15 = 0\}$

   Which are subsets of which? Consider all sixteen possibilities.

4. Consider the sets

   $A = \{n \in \mathbb{N} \mid n \leq 11\}$

   $B = \{n \in \mathbb{N} \mid n \text{ is even and } n \leq 20\}$

   $E = \{n \in \mathbb{N} \mid n \text{ is even}\}$

   Determine each of the following sets:

(a) $A \cup B$
(b) $A \cap B$
(c) $A \setminus B$
(d) $B \setminus A$
(e) $E \cap B$
(f) $B \setminus E$
(g) $E \setminus B$
(h) $\mathbb{N} \setminus E$

5. Prove (directly from the definitions of the operations) that $(A \cup B) \cap A^c \subseteq B$.

6. Prove or disprove each of the following:

   (a) $A \cap B = A \cap C$ implies $B = C$
   (b) $A \cup B = A \cup C$ implies $B = C$
   (c) ($A \cap B = A \cap C$ and $A \cup B = A \cup C$) implies $B = C$

7. Let $S = \{0, 1, 2, 3, 4\}$ and $T = \{0, 2, 4\}$.

   (a) How many elements are there in $S \times T$? How many in $T \times S$?
   (b) List the elements in $\{(m, n) \in S \times T \mid m < n\}$.
   (c) List the elements in $\{(m, n) \in T \times S \mid m < n\}$.
   (d) List the elements in $\{(m, n) \in S \times T \mid m + n \geq 3\}$.
   (e) List the elements in $\{(m, n) \in T \times S \mid mn \geq 4\}$.
   (f) List the elements in $\{(m, n) \in S \times S \mid m + n = 10\}$.

8. Let $S = \{1, 2, 3, 4, 5\}$ and $T = \{a, b, c, d\}$. For each question below: if the answer is "yes" give an example; if the answer is "no" explain briefly.

   (a) Are there any injective functions from S to T?
   (b) Are there any injective functions from T to S?
   (c) Are there any surjective functions from S to T?
   (d) Are there any surjective functions from T to S?
   (e) Are there any bijective functions from S and T?

9. Let $S = \{1, 2, 3, 4, 5\}$ and consider the following functions from S to S: $1_S(n) = n$, $f(n) = 6 - n$, $g(n) = max\{3, n\}$ and $h(n) = max\{1, n - 1\}$.

   (a) Write each of these functions as a set of ordered pairs.
   (b) Which of these functions are injective and which are surjective?

10. Consider the two functions from $\mathbb{N}^2$ to $\mathbb{N}$ defined by $f(m, n) = 2^m 3^n$ and $g(m, n) = 2^m 4^n$. Show that f is injective but that g is not injective. Is f surjective? Explain. (You may use that every $n \in \mathbb{N}$ with $n \geqslant 2$ has a unique prime factorisation.)

11. Show that if $f : A \to B$ and $g : B \to C$ are injective functions, then $g \circ f$ is injective.

12. Show that composition of functions is associative, that is, $h \circ (g \circ f) = (h \circ g) \circ f$.

13. Here are two 'shift functions' mapping $\mathbb{N}$ to $\mathbb{N}$:

$$\begin{aligned} R &: \mathbb{N} \to \mathbb{N}, \quad R(n) = n + 1 \\ L &: \mathbb{N} \to \mathbb{N}, \quad L(n) = \max\{1, n - 1\} \end{aligned}$$

   (a) Show that R is injective but not surjective.
   (b) Show that L is surjective but not injective.
   (c) Show that $L \circ R = \mathrm{Id}_{\mathbb{N}}$ but that $R \circ L \neq \mathrm{Id}_{\mathbb{N}}$.

Further reading for lecture 1

The extra material at the end of a lecture can include extra theory or references. It is NOT required! It's just for those who would like to know more.

▷ Some references for introductory set theory:

***The art of proof: basic training for deeper mathematics*,** by Beck and Geoghegan, chapter 5.

***Naive set theory*,** by Halmos.

***Russell's paradox*,** on Wikipedia.

***Zermelo-Fraenkel set theory*,** on Wikipedia.

▷ The axiomatic definition of the real numbers $\mathbb{R}$ and their construction from $\mathbb{Q}$ will be covered in the subject *MAST20033 Real Analysis: Advanced* (amongst others). An important difference between $\mathbb{Q}$ and $\mathbb{R}$ is that every bounded subset of $\mathbb{R}$ has a least upper bound. A standard construction of $\mathbb{R}$ from $\mathbb{Q}$ is via "Dedekind cuts" which, roughly speaking, carefully adds least upper bounds to $\mathbb{Q}$.

***The Art of Proof*,** by Beck and Geoghegan, chapter 8.

***Principles of Mathematical Analysis*,** by Rudin, chapter 1.

▷ Bijections are used in the definition of the **cardinality** (or size) of a set. Two sets are said to have the same cardinality if there exists a bijection from one to the other. The two sets $\{1, 2, 3\}$ and {Julia, Ada, Xav} have the same cardinality. It starts to get interesting when we consider infinite sets. Not all infinite sets have the same cardinality. The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ all have the same cardinality, but $\mathbb{R}$ does not. In particular, there is a bijection from $\mathbb{N}$ to $\mathbb{Q}$. That there is no bijection from $\mathbb{N}$ to $\mathbb{R}$ can be shown with an elegant argument known as 'Cantor diagonalisation'.

***The Art of Proof*,** by Beck and Geoghegan, chapter 13.

▷ Theorem, proposition, lemma, corollary,... What's the difference?

These are all statements of results that are then proven to be true. The difference is a little subjective, and the choice usually reflects the author's view of how important or interesting the result is. Theorems are results that are considered important. Propositions are less important but still interesting. Lemmas are usually shorter often technical results that are used in proving other statements. A corollary is a statement that follows easily from a previously proven theorem or proposition.

LECTURE **2**

Mathematical induction and logic

We continue with some background material on logic and proof by induction that we will need later when constructing proofs.

2.1 Some useful notation from logic

2.1.1 Propositions

We will be concerned with statements that are either true or false. They are called **propositions** (alternatively **statements**).

Example 2.1.

Propositions:

▷ $1 + 1 = 2$

▷ $1 + 1 = 3$

▷ For all integers $z \in \mathbb{Z}$, if z^2 is even then z is even.

▷ All maths lecturers are named Lawrence.

▷ Every even integer greater than 2 can be written as the sum of two primes.

Not propositions:

▷ 2^8

▷ z is even

▷ 'potato'

2.1.2 Operations on propositions (connectives)

Given two propositions p and q, we can combine them to form new propositions. The **conjunction** ('and') of p and q is denoted $p \wedge q$ and is defined to be true if both p and q are true, and false in all other cases. The **disjunction** ('or') of p and q is denoted $p \vee q$ and is defined to be false if both p and q are false, and true in all other cases. For each, the four possible cases can be listed in a **truth table**.

conjunction		
p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

disjunction		
p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

implication		
p	q	$p \implies q$
T	T	T
T	F	F
F	T	T
F	F	T

equivalence		
p	q	$p \iff q$
T	T	T
T	F	F
F	T	F
F	F	T

A statement of the form "if p then q" is called an **implication** (or a **conditional** statement). It is written as $p \implies q$. It is defined by the truth table above. Notice that if p is false, then $p \implies q$ is true whatever the truth value of q. The proof of Lemma 1.1 illustrates why this is the correct choice of definition to make.

A statement of the form "p if and only if q" is called an **equivalence**. It is written as $p \iff q$ and is defined by the truth table above.

Given a single proposition p, its **negation**, denoted $\neg p$ (or $\sim p$) is the statement that is false if p is true and is true if p is false.

Example 2.2. Let's construct a truth table for the proposition $\neg p \vee q$. Compare the last column with that for $p \implies q$.

p	q	$\neg p$	$\neg p \vee q$
T	T	F	T
T	F	F	F
F	T	T	T
F	F	T	T

We say that two statements p and q are **logically equivalent** if p is true precisely when q is true.[*] This is written $p \equiv q$.

Example 2.3. We observed in the previous example that $(p \implies q) \equiv (\neg p \vee q)$. To show this explicitly, we construct a truth table for $(p \implies q) \iff (\neg p \vee q)$ and observe that the equivalence has value T for all possible values of p and q.

p	q	$p \implies q$	$\neg p \vee q$	$(p \implies q) \iff (\neg p \vee q)$
T	T	T	T	T
T	F	F	F	T
F	T	T	T	T
F	F	T	T	T

Exercise 14. Use a truth table to show that $(p \implies q) \equiv (\neg q \implies \neg p)$. The second statement is called the **contrapositive** of the first statement.

The following can be established using truth tables.

Lemma 2.4: De Morgan

Let p and q be two statements. Then

1. $\neg(p \vee q) \equiv (\neg p) \wedge (\neg q)$
2. $\neg(p \wedge q) \equiv (\neg p) \vee (\neg q)$

2.1.3 Quantifiers

The symbol $\forall$ means 'for all' (or 'for each' or 'for every'). It is called the **universal quantifier**. The general form of a proposition formed using the universal quantifier is

$$\forall x \in A, p(x)$$

where, for a given x, $p(x)$ is a statement.

The symbol $\exists$ means 'there exist' (or 'for some'). It is called the **existential quantifier**. The statement

$$\exists\, x \in A, p(x)$$

is true if there is at least one element x in A such that the statement $p(x)$ is true.

Example 2.5. Here are some statements constructed using these quantifiers.

1. $\forall\, x \in \mathbb{R}, x^2 \geqslant 0$ (which is true)
2. $\forall\, x \in \mathbb{R}, (x^2 \in \mathbb{Q} \implies x \in \mathbb{Q})$ (which is false)
3. $\exists\, x \in \mathbb{R}, x^2 < 0$ (which is false)
4. $\forall\, x \in \mathbb{R}\ \exists\, y \in \mathbb{R}, x + y = 0$ (which is true)

[*] This is the same as saying that the statement $p \iff q$ is always true (i.e., is a *tautology*).

5. $\exists\, y \in \mathbb{R}\, \forall\, x \in \mathbb{R},\, x + y = 0$ (which is false)

Notice the difference between the final two examples above. Example 4 says that every real number has an additive inverse. For example, given $x = \pi$ we can take $y = (-1) \times \pi$. Example 5 says that there exists one real number that is the additive inverse of *every* real number.

Lemma 2.6: Negation of statements involving quantifiers

1. $\neg(\forall x \in A, p(x)) \equiv \exists x \in A, \neg p(x)$
2. $\neg(\exists x \in A, p(x)) \equiv \forall x \in A, \neg p(x)$

Example 2.7.

statement	negation
$\forall\, x \in \mathbb{R},\, x^2 \geqslant 0$	$\exists\, x \in \mathbb{R},\, x^2 < 0$
$\exists\, y \in \mathbb{R}\, \forall\, x \in \mathbb{R},\, x + y = 0$	$\forall\, y \in \mathbb{R}\, \exists\, x \in \mathbb{R},\, x + y \neq 0$
$\forall\, x \in U,\ x \in \emptyset \implies x \in A$	$\exists\, x \in U, (x \in \emptyset \wedge x \notin A)$
All maths lecturers are named Lawrence	There exists a maths lecturer whose name is not Lawrence
All flying pigs are purple	There exists a flying pig that is not purple

Which one of the two statements in the final row is true?

2.2 Induction

In this subject we will be assuming some standard properties of $\mathbb{N}$ such as the distributive law. An important property that we will use in many proofs is given in the following.

Theorem 2.8: Principle of mathematical induction

Let $P(n)$ be a (true or false) statement that depends on a natural number $n \in \mathbb{N}$. In order to prove that $P(n)$ is true for all values of $n \in \mathbb{N}$ it is sufficient to prove the following:

1) $P(1)$ is true ('base case')

2) $\forall n \in \mathbb{N}, P(n) \implies P(n+1)$ ('induction step')

To give a proof of this theorem we would need to consider the definition and construction of the natural numbers. It is equivalent to the so-called 'well ordering' property of $\mathbb{N}$. We will not give a proof, but shall regard the above as a fundamental property of $\mathbb{N}$ (and $\mathbb{Z}$).

Here are some examples of using mathematical induction as a method of proof.

Example 2.9.

1) **Claim.** For all $n \in \mathbb{N}$, the number $n^4 - 6n^3 + 11n^2 - 6n$ is divisible by 4.

Proof. Let $P(n)$ be the statement '$n^4 - 6n^3 + 11n^2 - 6n$ is divisible by 4'.† We check that both conditions of the above theorem are satisfied.

†Notice that $P(n)$ is a statement that is either true or false. It is not a polynomial, nor is it an integer. It would be an error to write something such as $P(n) = n^4 - 6n^3 + 11n^2 - 6n$

Base case: $P(1)$ is the statement '$1-6+11-6$ is divisible by 4'. Since $1-6+11-6=0$ and 0 is divisible by 4,[‡] the statement $P(1)$ is true.

Induction step: Let $n \in \mathbb{N}$ and suppose that $P(n)$ is true. That $P(n)$ is true means that there exists a $k \in \mathbb{Z}$ such that $n^4-6n^3+11n^2-6n=4k$. To show that $P(n+1)$ is true we need to show that $(n+1)^4-6(n+1)^3+11(n+1)^2-6(n+1)$ is divisible by 4. Note that

$$\begin{aligned}(n+1)^4-6(n+1)^3+11(n+1)^2-6(n+1) &= (n^4+4n^3+6n^2+4n+1)-6(n^3+3n^2+3n+1)\\ &\quad +11(n^2+2n+1)-6(n+1)\\ &= n^4-2n^3-n^2+2n\\ &= (n^4-6n^3+11n^2-6n)+4n^3-12n^2+8n\\ &= 4k+4(n^3-3n^2+2n)\\ &= 4(k+n^3-3n^2+2n)\end{aligned}$$

Therefore $P(n+1)$ is true. It follows from the principle of mathematical induction (Theorem 2.8) that $P(n)$ is true for all $n \in \mathbb{N}$. □

2) **Claim.** For all $n \geqslant 4$ we have $n! > 2^n$

Proof. Notice that the claim is that the inequality holds for all $n \geqslant 4$. In order to apply Theorem 2.8 in exactly the form given, we define $P(n)$ to be the statement that $(n+3)! > 2^{n+3}$. If we show that $P(n)$ is true for all $n \in \mathbb{N}$, we will have established the claim.

Base case: $P(1)$ is the statement that $4! > 2^4$. Since $4! = 24 > 16 = 2^4$, the statement $P(1)$ is true.

Induction step: Let $n \in \mathbb{N}$ and suppose that $P(n)$ is true, that is that $(n+3)! > 2^{n+3}$. We need to show that $P(n+1)$ is true, that is that $((n+1)+3)! > 2^{(n+1)+3}$. We have

$$\begin{aligned}((n+1)+3)! = (n+4)! &= (n+4)(n+3)!\\ &> (n+4)2^{n+3} && \text{(since } (n+3)! > 2^{n+3}\text{)}\\ &> 2 \times 2^{n+3} && \text{(since } n+4>2\text{)}\\ &= 2^{n+4} = 2^{(n+1)+3}\end{aligned}$$

Therefore $P(n+1)$ is true. It follows from the principle of mathematical induction (Theorem 2.8) that $P(n)$ is true for all $n \in \mathbb{N}$.

□

Here is another version of the induction statement in which the induction step is, in principle, easier to prove.

Theorem 2.10: 'complete induction'

Let $P(n)$ be a (true or false) statement that depends on a natural number $n \in \mathbb{N}$. In order to prove that $P(n)$ is true for all values of $n \in \mathbb{N}$ it is sufficient to prove the following:

1) $P(1)$ is true ('base case')
2) $\forall n \in \mathbb{N}$, $(P(1) \wedge \cdots \wedge P(n)) \implies P(n+1)$ ('induction step')
   (i.e., $\forall n \in \mathbb{N}$, if $P(k)$ is true for all $k \leqslant n$, then $P(n+1)$ is true)

Proof. We will show that this theorem follows from Theorem 2.8.[§]

[‡] Every integer $m \in \mathbb{Z}$ divides 0 since $0 = m \times 0$

[§] The converse is also true: Theorem 2.10 implies Theorem 2.8.

Let $P(n)$ be as in the statement of the current theorem and assume that both (1) and (2) hold. We need to show that $P(n)$ is true for all $n \in \mathbb{N}$. Let $Q(n)$ be the statement that 'P(k) is true for all $k \leqslant n$'. We want to check that the statements $Q(n)$ satisfy the conditions stated in Theorem 2.8.

Base case: Since $Q(1)$ is simply the statement that $P(1)$ is true, and were are assuming that (1) holds, we have that $Q(1)$ is true.

Induction step: We need to show that if $Q(n)$ is true, then $Q(n+1)$ is true. Assume then that $Q(n)$ is true. That is, that $P(1)$ is true, and $P(2)$ is true, and $P(3)$ is true,..., and $P(n)$ is true. In particular, we have that $P(n)$ is true. Therefore, since we are assuming that (2) (from the current theorem) holds, we have that $P(n+1)$ is true. Therefore, $P(1)$ is true, and $P(2)$ is true, and $P(3)$ is true,..., and $P(n)$ is true, and $P(n+1)$ is true. That is, $Q(n+1)$ is true.

It follows from the principle of mathematical induction (Theorem 2.8) that $Q(n)$ is true for all $n \in \mathbb{N}$. Therefore, $P(n)$ is true for all $n \in \mathbb{N}$. □

Example 2.11. Complete induction can be used to prove that every natural number $n \in \mathbb{N}_{\geqslant 2}$ can be written as a product of primes.

2.3 Exercises

15. Translate the following into mathematical notation.
    (a) The square of 10 is 50 and the cube of 5 is 12.
    (b) If 7 is an integer, then 6 is not an integer.

16. Construct truth tables for the following statements.
    (a) $(p \wedge q) \vee (\neg p \wedge \neg q)$
    (b) $(\neg q \wedge (p \implies q)) \implies \neg p$

17. Use a truth table to show that $(p \iff q)$ is logically equivalent to $(\neg p \iff \neg q)$.

18. Translate the following into mathematical notation.
    (a) All rational numbers are larger than 6.
    (b) There is a real-number solution to $x^2 + 3x - 7 = 0$.
    (c) There is a natural number whose cube is 8.
    (d) The set of all integers that aren't multiples of 7.

19. Find the negation of the following propositions.
    (a) $\forall x \in \mathbb{R}, \;\; x^2 = 10$
    (b) $\exists y \in \mathbb{N}, \;\; y < 0$
    (c) $\exists a \in \mathbb{N}, \;\; \forall x \in \mathbb{R}, \;\; ax = 4$
    (d) $\forall y \in \mathbb{Q}, \;\; \exists x \in \mathbb{R}, \;\; xy = 30$

20. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Use induction to prove that $A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$ for all $n \in \mathbb{N}$.

21. Use induction to show that $k^4 - 6k^3 + 11k^2 - 6k$ is divisible by 4 for all $k \in \mathbb{N}$.

Further reading for lecture 2 (for interest)

▷ More on induction and well-ordering of $\mathbb{N}$

The Art of Proof by Beck and Geoghegan, chapter 2.

▷ Introductory logic

The Art of Proof by Beck and Geoghegan, chapter 3.

▷ Here is a useful minor reformulation of Theorem 2.8. The difference is that we start at any integer as the base case (in place of 1).

Theorem. *Let $P(n)$ be a (true or false) statement that depends on an integer $n \in \mathbb{Z}$ and let $n_0 \in \mathbb{Z}$. In order to prove that $P(n)$ is true for all values of $n \geqslant n_0$ it is sufficient to prove the following:*

1) $P(n_0)$ is true *('base case')*

2) $\forall n \in \mathbb{Z}_{\geqslant n_0}, P(n) \implies P(n+1)$ *('induction step')*

Proof. We proof that this theorem is implied by Theorem 2.8.

Let $P(n)$ and n_0 be as in the statement of the current theorem and suppose that (1) and (2) both hold. For any $n \in \mathbb{N}$, define $Q(n)$ to be the statement $P(n_0 + n - 1)$. Note that $n_0 + n - 1 \geqslant n_0$, since $n \geqslant 1$. We will use Theorem 2.8 to show that $Q(n)$ holds for all $n \in \mathbb{N}$.

Base case: $Q(1)$ is the statement $P(n_0)$, which is true by condition (1).

Induction step: Suppose that $Q(n)$ is true for some $n \in \mathbb{N}$. Then $P(n + n_0 - 1)$ is true, and therefore $P(n + n_0)$ is true by condition (2). Therefore $Q(n+1)$ is true.

It follows from the principle of mathematical induction (Theorem 2.8) that $Q(n)$ is true for all $n \in \mathbb{N}$. Therefore $P(n)$ is true for all $n \geqslant n_0$. □

Matrices

Matrices are a fundamental tool in linear algebra. We recall some definitions including the usual arithmetic binary operations and the unary operation of transposition

Definition 3.1: Matrix

Let $m, n \in \mathbb{N}$. A **matrix of size $m \times n$** is a rectangular array of numbers having m (horizontal) **rows** and n (vertical) **columns**. The numbers in the array are called the **entries** of the matrix. For the moment, the entries are from $\mathbb{Z}$ or $\mathbb{R}$ or $\mathbb{C}$, but later we will allow other types of entries.

For a matrix A of size $m \times n$, we denote by A_{ij} the entry in the i-th row and j-th column of A

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} \quad \text{or} \quad A = [A_{ij}]$$

We denote by $\mathcal{M}_{m,n}(\mathbb{R})$ the set of all matrices of size $m \times n$ having real entries.
The notations $\mathcal{M}_{m,n}(\mathbb{C})$, $\mathcal{M}_{m,n}(\mathbb{Q})$, and simply $\mathcal{M}_{m,n}$ are used similarly. For the moment, when $\mathbb{F}$ is used it represents one of: $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. However, all results (and proofs) remain valid for any *field*[a].

[a] We will see the definition of a field in Lecture 10

Example 3.2. $A = \begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2i \end{bmatrix} \in \mathcal{M}_{2,3}(\mathbb{C})$, $A_{12} = 3.1$, $A_{21} = -\frac{1}{2}$

Definition 3.3

We fix some terminology and notation for particular kinds of matrices.

- ▷ A matrix with the same number of rows as columns is called a **square matrix**
- ▷ A matrix with only one row is called a **row matrix**
- ▷ A matrix with only one column is called a **column matrix**
- ▷ A matrix with all elements equal to zero is a **zero matrix**, eg., $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right]$
  It is often denoted simply by 0 when the size is clear from context.
- ▷ A matrix A with $A_{ij} = 0$ if $i \neq j$ is called a **diagonal matrix** eg., $\left[\begin{smallmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{smallmatrix}\right]$
- ▷ A square matrix A satisfying $A_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ is called an **identity matrix**. The identity matrix of size $n \times n$ is denoted I_n. For example, $I_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, $I_3 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]$

3.1 Operations on matrices

Definition 3.4: matrix addition

Given two matrices of the same size $A, B \in \mathcal{M}_{m,n}(\mathbb{F})$ we define $A+B \in \mathcal{M}_{m,n}(\mathbb{F})$ by $(A+B)_{ij} = A_{ij} + B_{ij}$. That is,

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \dots & B_{mn} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \dots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \dots & A_{2n}+B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \dots & A_{mn}+B_{mn} \end{bmatrix}$$

Note. 1. Only matrices of the same size can be added together!

2. The addition $A_{ij} + B_{ij}$ is in the field $\mathbb{F}$.

Example 3.5. $\begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2i \end{bmatrix} + \begin{bmatrix} 0 & 0 & 6 \\ 2 & 3 & 8 \end{bmatrix} = \begin{bmatrix} \pi & 3.1 & 6 \\ \frac{3}{2} & 7 & 8+2i \end{bmatrix}$

Definition 3.6: scalar multiplication for matrices

Given a matrix $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $k \in \mathbb{F}$, define a matrix $kA \in \mathcal{M}_{m,n}(\mathbb{F})$ by $(kA)_{ij} = k \times A_{ij}$. That is,

$$k \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} = \begin{bmatrix} kA_{11} & kA_{12} & \dots & kA_{1n} \\ kA_{21} & kA_{22} & \dots & kA_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ kA_{m1} & kA_{m2} & \dots & kA_{mn} \end{bmatrix}$$

Example 3.7. $2\begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2i \end{bmatrix} = \begin{bmatrix} 2\pi & 6.2 & 0 \\ -1 & 8 & 4i \end{bmatrix}$

Remark. We write $A - B$ to mean $A + (-1)B$

There are other useful operations with matrices.

Definition 3.8: matrix multiplication

Given matrices $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{n,p}(\mathbb{F})$ we define their product, AB, to be the matrix in $\mathcal{M}_{m,p}(\mathbb{F})$ given by

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

Note. Two matrices A and B can only be multiplied together (in that order) if the number of columns of A is equal to the number of rows of B.

Exercise 22. Using the definition of matrix multiplication, show that for any matrix $A \in \mathcal{M}_{m,n}(\mathbb{F})$, $B \in \mathcal{M}_{n,m}(\mathbb{F})$, and $k \in \mathbb{F}$ we have:

(a) $AI_n = A$ and $I_mA = A$

(b) $A(kB) = k(AB)$

Example 3.9. 1. $\begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2i \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 0 & 6 \\ 7 & 2 \end{bmatrix} = \begin{bmatrix} 2\pi & 5\pi + 18.6 \\ -1+14i & \frac{43}{2} + 4i \end{bmatrix}$

2. $\begin{bmatrix} 2 & 5 \\ 0 & 6 \\ 7 & 2 \end{bmatrix} \begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2i \end{bmatrix} = \begin{bmatrix} 2\pi - \frac{5}{2} & 26.2 & 10i \\ -3 & 24 & 12i \\ 7\pi - 1 & 29.7 & 4i \end{bmatrix}$

3. $\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Remark.

1. Matrix multiplication is *not commutative*!
   The order matters. For two matrices A and B it's possible that AB is defined, but that BA is not. Even when both AB and BA are defined, they are not necessarily equal.

2. Matrix multiplication *is associative*.
   For any $A \in \mathcal{M}_{m,n}(\mathbb{F})$, $B \in \mathcal{M}_{n,p}(\mathbb{F})$, and $C \in \mathcal{M}_{p,q}(\mathbb{F})$ we have $(AB)C = A(BC)$. This can be proved directly from the definition of matrix multiplication.

Definition 3.10: transpose

Given a matrix $A \in \mathcal{M}_{m,n}$ we define its **transpose** to be the matrix $A^T \in \mathcal{M}_{n,m}$ given by $(A^T)_{ij} = A_{ji}$. That is, A^T is obtained from A by interchanging the rows and columns of A. A matrix is called **symmetric** if $A^T = A$.

Example 3.11. $A = \begin{bmatrix} \pi & 3.1 & 0 \\ -\frac{1}{2} & 4 & 2 \end{bmatrix}$, $A^T = \begin{bmatrix} \pi & -\frac{1}{2} \\ 3.1 & 4 \\ 0 & 2 \end{bmatrix}$

Exercise 23. Using the definitions of matrix addition and transpose prove that for all $A, B \in \mathcal{M}_{m,n}(\mathbb{F})$, $(A+B)^T = A^T + B^T$.

Lemma 3.12

Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{n,p}(\mathbb{F})$. Then $(AB)^T = B^T A^T$.

Proof. Note first that $(AB)^T$ and $B^T A^T$ are both of size $p \times m$. To show that they are equal we need to show that

$$\forall i \in \{1, \dots, p\}\ \forall j \in \{1, \dots, n\}, \quad ((AB)^T)_{ij} = (B^T A^T)_{ij}$$

Let $i \in \{1, \dots, p\}$ and $j \in \{1, \dots, n\}$. Then

$$\begin{aligned}
((AB)^T)_{ij} &= (AB)_{ji} && \text{(definition of transpose)} \\
&= \sum_{k=1}^{n} A_{jk} B_{ki} && \text{(definition of multiplication)} \\
&= \sum_{k=1}^{n} B_{ki} A_{jk} && \text{(multiplication in } \mathbb{F} \text{ is commutative)} \\
&= \sum_{k=1}^{n} (B^T)_{ik} (A^T)_{kj} && \text{(definition of transpose)} \\
&= (B^T A^T)_{ij} && \text{(definition of multiplication)}
\end{aligned}$$

□

3.2 Exercises

24. Suppose that A, B, C, and D are matrices with sizes given by:
$A \in \mathcal{M}_{2,3}$, $B \in \mathcal{M}_{1,3}$, $C \in \mathcal{M}_{2,2}$, $D \in \mathcal{M}_{2,1}$.

    Determine which of the following expressions are defined. For those which are defined, give the size of the resulting matrix.

    (a) CA
    (b) BD
    (c) DB
    (d) CDA^T
    (e) $CA + DB$
    (f) $AB^TD^T + C$

25. Give examples of matrices $A \in \mathcal{M}_{3,3}(\mathbb{C})$ such that A is:

    (a) a diagonal matrix; that is, $A_{ij} = 0$ if $i \neq j$
    (b) a scalar matrix; that is, A is a diagonal matrix with $A_{ii} = A_{jj}$ for all i, j
    (c) a symmetric matrix; that is, $A_{ij} = A_{ji}$ for all i, j
    (d) an upper triangular matrix; that is, $A_{ij} = 0$ if $i > j$

26. Let

$$A = \begin{bmatrix} -1 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & -2 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 4 & -2 \\ 3 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

    Find

    (a) $A + B$
    (b) $2A - 3B$
    (c) $A - \lambda I \quad (\lambda \in \mathbb{R})$
    (d) A^T
    (e) AB
    (f) BA

27. Calculate AB, BC, A^TC^T and $(CA)^T$ given that

$$A = \begin{bmatrix} 1 & 3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 4 \end{bmatrix} \qquad C = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

28. Evaluate the following matrix products:

    (a) $\begin{bmatrix} 3 & 4 & 2 \\ 1 & 3 & 6 \\ 7 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 0 \end{bmatrix}$

    (b) $\begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -1 & -2 \end{bmatrix}$

    (c) $\begin{bmatrix} 7+i & 6 & -4+3i \end{bmatrix} \begin{bmatrix} 0 \\ 1-i \\ 1 \end{bmatrix}$

    (d) $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 7 & 6 & -4 \end{bmatrix}$

    (e) $\begin{bmatrix} 3 & -6 & 0 \\ 0 & 2 & -2 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$

    (f) $\begin{bmatrix} 4 & 3 \\ 6 & 6 \\ 8 & 9 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \end{bmatrix}$

29. Consider the following matrices:

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 3 & 1 & -1 \\ -1 & 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 0 & 0 \\ 2 & 3 & -1 \\ -2 & 3 & 4 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 1 & -2 \\ 2 & 3 & 5 \\ 0 & 1 & 2 \end{bmatrix}$$

    Verify that

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

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

(c) $A(B + C) = AB + AC$

30. Find examples of the following:

    (a) a non-zero matrix $A \in \mathcal{M}_{2,2}(\mathbb{R})$ with $A^2 = 0$

    (b) a matrix $B \in \mathcal{M}_{2,2}(\mathbb{R})$ with $B^2 = -I_2$

    (c) matrices $C, D \in \mathcal{M}_{2,2}(\mathbb{R})$ with no zero entries but with $CD = 0$

31. (a) Show that if the matrix products AB and BA are both defined, then AB and BA are square matrices.

    (b) Show that if A is an $m \times n$ matrix and $A(BA)$ is defined, then B is an $n \times m$ matrix.

32. Verify, using appropriate trigonometric identities, that

$$\begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) \\ \sin(\theta_1) & \cos(\theta_1) \end{bmatrix} \begin{bmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{bmatrix} = \begin{bmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{bmatrix}$$

33. Suppose that a 2 by 2 matrix $A \in \mathcal{M}_{2,2}(\mathbb{C})$ satisfies $AB = BA$ for every 2 by 2 matrix B. That is A satisfies

$$\forall B \in \mathcal{M}_{2,2}(\mathbb{C}),\ AB = BA$$

In this exercise we show that A must be equal to zI_2 for some $z \in \mathbb{C}$.

Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. By considering the two cases $B = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, show that $a = d$ and $b = c = 0$.

34. Consider the following matrices:

$$C = \begin{bmatrix} 1 & -1 \\ 5 & -4 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

By setting up and solving appropriate simultaneous equations, find:

    (a) all matrices $A \in \mathcal{M}_{2,2}(\mathbb{C})$ that satisfy $AC = CA$

    (b) all matrices $A \in \mathcal{M}_{3,3}(\mathbb{C})$ that satisfy $AD = DA$

35. Let $A, B \in \mathcal{M}_{n,n}(\mathbb{C})$. Suppose that $A^2 = A$. Show that $(AB - ABA)^2 = 0$.
(Note we may not assume that $n = 2$ nor that $AB = BA$.)

36. Let $A, B \in \mathcal{M}_{n,n}$ be symmetric matrices. Show that AB is symmetric if and only if $AB = BA$.

37. (a) Give an example of three matrices $A, B \in \mathcal{M}_{2,2}$ and $C \in \mathcal{M}_{2,1}$ such that $C \neq 0$ and $AC = BC$ but $A \neq B$.

    (b) Let $A, B \in \mathcal{M}_{m,n}(\mathbb{C})$. Suppose that $AC = BC$ for all $C \in \mathcal{M}_{n,1}(\mathbb{C})$. Show that $A = B$.

Further material for lecture 3

- Matrices

  Elementary Linear Algebra by Anton and Rorres, §1.3
  Linear Algebra Done Right by Axler, §3.C

- Some histrory

  Matrices and determinants on MacTutor

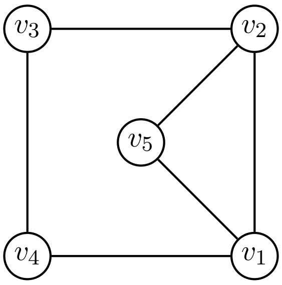
- A matrix $A \in \mathcal{M}_{m,n}(\mathbb{F})$ determines (and is determined by) a function

  $$f : \{1, \dots, m\} \times \{1, \dots, n\} \to \mathbb{F}, \qquad f(i, j) = A_{i,j}$$

  The only difference is notation.

- **Adjacency matrix of a graph**
  Given a finite graph with vertices $v_1, \dots, v_n$ we define a matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$ by

  $$A_{i,j} = \begin{cases} 1 & \text{if the vertices } v_i \text{ and } v_j \text{ are connected by an edge} \\ 0 & \text{if the vertices } v_i \text{ and } v_j \text{ are not connected by an edge} \end{cases}$$



$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad A^3 = \begin{bmatrix} 2 & 6 & 1 & 5 & 4 \\ 6 & 2 & 5 & 1 & 4 \\ 1 & 5 & 0 & 4 & 2 \\ 5 & 1 & 4 & 0 & 2 \\ 4 & 4 & 2 & 2 & 2 \end{bmatrix}$$

Note. Given any $k \in \mathbb{N}$, the i,j-th entry of A^k gives the number of edge paths of length k from v_i to v_j. (If you're feeling adventurous, you could try proving this using induction on k.) For example, there are six edge paths of length three from v_1 to v_2 in the above graph.

Matrix inverses and linear systems

4.1 Inverse of a matrix

Matrices follow many of the algebraic properties that we are familiar with from the real numbers (such as the distributive law). It's natural to think about "dividing by a matrix" in the sense of multiplying by the multiplicative inverse.

An important property of the real numbers is that every non-zero real number has a multiplicative inverse, that is, $\forall\, a \in \mathbb{R} \setminus \{0\}\ \exists\, b \in \mathbb{R}, ab = 1$. Some, *but not all*, non-zero square matrices have a multiplicative inverse in the same sense.

Definition 4.1: Matrix inverse

A square matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is called **invertible** if there exists a matrix $B \in \mathcal{M}_{n,n}(\mathbb{F})$ such that $AB = I_n$ and $BA = I_n$. The matrix B is called the **inverse** of A and is denoted A^{-1}. If A is not invertible, we say that A is **singular**.

Remark. Calling B *the* inverse of A needs some justification. It's possible to show that there can be at most one matrix that satisfies the above property. That is, if $B, C \in \mathcal{M}_{n,n}(\mathbb{F})$ are such that $AB = I_n$ and $BA = I_n$ and $AC = I_n$ and $CA = I_n$, then $B = C$.

Example 4.2.

1. $\begin{bmatrix} 2 & i \\ i & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & i \\ i & -2 \end{bmatrix}$
2. $\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ is singular
3. $\begin{bmatrix} 0 & -3 & -2 \\ 1 & -4 & -2 \\ -3 & 4 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & -5 & -2 \\ 5 & -6 & -2 \\ -8 & 9 & 3 \end{bmatrix}$

It's easy to check that the given inverses are correct by simply multiplying and verifying that the result is the identity matrix. We will describe a method for calculating the inverse of a matrix in a later section.

Remark. Notice that it follows immediately from the definition that:

1. For A to be invertible it must be square
2. (If it exists) A^{-1} has the same size as A
3. If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$
4. $I_n^{-1} = I_n$

We note some other useful results about invertibility.

Lemma 4.3: Product of invertible matrices is invertible

If A and C are invertible matrices of the same size, then AC is invertible and

$$(AC)^{-1} = C^{-1}A^{-1}$$

Proof. Let $A, C \in \mathcal{M}_{n,n}(\mathbb{F})$ be two invertible matrices. We need to verify that $C^{-1}A^{-1}$ satisfies the conditions given in the definition of the matrix inverse.

$$\begin{aligned}(AC)(C^{-1}A^{-1}) &= A(CC^{-1})A^{-1} && \text{(associativity)}\\ &= AI_nA^{-1} && \text{(since } CC^{-1} = I_n\text{)}\\ &= AA^{-1} && \text{(since } AI_n = A\text{)}\\ &= I_n\end{aligned}$$

and, similarly

$$(C^{-1}A^{-1})(AC) = C^{-1}(A^{-1}A)C = C^{-1}I_nC = C^{-1}C = I_n$$

□

Exercise 38. Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. For $k \in \mathbb{N}$ we define A^k to be the product of k copies of A, that is,

$$A^k = \underbrace{A \times A \times \cdots \times A}_{k \text{ copies}}$$

Suppose that $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is invertible. Show that

(a) $(A^k)^{-1} = (A^{-1})^k$ for all $k \in \mathbb{N}$.

(b) If $a \in \mathbb{F}$ with $a \neq 0$, then aA is invertible and $(aA)^{-1} = \frac{1}{a}A^{-1}$.

(c) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

The following technical result will be useful later.

Lemma 4.4

Let $A, B \in \mathcal{M}_{n,n}$ be two square matrices and let $i \in \{1, 2, \ldots, n\}$. If all entries in the i-th row of A are equal to zero, then all entries in the i-th row of AB are zero.

In particular, if a square matrix has a row consisting entirely of zeros, then it is singular.

Proof. Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$ be a square matrix with all entries in the i-th row of A equal to zero. That is, $\forall j \in \{1, \ldots, n\}$, $A_{ij} = 0$. Let $B \in \mathcal{M}_{n,n}(\mathbb{F})$. For the same fixed i, we have (for all j)

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} = \sum_{k=1}^{n} 0 \times B_{kj} = \sum_{k=1}^{n} 0 = 0$$

Therefore, the i-th row of AB has all entries equal to zero. In particular, $(AB)_{ii} \neq 1$ and hence $AB \neq I_n$. Therefore, no matrix B can satisfy the properties needed to be the inverse of A. □

4.2 Linear Systems

We would like to have an efficient method to solve simultaneous linear equations. Suppose that we have n vairiables $x_1, \ldots, x_n$ and m linear equations that they should simultaneously satisfy.

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

This is sometimes called a **linear system**.

Letting

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

the linear system $(*)$ can be written as

$$AX = B \tag{$\dagger$}$$

Example 4.5. The linear system below

$$\begin{aligned} x + (1+i)y + 7z &= 1 - i \\ ix + 2y - z &= 2 + i \end{aligned}$$

can be written as

$$\begin{bmatrix} 1 & 1+i & 7 \\ i & 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1-i \\ 2+i \end{bmatrix}$$

Given $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$ we would like to find all $X \in \mathcal{M}_{n,1}(\mathbb{F})$ such that the equation $AX = B$ is satisfied. Thinking about our experience with solving simple equations of the form $2x = 7$, our first thought might be to multiply on both sides by A^{-1}. The problem is that A need not be invertible (or even square). However, if A does happen to be invertible, then we have the following.

Proposition 4.6

Let $A \in \mathcal{M}_{m,m}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$. If A is invertible, then the equation $AX = B$ has a unique solution and it is given by $X = A^{-1}B$.

Proof. First note that $A^{-1}B$ is a solution since $AA^{-1}B = I_m B = B$. Now suppose that X is any solution. Then we have

$$AX = B \implies A^{-1}AX = A^{-1}B \implies I_m X = A^{-1}B \implies X = A^{-1}B$$

Therefore, $X = A^{-1}B$ is the only solution. □

The general case of a linear system (in which A is not necessarily invertible) is discussed in the next section. The technique will rest on the following observation. Suppose we have $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$ and an *invertible* matrix $E \in \mathcal{M}_{m,m}(\mathbb{F})$. Define $A' = EA$ and $B' = EB$. Then for all $X \in \mathcal{M}_{n,1}(\mathbb{F})$ we have

$$AX = B \iff A'X = B'$$

The goal will be to arrange for the new linear system $A'X = B'$ to be as simple as possible.

4.3 Exercises

39. Let $a, b, c, d \in \mathbb{C}$.

 (a) Suppose that $ad - bc \neq 0$. Show (using the definition of inverse) that the inverse of the matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathcal{M}_{2,2}(\mathbb{C})$ is the matrix

$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{C})$$

(b) Suppose now that $ad - bc = 0$.

i) Show that $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \left[\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$

ii) Use the above observation to show that the matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ has no inverse.

40. Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$. Suppose that there exists $B \in \mathcal{M}_{n,k}(\mathbb{C})$ such that $AB = 0$ and $B \neq 0$. Show that A is not invertible. (Be careful, A need not be equal to 0.)

41. (a) Show that if a square matrix A satisfies $A^2 - 4A + 3I = 0$ then $A^{-1} = \frac{1}{3}(4I - A)$.

(b) Veify these relations in the case that $A = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$

42. Let

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

Show that $A^{-1} = -\frac{1}{3}(A^2 - 2A - 4I)$.

43. Write the following systems of linear equations in the form $AX = B$. Then, by using the inverse of A, solve the system (i.e., find X).

(a)

$$\begin{aligned} 2x - 3y &= 3 \\ 3x - 5y &= 1 \end{aligned}$$

(b)

$$\begin{aligned} x + z &= -1 \\ 2x + 3y + 4z &= 3 \\ -x - 2z &= 3 \end{aligned}$$

(Your answer to the previous exercise will be useful here.)

44. Suppose $A, B, P \in \mathcal{M}_{n,n}(\mathbb{F})$ are such that P is invertible and $A = PBP^{-1}$. Show that

$$A^k = PB^kP^{-1}$$

(for all $k \in \mathbb{N}$).

45. Let A be a square matrix. Show that if A^2 is invertible, then A is invertible.
(It is possible to give a proof that uses only what we've seen so far. We will see later how this can also be shown using the determinant.)

Extra material for lecture 4

- ▷ Matrix inverse

  Elementary Linear Algebra by Anton and Rorres, §1.3
  Linear Algebra Done Right by Axler, §3.C

- ▷ Uniqueness of matrix inverse

  Suppose that $B, C \in \mathcal{M}_{n,n}(\mathbb{F})$ are such that $AB = I_n$ and $BA = I_n$ and $AC = I_n$ and $CA = I_n$. We would then have $AB = AC$ and

$$\begin{aligned} AB = AC &\implies B(AB) = B(AC) && \text{(multiplying on the left by } B\text{)} \\ &\implies (BA)B = (BA)C && \text{(associativity)} \\ &\implies I_nB = I_nC && \text{(since } BA = I\text{)} \\ &\implies B = C \end{aligned}$$

  Therefore, if a matrix does have an inverse, it is unique.

- ▷ Linear systems

  Elementary Linear Algebra by Anton and Rorres, §1.1

LECTURE 5

Row operations and elementary matrices

5.1 Row operations

For a linear system $AX = B$ in which the matrix A is not invertible Proposition 4.6 does not apply. To handle the general case we introduce the notion of elementary row operations and the corresponding elementary matrices. They will turn out to be useful in other contexts, including when calculating the inverse of a matrix.

Given a linear system there are certain operations that can be carried out without changing the set of solutions. For example, changing the order in which the equations are listed does not alter the set of solutions. Similarly, multiplying one of the equations by a (non-zero) constant does change the solutions. This leads us to define the following operations on matrices.

Definition 5.1: Elementary row operations

An **elementary row operation** on a matrix $A \in \mathcal{M}_{m,n}(\mathbb{F})$ is one of the following:

1. Interchanging two rows.
2. Multiplying a row by a non-zero element of $\mathbb{F}$.
3. Adding a multiple of one row to another.

Note that applying one of the above row operations does not change the size of the matrix. We say that two matrices are **row equivalent**[*]if one can be obtained from the other by a sequence of row operations. If $A, B \in \mathcal{M}_{m,n}(\mathbb{F})$ are row equivalent, we write $A \sim B$.

Example 5.2. $\begin{bmatrix} 0 & i & 2 \\ 1 & 3 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2+6i \\ 0 & 1 & -2i \end{bmatrix}$ since one can be obtained from the other as follows

$$\begin{bmatrix} 0 & i & 2 \\ 1 & 3 & -2i \end{bmatrix} \xrightarrow{R1 \leftrightarrow R2} \begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix} \xrightarrow{(-i) \times R2} \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix} \xrightarrow{R1 - 3R2} \begin{bmatrix} 1 & 0 & -2+6i \\ 0 & 1 & -2i \end{bmatrix}$$

Definition 5.3

An **elementary matrix** is a matrix obtained from an identity matrix I_n by performing a *single* elementary row operation.

It follows from the definition that elementary matrices are always square.

Example 5.4. The matrices $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$, and $\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$ are elementary matrices. To justify this note that

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{R1 \leftrightarrow R2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{(-i) \times R2} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{R1 - 3R2} \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$$

[*] Row equivalence is an example of what is called an *equivalence relation*.

The connection between elementary matrices and elementary row operations is given by the following result.

Lemma 5.5

Let $A, B \in \mathcal{M}_{m,n}(\mathbb{F})$. Suppose that B is obtained from A by applying a single row operation and let $E \in \mathcal{M}_{m,m}(\mathbb{F})$ be the elementary matrix obtained by applying the same row operation to I_m. Then $B = EA$.

Proof. We consider the three kinds of row operation separately.

Suppose first that the row operation swaps rows p and q. Then we have

$$B_{ij} = \begin{cases} A_{ij} & \text{if } i \notin \{p,q\} \\ A_{qj} & \text{if } i = p \\ A_{pj} & \text{if } i = q \end{cases} \qquad E_{ij} = \begin{cases} I_{ij} & \text{if } i \notin \{p,q\} \\ I_{qj} & \text{if } i = p \\ I_{pj} & \text{if } i = q \end{cases}$$

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} \sum_{k=1}^{m} I_{ik}A_{kj} & \text{if } i \notin \{p,q\} \\ \sum_{k=1}^{m} I_{qk}A_{kj} & \text{if } i = p \\ \sum_{k=1}^{m} I_{pk}A_{kj} & \text{if } i = q \end{cases} = \begin{cases} A_{ij} & \text{if } i \notin \{p,q\} \\ A_{qj} & \text{if } i = p \\ A_{pj} & \text{if } i = q \end{cases} = B_{ij}$$

Suppose now that the row operation is to multiply the p-th row by $\lambda \in \mathbb{F} \setminus \{0\}$. Then we have

$$B_{ij} = \begin{cases} A_{ij} & \text{if } i \neq p \\ \lambda A_{ij} & \text{if } i = p \end{cases} \qquad E_{ij} = \begin{cases} I_{ij} & \text{if } i \neq p \\ \lambda I_{ij} & \text{if } i = p \end{cases}$$

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} \sum_{k=1}^{m} I_{ik}A_{kj} & \text{if } i \neq p \\ \sum_{k=1}^{m} \lambda I_{ik}A_{kj} & \text{if } i = p \end{cases} = \begin{cases} A_{ij} & \text{if } i \neq p \\ \lambda A_{ij} & \text{if } i = p \end{cases} = B_{ij}$$

Therefore $EA = B$ in this case also.

Finally, suppose that the row operation replaces row p by itself plus $\lambda \in \mathbb{F}$ times row q. We have

$$B_{ij} = \begin{cases} A_{ij} & \text{if } i \neq p \\ A_{pj} + \lambda A_{qj} & \text{if } i = p \end{cases} \qquad E_{ij} = \begin{cases} I_{ij} & \text{if } i \neq p \\ I_{pj} + \lambda I_{qj} & \text{if } i = p \end{cases}$$

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} \sum_{k=1}^{m} I_{ik}A_{kj} & \text{if } i \neq p \\ \sum_{k=1}^{m} (I_{pk} + \lambda I_{qk})A_{kj} & \text{if } i = p \end{cases} = \begin{cases} A_{ij} & \text{if } i \neq p \\ A_{pj} + \lambda A_{qj} & \text{if } i = p \end{cases} = B_{ij}$$

Again, this shows that $EA = B$.

□

Corollary 5.6

Elementary matrices are invertible.

Proof. Given any row operation ρ there is a row operation ρ' which undoes the effect of ρ. Let the corresponding elementary matrices be E and E'. Then applying the lemma with $A = I$ we have $E'EI = I$ and $EE'I = I$. Hence $EE' = I$ and $E'E = I$. Therefore E is invertible and $E^{-1} = E'$. □

Example 5.7. Considering the row operations and corresponding elementary matrices from Examples 5.2 and 5.4 we have:

row operation $A \xrightarrow{\rho} B$	elementary matrix E	$EA = B$
$\begin{bmatrix} 0 & i & 2 \\ 1 & 3 & -2i \end{bmatrix} \xrightarrow{R1 \leftrightarrow R2} \begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix}$	$E_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & i & 2 \\ 1 & 3 & -2i \end{bmatrix} = \begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix}$
$\begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix} \xrightarrow{(-i) \times R2} \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix}$	$E_2 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix}$
$\begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix} \xrightarrow{R1 - 3R2} \begin{bmatrix} 1 & 0 & -2+6i \\ 0 & 1 & -2i \end{bmatrix}$	$E_3 = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2+6i \\ 0 & 1 & -2i \end{bmatrix}$

Notice that

$$E_3E_2E_1 \begin{bmatrix} 0 & i & 2 \\ 1 & 3 & -2i \end{bmatrix} = E_3E_2 \begin{bmatrix} 1 & 3 & -2 \\ 0 & i & 2 \end{bmatrix} = E_3 \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2i \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2+6i \\ 0 & 1 & -2i \end{bmatrix}$$

The order in which the E_i have been multiplied is important!

Exercise 46. For each of the row operations ρ_i in Example 5.7 write down a row operation ρ_i' that undoes the effect of ρ_i. Write down the elementary matrix E_i' that corresponds to ρ'. Verify that $E_iE_i' = I_2$.

$\rho_1' = \rho_1$, $E_1' = E_1$, ρ_2' is $i \times R2$, $E_2' = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$, ρ_3' is $R1 + 3R2$, $E_3' = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

Lemma 5.8

Let $A, B \in \mathcal{M}_{m,n}(\mathbb{F})$. If A and B are row equivalent, then there exists an invertible matrix $E \in \mathcal{M}_{m,m}(\mathbb{F})$ such that $B = EA$.

Proof. Since A and B are row equivalent, there is a sequence of elementary row operations that transforms A to B

$$A \xrightarrow{\rho_1} A_1 \xrightarrow{\rho_2} A_2 \xrightarrow{\rho_3} \cdots \xrightarrow{\rho_k} A_k = B$$

Let E_i be the elementary matrix corresponding to the row operation ρ_i and define $E = E_kE_{k-1} \dots E_2E_1$. Applying Lemma 5.5 we have

$$B = E_kE_{k-1} \dots E_2E_1A = EA$$

Since each E_i is invertible (Corollary 5.6), E is invertible (Lemma 4.3.) □

5.2 Row echelon form

We now define the first version of a "simplified" matrix that will be useful for solving linear systems and for other applications (such as finding bases and calculating rank).

Definition 5.9: Row echelon form (REF)

The leftmost non-zero entry in a row is called the **leading entry** of that row.

A matrix is in **row echelon form** if it satisfies the following conditions:

1. For any two non-zero rows, the leading entry of the lower row is further to the right than the leading entry in the higher row.

2. Any row that consists entirely of zeros is lower than every non-zero row.

Note. Some authors add the condition that to be in row echelon form all leading entries should be equal to 1. We are *not* including this requirement for what we call row echelon form. The extra condition is not needed for any of our applications.

Examples 5.10.

$$\begin{bmatrix} 0 & 1 & -2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 & 2 & 3 \\ 0 & 4 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix} \qquad \text{are in row echelon form}$$

$$\begin{bmatrix} 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 3 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -3 & 6 & -4 & 9 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 4 & 1 & 2 \end{bmatrix} \qquad \text{are not in row echelon form}$$

5.3 Exercises

47. For each of the following row operations find the corresponding elementary matrix.

(a) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{R_2 - 3R_1} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \end{bmatrix} \xrightarrow{R_2 - 3R_1} \begin{bmatrix} 1 & 2 & 5 \\ 0 & -2 & -9 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 2 \\ 5 & 6 \\ 3 & 4 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{bmatrix} \xrightarrow{R_2 \times (-2)} \begin{bmatrix} 3 & 1 & 4 \\ -2 & -10 & -18 \\ 2 & 6 & 5 \end{bmatrix}$

48. Let $A \in \mathcal{M}_{3,5}(\mathbb{C})$. Suppose that B is obtained from A by the following sequence of row operations in the given order.

1) $R_1 \leftrightarrow R_2$ 2) $R_3 - 2R_1$ 3) $R_1 \times 3$

Find a single matrix E such that $B = EA$.

49. Which of the following matrices are in row echelon form?

(a) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 4 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & 2 & 4 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 3 & 0 & 2 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$

Extra material for lecture 5

- ▷ **row operations, row echelon form**

  Elementary Linear Algebra by Anton and Rorres, §1.1, §1.5

- ▷ **equivalence relations**

  The Art of Proof by Beck and Geoghegan, §6.1.

- ▷ **column operations**

  Elementary column operations can be defined in way analogous to row operations. The elementary matrices corresponding to column operations are multiplied on the right rather than on the left. Here's an example to illustrate.

$$\begin{bmatrix} 2 & 1 & 3 \\ 5 & 4 & 6 \end{bmatrix} \xrightarrow{C_1 \leftrightarrow C_2} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \xrightarrow[C_3-3C_1]{C_2-2C_1} \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \end{bmatrix} \xrightarrow{C_3-2C_2} \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 3 \\ 5 & 4 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \end{bmatrix}$$

LECTURE 6

Gaussian elimination and types of solution set

We look at how to solve a linear system using 'Gaussian elimination' to put a corresponding matrix into row echelon form. We also look at how to determine if a linear system has no solutions, a unique solution, or more than one solution.

6.1 Gaussian elimination

Any matrix can be put into row echelon form by performing a sequence of row operations as follows.

Algorithm 6.1: Gaussian elimination[*](To put a matrix into REF)

1. Consider the first column that is not all zeros. Interchange rows (if necessary) to bring a non-zero entry to the top of that column. (The 'leading entry'.)
2. Add suitable multiples of the top row to lower rows so that all entries below the leading entry are zero.
3. Start again with Step 1 applied to the matrix without the first row.
   (stop if there are no more rows)

Example 6.2. Here's an example of applying the above procedure. It's a good idea to record the row operations being used at each step.

$$\begin{bmatrix} 3 & 2 & -1 & -15 \\ 1 & 1 & -4 & -30 \\ 3 & 1 & 3 & 11 \\ 3 & 3 & -5 & -41 \end{bmatrix} \xrightarrow{R2-\frac{1}{3}R1} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 3 & 1 & 3 & 11 \\ 3 & 3 & -5 & -41 \end{bmatrix} \xrightarrow{R3-R1} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & -1 & 4 & 26 \\ 3 & 3 & -5 & -41 \end{bmatrix} \xrightarrow{R4-R1} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & -1 & 4 & 26 \\ 0 & 1 & -4 & -26 \end{bmatrix}$$

$$\xrightarrow{R3+3R2} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & 0 & -7 & -49 \\ 0 & 1 & -4 & -26 \end{bmatrix} \xrightarrow{R4-3R2} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & 0 & -7 & -49 \\ 0 & 0 & 7 & 49 \end{bmatrix}$$

$$\xrightarrow{R4+R3} \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & 0 & -7 & -49 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The final matrix is in row echelon form. All matrices above are row equivalent to one another.

6.2 Using Gaussian elimination to solve a linear system

Given $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$, to solve the linear system $AX = B$, we can proceed as follows.

1. Form a matrix $[A|B] \in \mathcal{M}_{m,n+1}(\mathbb{F})$ by adjoining B as an extra (final) column to A.
   (This matrix $[A|B]$ is sometimes called an **augmented matrix**.)

[*] Named after the 19th century German mathematician Carl Friedrich Gauss. The technique had been previously discovered by Chinese mathematicians during the Han dynasty (206 BCE – 220 CE).

2. Apply Gaussian elimination starting with $[A|B]$ to obtain a row echelon matrix $[A'|B']$.
3. Solve the new, simplified set of equations $A'X = B'$ starting from the last equation and working up. (This is sometimes called **back substitution**)

Remark. Why does this work? We know that $[A'|B'] = E[A|B]$ for some invertible matrix E by Lemma 5.8. Therefore $A' = EA$ and $B' = EB$. Therefore $AX = B$ and $A'X = B'$ have the same set of solutions (see the comment at the end of section 4.2).

Example 6.3. Let's use this technique to find all solutions to the following linear system.

$$\begin{aligned} 3x + 2y - z &= -15 \\ x + y - 4z &= -30 \\ 3x + y + 3z &= 11 \\ 3x + 3y - 5z &= -41 \end{aligned} \tag{$*$}$$

Let $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & 1 & -4 \\ 3 & 1 & 3 \\ 3 & 3 & -5 \end{bmatrix}$, $B = \begin{bmatrix} -15 \\ -30 \\ 11 \\ -41 \end{bmatrix}$.

Then

$$[A|B] = \begin{bmatrix} 3 & 2 & -1 & -15 \\ 1 & 1 & -4 & -30 \\ 3 & 1 & 3 & 11 \\ 3 & 3 & -5 & -41 \end{bmatrix} \sim \begin{bmatrix} 3 & 2 & -1 & -15 \\ 0 & \frac{1}{3} & \frac{-11}{3} & -25 \\ 0 & 0 & -7 & -49 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ (as shown in Example 6.2)}$$

The new linear system is

$$\begin{aligned} 3x + 2y - \phantom{\tfrac{11}{3}}z &= -15 \\ \tfrac{1}{3}y - \tfrac{11}{3}z &= -25 \\ -7z &= -49 \end{aligned}$$

We then get

$$\begin{aligned} z &= 7 && \text{(from the last equation)} \\ y &= 11z - 75 && \text{(from the second last equation)} \\ &= 2 && \text{(since } z = 7\text{)} \\ x &= \tfrac{-2}{3}y + \tfrac{1}{3}z - 5 && \text{(from the first equation)} \\ &= -4 && \text{(since } y = 2 \text{ and } z = 7\text{)} \end{aligned}$$

So the original linear system $(*)$ has a unique solution and it is given by $(x, y, z) = (-4, 2, 7)$.

6.3 Inconsistent linear systems

Definition 6.4

A linear system is called **inconsistent** if it has no solutions.

Example 6.5. Find all solutions of the system

$$\begin{aligned} x - y + z &= 3 \\ x - 7y + 3z &= -11 \\ 2x + y + z &= 16 \end{aligned}$$

We form the matrix $[A|B]$ and reduce to row echelon form.

$$[A|B] = \begin{bmatrix} 1 & -1 & 1 & 3 \\ 1 & -7 & 3 & -11 \\ 2 & 1 & 1 & 16 \end{bmatrix} \xrightarrow[R3-2R1]{R2-R1} \begin{bmatrix} 1 & -1 & 1 & 3 \\ 0 & -6 & 2 & -14 \\ 0 & 3 & -1 & 10 \end{bmatrix} \xrightarrow{R3+\frac{1}{2}R2} \begin{bmatrix} 1 & -1 & 1 & 3 \\ 0 & -6 & 2 & -14 \\ 0 & 0 & 0 & 3 \end{bmatrix} = [A'|B']$$

The new linear system is

$$\begin{aligned} x - \ y + \ z &= 3 \\ -6y + 2z &= -14 \\ 0 &= 3 \end{aligned}$$

This system is clearly never satisfied no matter what values of x, y, z we take!
The linear system is inconsistent.

Note. As we saw in the above example, a linear system is inconsistent if there is a row of the form $[0 \cdots 0 \,|\, k]$ (with $k \neq 0$) in the row echelon form. Given the way in which row echelon form is defined, such a row occurs precisely when there is a leading entry in the final column of the row echelon form matrix $[A'|B']$. We shall see in the next section that this is the only situation in which a linear system is inconsistent.

6.4 Consistent linear systems

Definition 6.6

A linear system is called **consistent** if it has at least one solution.

In Example 6.3 we saw a consistent linear system for which there was a unique solution. Here is an example in which there turns out to be more than one solution.

Example 6.7. Find all solutions of the system

$$\begin{aligned} x + \ y + \ z &= 4 \\ 2x + \ y + 2z &= 9 \\ 3x + 2y + 3z &= 13 \end{aligned}$$

$$[A|B] = \begin{bmatrix} 1 & 1 & 1 & 4 \\ 2 & 1 & 2 & 9 \\ 3 & 2 & 3 & 13 \end{bmatrix} \xrightarrow[R3-3R1]{R2-2R1} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{bmatrix} \xrightarrow{R3-R2} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The new system is

$$\begin{aligned} x + \ y + \ z &= 4 \\ -y \qquad &= 1 \end{aligned}$$

From which we get

$$\begin{aligned} y &= -1 && \text{(from the last (non -trivial) equation)} \\ x &= 4 - y - z && \text{(from the first equation)} \\ &= 5 - z \end{aligned}$$

For any value of z, we get a solution by taking $y = -1$ and $x - 5 - z$. That is, the set of all solutions is

$$S = \{(5 - z, -1, z) \mid z \in \mathbb{F}\}$$

Lemma 6.8

Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$. Suppose that $[A|B] \sim [A'|B']$ and that $[A'|B']$ is in row echelon form. Define $r \in \mathbb{N} \cup \{0\}$ to be the number of non-zero rows in $[A'|B']$.

1. The system $AX = B$ is consistent iff there is no leading entry in the final column of $[A'|B']$.

Suppose now that the system is consistent.

2. Then $r \leqslant n$ and the full solution will require $n - r$ parameters.

Sketch of proof. We have already noted above that if there is a leading entry in the final column, then the system is inconsistent. If there is no leading entry in the final column, then a solution can always be found using back substitution as above. The linear system $A'X = B'$ yields r (non-trivial) equations. For each equation we can write the variable x_i, that corresponds to the leading entry, in terms of the x_j having $j > i$. The $n - r$ variables that do not correspond to leading entries can be chosen as parameters. □

Note. 1. If $r = n$ (as in Example 6.3), the solution requires no parameters. That is, there is a unique solution.

2. If $r < n$, then there will be more than one solution. If $\mathbb{F}$ is infinite (eg, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$), there will be infinitely many solutions.

6.5 Exercises

50. Use Gaussian elimination to put the following matrices into row echelon form.

(a) $\begin{bmatrix} 1 & 1 & -8 & -14 \\ 3 & -4 & -3 & 0 \\ 2 & -1 & -7 & -10 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 5 & 7 & 6 \\ 0 & 0 & 5 & 2 & 4 \end{bmatrix}$

51. Solve the linear system whose augmented matrix can be reduced to the following row echelon form.

$$\left[\begin{array}{ccc|c} 1 & -3 & 4 & 7 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

52. Use Gaussian elimination to solve the following linear systems of equations:

(a) $\begin{aligned} x + z &= 0 \\ x + y &= 0 \\ y + z &= 0 \end{aligned}$ (b) $\begin{aligned} 2x + 3y &= 1 \\ x + \ y &= 1 \end{aligned}$ (c) $\begin{aligned} x - \ y + \ z &= 1 \\ 2x + \ y - \ z &= 3 \\ 3x + 2y - 3z &= 2 \end{aligned}$

53. Solve the systems of linear equations whose augmented matrices can be reduced to the following row echelon forms.

(a) $\left[\begin{array}{ccc|c} 1 & -3 & 7 & 1 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$ (b) $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$

54. Use Gaussian elimination to solve the following linear systems of equations:

(a) $$\begin{aligned} 4x - 2y &= 5 \\ -6x + 3y &= 1 \end{aligned}$$

(b) $$\begin{aligned} x - 4y &= 1 \\ -2x + 8y &= -2 \end{aligned}$$

(c) $$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1 \\ 2x_1 + 3x_2 + 3x_3 \phantom{{}+x_4} &= 1 \\ -x_1 - 2x_2 - 2x_3 + x_4 &= 0 \\ - x_2 - x_3 + 2x_4 &= 1 \end{aligned}$$

55. Determine the conditions on $a, b, c \in \mathbb{C}$ so that the system is consistent:

(a) $$\begin{aligned} x + 2y - 3z &= a \\ 3x - y + 2z &= b \\ x - 5y + 8z &= c \end{aligned}$$

(b) $$\begin{aligned} x + 2y + 4z &= a \\ 2x + 3y - z &= b \\ 3x + y + 2z &= c \end{aligned}$$

56. Determine the values of the constant $k \in \mathbb{C}$ for which the system has

(i) no solutions,

(ii) a unique solution,

(iii) an infinite number of solutions.

Find the solutions when they exist.

(a) $$\begin{aligned} 2x + 3y + z &= 11 \\ x + y + z &= 6 \\ 5x - y + 11z &= k \end{aligned}$$

(b) $$\begin{aligned} x_1 + x_3 &= 1 \\ x_2 + x_3 &= 2 \\ 2x_2 + kx_3 &= k \end{aligned}$$

(c) $$\begin{aligned} x + y + 2z &= 9 \\ x - y + z &= 2 \\ 4x + 2y + (k - 22)z &= k \end{aligned}$$

57. Determine the values for a, b and c for which the parabola $y = ax^2 + bx + c$ passes through the points:

(a) $(0, -3), (1, 0)$ and $(2, 5)$

(b) $(-1, 1), (1, 9)$ and $(2, 16)$

Extra material for lecture 6

▷ **Gaussian elimination**

Elementary Linear Algebra by Anton and Rorres, §1.2

▷ A **homogeneous linear system** is a system of linear equations in which each equation is equal to 0. Otherwise, the system is called **non-homogeneous**. We explore the relationship between the solution set of a non-homogeneous linear system and the solution set of the associated homogeneous linear system.
Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$. Define

$$S_{\text{hom}} = \{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = 0\} \quad \text{and} \quad S_{\text{inhom}} = \{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = B\}$$

Let $X_p \in S_{\text{inhom}}$ be a fixed solution of the non-homogeneous system. Show that

$$S_{\text{inhom}} = X_p + S_{\text{hom}}$$

(Notation: $X_p + S_{\text{hom}} = \{X_p + X \mid X \in S_{\text{hom}}\}$)

For example, the following linear system is inhomogeneous.

$$\begin{aligned} x + y + 3z &= 2 \\ x - y + 7z &= 4 \end{aligned}$$

A particular solution is $(x, y, z) = (3, -1, 0)$
The corresponding homogeneous system is

$$\begin{aligned} x + y + 3z &= 0 \\ x - y + 7z &= 0 \end{aligned}$$

The general solution to the homogeneous system is:

$$(x, y, z) = t(-5, 2, 1) \quad t \in \mathbb{R}$$

That is,

$$S_{\text{hom}} = \{t(-5, 2, 1) \mid t \in \mathbb{R}\}$$

The general solution to the inhomogeneous system is:

$$(x, y, z) = (3, -1, 0) + t(-5, 2, 1) \quad t \in \mathbb{R}$$

That is,

$$S_{\text{inhom}} = \{(3, 1, 0) + t(-5, 2, 1) \mid t \in \mathbb{R}\} = (3, 1, 0) + S_{\text{hom}}$$

LECTURE 7

Reduced row echelon form and matrix inverses

7.1 Reduced row echelon form

We further refine the simplified form of a matrix. For linear systems this corresponds to performing back substitution while still in matrix form. The new form has the advantage of being uniquely determined by the original matrix.

Definition 7.1: Reduced row echelon form (RREF)

A matrix is in **reduced row echelon form** if the following three conditions are satisfied.

1. It is in row echelon form.
2. Each leading entry is equal to 1 (sometimes called a **leading 1**).
3. In each column containing a leading 1, all other entries are zero.

Example 7.2. 1) $\begin{bmatrix} 1 & -2 & 3 & -4 & 5 \end{bmatrix}$, $\begin{bmatrix} 1 & 2+i & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 2+i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ are in RREF

2) $\begin{bmatrix} 1 & 0 & 1-i & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & i \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 & 2 & 4 \\ 0 & 1 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -4 & 9 \end{bmatrix}$ are not in RREF

Any matrix can be put into reduced row echelon form as follows.

Algorithm 7.3: Gauss-Jordan elimination (To put a matrix into RREF)

1. First use Gaussian elimination (Algorithm 6.1) to put the matrix in row echelon form.
2. Multiply rows by appropriate numbers (type 2 row ops) to create the leading 1′s.
3. Working from the bottom row upward, use row operations of type 3 to create zeros above the leading entries.

Example 7.4.

$$\begin{bmatrix} -2 & 0 & 6 & 8 & 14 \\ 1 & 0 & -5 & -8 & -13 \\ 2 & 0 & -2 & 0 & -2 \\ 2 & 0 & -5 & -6 & -11 \end{bmatrix} \xrightarrow[\substack{R3+R2 \\ R4+R1}]{R2+\frac{1}{2}R1} \begin{bmatrix} -2 & 0 & 6 & 8 & 14 \\ 0 & 0 & -2 & -4 & -6 \\ 0 & 0 & 4 & 8 & 12 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix} \xrightarrow[R4+\frac{1}{2}R2]{R3+2R2} \begin{bmatrix} -2 & 0 & 6 & 8 & 14 \\ 0 & 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[-\frac{1}{2}R2]{-\frac{1}{2}R1} \begin{bmatrix} 1 & 0 & -3 & -4 & -7 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R1+3R2} \begin{bmatrix} 1 & 0 & 0 & 2 & 2 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The final matrix is in reduced row echelon form.

Consider the linear system

$$\begin{aligned} -2x_1 + 6x_3 + 8x_4 &= 14 \\ x_1 - 5x_3 - 8x_4 &= -13 \\ 2x_1 - 2x_3 &= -2 \\ 2x_1 - 5x_3 - 6x_4 &= -11 \end{aligned}$$

The above row operations tell us that the set of solutions is the same as that for the system

$$\begin{aligned} x_1 + 2x_4 &= 14 \\ x_3 + 2x_4 &= 3 \end{aligned}$$

This gives

$$\begin{aligned} x_1 &= -2x_4 + 14 \\ x_3 &= -2x_4 + 3 \end{aligned}$$

We choose $s = x_2$ and $t = x_4$ as parameters (since the corresponding columns have no leading entry). The set of solutions is given by

$$\{(-2t + 14, s, -2t + 3, t) \mid s, t \in \mathbb{F}\}$$

7.2 Using row operations to find the inverse of a matrix

Our method for finding the inverse of a matrix is based on the following result.

Theorem 7.5

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. Then A is invertible if and only if $A \sim I_n$.

Proof. Suppose first that $A \sim I_n$. By Lemma 5.8 there is an invertible matrix $E \in \mathcal{M}_{n,n}(\mathbb{F})$ such that $EA = I_n$. Since E is invertible we have

$$EA = I_n \implies E^{-1}EA = E^{-1}I_n \implies A = E^{-1}$$

Therefore A is invertible (since E^{-1} is invertible) and $A^{-1} = E$.

Now suppose that A is invertible. We want to show that $A \sim I_n$. Let $R \in \mathcal{M}_{n,n}(\mathbb{F})$ be a matrix in reduced row echelon form such that $A \sim R$ (existence is ensured by Algorithm 7.3). By Lemma 5.8 there is an invertible matrix $E \in \mathcal{M}_{n,n}(\mathbb{F})$ such that $EA = R$. Since both E and A are invertible, R is invertible (Lemma 4.3). Since R is invertible it does not have a zero row (Lemma 4.4). From the definition of reduced row echelon form we can see that a square matrix that is in reduced row echelon form and does not have a zero row must be the identity matrix. Therefore, $R = I_n$ and hence $A \sim I_n$. □

Corollary 7.6

If $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is invertible, then it can be written as a product of elementary matrices. □

Exercise 58. Suppose that $A, B \in \mathcal{M}_{n,n}(\mathbb{F})$ are such that $AB = I_n$. Show that A is invertible and that $A^{-1} = B$. (Hint: as in the above proof, let $E \in \mathcal{M}_{n,n}(\mathbb{F})$ be invertible such that $R = EA$ is in reduced row echelon form. Show that $RB = E$ and that therefore R does not have a row of zeros. Deduce that $R = I_n$.)

Using the idea of the above theorem (and its proof) we have a way of finding the inverse of a square matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$. First we use row operations to put A into reduced row echelon form R. If $R \neq I_n$, then A is not invertible. If $R = I_n$, then A is invertible and its inverse is the matrix E. A convenient way of doing this is the following.

Algorithm 7.7: Finding the inverse of a square matrix (if it exists)

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$.

1. Form the matrix $[A \,|\, I_n] \in \mathcal{M}_{n,2n}(\mathbb{F})$
2. Use row operations to put the matrix into reduced row echelon form
$$[A \,|\, I_n] \sim [R \,|\, B]$$
(where $R, B \in \mathcal{M}_{n.n}(\mathbb{F})$)
3. If $R \neq I_n$, then A in *not* invertible.
If $R = I_n$, then A is invertible and $A^{-1} = B$

Remark. If we have row operations ρ_i (with corresponding elementary matrices E_i) such that

$$[A|I_n] \xrightarrow{\rho_1} [A_1 \,|\, B_1] \xrightarrow{\rho_2} [A_2 \,|\, B_2] \xrightarrow{\rho_3} \cdots \xrightarrow{\rho_k} [I_n \,|\, B]$$

then we have $A^{-1} = B = E_k \dots E_2 E_1$.

Example 7.8. Let's find the inverse of the matrix $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{C})$.

$$[A|I_3] = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 3 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R2+R1} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 3 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{R3-R2} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

At this point we know that the matrix A is invertible.

$$\xrightarrow[R2-2R3]{R1-R3} \begin{bmatrix} 1 & 2 & 0 & 2 & 1 & -1 \\ 0 & 1 & 0 & 3 & 3 & -2 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

$$\xrightarrow{R1-2R2} \begin{bmatrix} 1 & 0 & 0 & -4 & -5 & 3 \\ 0 & 1 & 0 & 3 & 3 & -2 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

Therefore $A^{-1} = \begin{bmatrix} -4 & -5 & 3 \\ 3 & 3 & -2 \\ -1 & -1 & 1 \end{bmatrix}$. In addition, from the row operations used we have that

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

7.3 Exercises

59. For each of the following linear systems, use row reduction to decide whether the system has

    (i) no solution, (ii) a unique solution, (iii) more than one solution.

    Solve the systems where possible.

    (a)
$$\begin{aligned} 3x - 2y + 4z &= 3\\ x - y + z &= 7\\ 4x - 3y + 5z &= 1 \end{aligned}$$

    (b)
$$\begin{aligned} x + 2y - z &= -1\\ 2x + 7y - z &= 3\\ -3x - 12y + z &= 0 \end{aligned}$$

    (c)
$$\begin{aligned} 3x - 4y + z &= 2\\ -5x + 6y + 10z &= 7\\ 8x - 10y - 9z &= -5 \end{aligned}$$

    (d)
$$\begin{aligned} 2x - 3y + 5z &= 10\\ 4x + 7y - 2z &= -5\\ 2x - 4y + 25z &= 31 \end{aligned}$$

60. Using row reduction, find the set of solutions to the following system of equations:

$$\begin{aligned} 2x_1 + x_2 + 3x_3 + x_4 &= 3\\ x_1 + x_2 + x_3 - x_4 &= 6\\ x_1 - x_2 + 3x_3 + 5x_4 &= -12\\ 4x_1 + x_2 + 7x_3 + 5x_4 &= -3 \end{aligned}$$

61. Determine the values of $k \in \mathbb{C}$ for which the system of equations has

    (i) no solution,

    (ii) a unique solution,

    (iii) more than one solution.

    (a)
$$\begin{aligned} kx + y + z &= 1\\ x + ky + z &= 1\\ x + y + kz &= 1 \end{aligned}$$

    (b)
$$\begin{aligned} 2x + (k-4)y + (3-k)z &= 1\\ 2y + (k-3)z &= 2\\ x - 2y + z &= 1 \end{aligned}$$

    (c)
$$\begin{aligned} x + 2y + kz &= 1\\ 2x + ky + 8z &= 3 \end{aligned}$$

    (d)
$$\begin{aligned} x - 3z &= -3\\ 2x + ky - z &= -2\\ x + 2y + kz &= 1 \end{aligned}$$

    Find the solutions when they exist.

62. Determine the conditions on $a, b, c \in \mathbb{C}$ so that the system has a solution:

    (a)
$$\begin{aligned} x + 2y - 3z &= a\\ 3x - y + 2z &= b\\ x - 5y + 8z &= c \end{aligned}$$

    (b)
$$\begin{aligned} x - 2y + 4z &= a\\ 2x + 3y - z &= b\\ 3x + y + 2z &= c \end{aligned}$$

    Find the solutions when they exist.

63. The equation of an arbitrary circle in the x-y plane can be written in the form

$$x^2 + y^2 + ax + by + c = 0$$

    where a, b, c are real constants. Find the equation of the unique circle that passes through the three points $(-2, 7), (-4, 5), (4, -3)$.

64. A traveller who just returned from Europe spent the following amounts.

    For housing: \$30/day in England, \$20/day in France, \$20/day in Spain

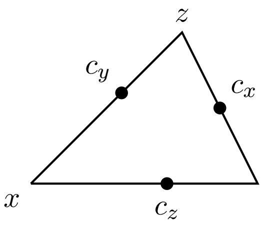
    For food: \$20/day in England, \$30/day in France, \$20/day in Spain

    For incidental expenses: \$10/day in each country.

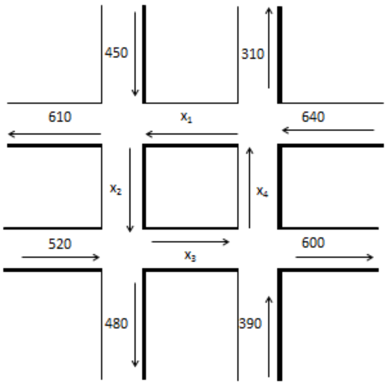
    The traveller's records of the trip indicate a total of \$340 spent for housing, \$320 for food, \$140 for incidental expenses while travelling in these countries. Calculate the number of days spent in each country or show that the records must be incorrect.

65. Frank's, Dave's and Phil's ages are not known, but are related as follows. The sum of Dave's and Phil's ages is 13 more than Frank's. Frank's age plus Phil's age is 19 more than Dave's age. If the sum of their ages is 71, how old are Frank, Dave and Phil?

66. Consider a triangle with vertices x, y and z as shown. Show that there exist unique points c_x, c_y, c_z (on the sides indicated) with the property that:
    $d(x, c_y) = d(x, c_z)$ and $d(y, c_x) = d(y, c_z)$ and $d(z, c_x) = d(z, c_y)$.



67. Consider the traffic flow diagram on the right. The number of cars x_1, x_2, x_3 and x_4 entering an intersection must equal the number of cars leaving the intersection.
    Determine the *smallest nonnegative* values of x_1, x_2, x_3 and x_4.



68. Following the algorithm described in lectures, reduce the following matrices to reduced row echelon form. Keep a record of the elementary row operations you use.

    (a) $\begin{bmatrix} 4 & -8 & 16 \\ 1 & -3 & 6 \\ 2 & 1 & 1 \end{bmatrix}$

    (b) $\begin{bmatrix} 2+i & 2+i & 5 & 6+i \\ 1-2i & 1-2i & -2+i & 2-i \end{bmatrix}$

    (c) $\begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 2 & 2 \\ 0 & 2 \end{bmatrix}$

    (d) $\begin{bmatrix} 0 & 2 & 1 & 4 \\ 0 & 0 & 2 & 6 \\ 1 & 0 & -3 & 2 \end{bmatrix}$

    (e) $\begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 1 \\ 3 & 6 & 1 & 1 \end{bmatrix}$

    (f) $\begin{bmatrix} 0 & 0 & 2 & 7 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -4 & 5 \\ -2 & 2 & -5 & 4 \end{bmatrix}$

69. For each of the following matrices, decide whether or not the matrix is invertible and, if it is, find the inverse.

    (a) $\begin{bmatrix} 2 & 0 \\ -3 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$

70. Consider the matrix $A = \begin{bmatrix} 1 & 0 \\ -5 & 2 \end{bmatrix}$.

    (a) Write A^{-1} as a product of two elementary matrices.

(b) Write A as a product of two elementary matrices.

71. (a) Find the inverse of the matrix

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 2 & 3 & 2 & 2 \\ 4 & 2 & 4 & 3 \\ 6 & 3 & 3 & 5 \end{bmatrix}$$

(b) Check your answer by matrix multiplication.

(c) Write A as a product of elementary matrices.

(d) Use your answer to part (a) to solve the system

$$\begin{aligned} 2x + y + z + w &= 3 \\ 2x + 3y + 2z + 2w &= 5 \\ 4x + 2y + 4z + 3w &= 6 \\ 6x + 3y + 3z + 5w &= 9 \end{aligned}$$

Extra material for lecture 7

- **reduced row echelon form**

  Elementary Linear Algebra by Anton and Rorres, §1.2, §1.5

- **calculating matrix inverses**

  Elementary Linear Algebra by Anton and Rorres, §1.5

LECTURE 8

Rank and determinant of a matrix

8.1 Uniqueness of the reduced row echelon form

Proposition 8.1

The reduced row echelon form of a matrix is unique.

The proof is presented (for completeness) in an appendix at the end of this lecture.

Note that this is the same as saying that if R and S are two matrices each in RREF and if $R \sim S$, then we must have $R = S$.

8.2 Rank of a matrix

Knowing that the reduced row echelon form of A is unique enables us to make the following definition.

Definition 8.2

The **rank** of a matrix is the number of non-zero rows in its reduced row echelon form. We denote the rank of a matrix A by $\operatorname{rank}(A)$.

Remark.

1. Although the row echelon form is not unique, it has the same number of non-zero rows as its reduced row echelon form.
2. If two matrices are row equivalent, then they have the same rank.
3. If A has m rows, then $\operatorname{rank}(A) \leqslant m$.
4. If a square matrix $R \in \mathcal{M}_{n,n}(\mathbb{F})$ has rank n and is in reduced row echelon form, then $R = I_n$.

Proposition 8.3

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. Then A is invertible if and only if $\operatorname{rank}(A) = n$.

Proof. Assume first that A is invertible. Then $A \sim I_n$ by Theorem 7.5. Therefore $\operatorname{rank}(A) = \operatorname{rank}(I_n) = n$.

For the converse, assume now that $\operatorname{rank}(A) = n$. Let R be the reduced row echelon form of A. Then $\operatorname{rank}(R) = \operatorname{rank}(A) = n$. Therefore R is an $n \times n$ matrix, it is in reduced row echelon form and has no zero rows (since it has rank n). It must be the case that $R = I_n$ because I_n is the only $n \times n$ matrix that is in RREF and has rank n. Therefore A is invertible by Theorem 7.5. □

8.3 Determinant

Recall that for a 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ the quantity $ad - bc$ can be used to determine whether or not the matrix is invertible. It is also equal to the area of the parallelogram defined by the vectors $(a, b), (c, d) \in \mathbb{R}^2$. The determinant is a generalisation of this quantity to $n \times n$ matrices.

The determinant is a number (i.e., element of $\mathbb{F}$) associated to a square matrix (in $\mathcal{M}_{n,n}(\mathbb{F})$). It gives a lot of information about the matrix. For example, a square matrix is invertible precisely when its determinant is non-zero.

Rather than beginning with an explicit formula for the determinant, we list its important properties. The first three of which are, in fact, enough to determine the value of the determinant.

Properties of determinants

Given a matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$, the determinant of A is a number $\det(A) \in \mathbb{F}$ that satisfies:

1. $\det(I_n) = 1$
2. If B is obtained from A by swapping two rows, then $\det(B) = -\det(A)$.
3. The determinant depends linearly on the first row, that is:

(a) $$\det \begin{bmatrix} kA_{11} & kA_{12} & \cdots & kA_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} = k \det \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$

(b) $$\det \begin{bmatrix} A_{11} + A'_{11} & A_{12} + A'_{12} & \cdots & A_{1n} + A'_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} = \det \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} + \det \begin{bmatrix} A'_{11} & A'_{12} & \cdots & A'_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$

Note. The above properties tell us exactly what the effect of a row operation is on the determinant.

Example 8.4.

$$\begin{aligned} \det \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 1 & 0 \end{bmatrix} &= 2 \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 1 & 0 \end{bmatrix} = -2 \det \begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = -6 \det \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ &= 6 \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = -6 \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = -6 \end{aligned}$$

The following further properties can all be derived from the first three[*].

Properties of determinants

4. If two rows are equal, then the determinant is zero
5. If B is obtained from A by applying a row operation of the third kind (i.e., replacing a row by itself plus a multiple of another row), then $\det(B) = \det(A)$
6. If A has a row of zeros, then $\det(A) = 0$

[*] In fact, for some fields (those of characteristic 2) property 4 does not follow from the first 3.

7. $\det \begin{bmatrix} d_1 & * & \cdots & * \\ 0 & d_2 & \cdots & * \\ & & \ddots & \\ 0 & \cdots & 0 & d_n \end{bmatrix} = d_1 d_2 \dots d_n$

8. $\det(A) = 0$ if and only if A is singular

9. $\det(AB) = \det(A)\det(B)$

10. $\det(A^T) = \det(A)$

Note. Another common notation for the determinant of a matrix A which we will sometimes use is to write $|A|$ in place of $\det(A)$

Example 8.5. We can use row operations to calculate the determinant of a matrix by putting it into row echelon form.

$$\begin{vmatrix} 1 & 2 & -2 & 0 \\ 2 & 3 & -4 & 1 \\ -1 & -2 & 0 & 2 \\ 0 & 2 & 5 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -2 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & 2 \\ 0 & 2 & 5 & 3 \end{vmatrix} \qquad \text{(property 5: } R_2 - 2R_1, R_3 + R_1\text{)}$$

$$= \begin{vmatrix} 1 & 2 & -2 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 5 & 5 \end{vmatrix} \qquad \text{(property 5: } R_4 + 2R_2\text{)}$$

$$= \begin{vmatrix} 1 & 2 & -2 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 0 & 10 \end{vmatrix} \qquad \text{(property 5: } R_4 + \tfrac{5}{2}R_3\text{)}$$

$$= 1 \times (-1) \times (-2) \times 10 = 20 \qquad \text{(property 7)}$$

8.4 Exercises

72. Find the rank of the matrices given in Exercise 50.

73. Find the rank of the matrices given in Exercise 68.

74. (a) Determine the rank of the following matrices: (i) $\begin{bmatrix} 1 & 2 & -1 \\ 3 & -6 & 2 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

    (b) Determine the rank of the matrix $\begin{bmatrix} 1 & 1 & k \\ 1 & k & 1 \\ k & 1 & 1 \end{bmatrix}$ when: (i) $k = 1$, (ii) $k = -2$, (iii) $k \notin \{1, -2\}$.

75. Use row operations to calculate the determinants of the following matrices:

    (a) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 7 \\ 1 & 4 & 13 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 & 1 \\ 3 & 0 & -1 \\ 4 & 5 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} \frac{1}{3} & \frac{3}{5} & \frac{2}{5} \\ \frac{3}{8} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{2} \end{bmatrix}$

76. Let $a, b, c, d, e, f, g, h, i \in \mathbb{F}$ be such that

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 1$$

Find the following determinants:

(a) $\begin{vmatrix} a & b & c \\ g & h & i \\ d & e & f \end{vmatrix}$

(b) $\begin{vmatrix} a & -b & c \\ d & -e & f \\ g & -h & i \end{vmatrix}$

(c) $\begin{vmatrix} d & e & f \\ 3g & 3h & 3i \\ a & b & c \end{vmatrix}$

(d) $\begin{vmatrix} 2a & 2b & 2c \\ 2d & 2e & 2f \\ 2g & 2h & 2i \end{vmatrix}$

(e) $\begin{vmatrix} a & b & c \\ d+a & e+b & f+c \\ g-2a & h-2b & i-2c \end{vmatrix}$

77. A matrix P is called **idempotent** if $P^2 = P$.

(a) Show that if P is idempotent and $\det(P) \neq 0$, then $P = I$.

(b) Give an example of an idempotent matrix that is neither a zero matrix nor an identity matrix.

Extra material for lecture 8

▷ **Proof of uniqueness of RREF**

Proof. We need to show that for any $A \in \mathcal{M}_{m,n}(\mathbb{F})$ if $A \sim R_1$ and $A \sim R_2$, and both R_1 and R_2 are in reduced row echelon form, then $R_1 = R_2$.

The idea is to show that if $R_1 \neq R_2$ (and they are both in RREF) then there exists $X \in \mathcal{M}_{n,1}(\mathbb{F})$ such that $R_1X \neq 0$ and $R_2X = 0$. This would establish that $R_1 \neq R_2 \implies R_1 \not\sim R_2$ (as required).

Suppose, for a contradiction, that $R_1 \neq R_2$. Let $j \in \{1,\dots,n\}$ be minimal such that j-th column of R_1 is not equal to the j-th column of R_2. Form a new matrix S_1 from R_1 as follows. Drop all columns to the right of the j-th column and drop all the columns on its left that do not contain a leading entry. Define S_2 as the matrix similarly obtained from R_2. For example

$$R_1 = \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 1 & 2 & 0 & 3 & 6 \\ 0 & 0 & 1 & 4 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad S_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \qquad S_2 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that S_1 and S_2 will be in reduced row echelon form and $S_1 \sim S_2$. Suppose first that the j-th column of R_1 (which is the last column of S_1) contains a leading entry.

Then we have

$$S_1 = \left[\begin{array}{c|c} \begin{matrix} \\ I_k \\ \\ \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \begin{matrix} \\ \\ 0 \\ \\ \end{matrix} & \begin{matrix} 1 \\ 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right] \qquad S_2 = \left[\begin{array}{c|c} \begin{matrix} \\ I_k \\ \\ \end{matrix} & \begin{matrix} b_1 \\ \vdots \\ b_k \end{matrix} \\ \hline \begin{matrix} \\ \\ 0 \\ \\ \end{matrix} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right]$$

But then $S_1 \begin{bmatrix} b_1 & \cdots & b_k & -1 \end{bmatrix}^T \neq 0$ whereas $S_2 \begin{bmatrix} b_1 & \cdots & b_k & -1 \end{bmatrix}^T = 0$ which is not possible since $S_1 \sim S_2$. The same argument applies in the case in which the j-th column of R_2 contains a leading entry. If neither the j-th column of R_1 nor the j-th column of R_2 contains a leading entry we have

$$S_1 = \left[\begin{array}{c|c} \begin{matrix} \\ I_k \\ \\ \end{matrix} & \begin{matrix} a_1 \\ \vdots \\ a_k \end{matrix} \\ \hline \begin{matrix} \\ 0 \\ \\ \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right] \qquad S_2 = \left[\begin{array}{c|c} \begin{matrix} \\ I_k \\ \\ \end{matrix} & \begin{matrix} b_1 \\ \vdots \\ b_k \end{matrix} \\ \hline \begin{matrix} \\ 0 \\ \\ \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right]$$

Considered as the augmented matrix of a linear system, each has a unique solution. The solutions must be equal since $S_1 \sim S_2$. Therefore $a_i = b_i$ for all i. But then $S_1 = S_2$, which contradicts the assumption that their final columns are different.

□

Determinants (continued)

9.1 Derivation of properties

We claim that each of the properties given for the determinant can be derived from the first 3.

Exercise 78. Show that properties 4,5, and 6 follow from the first three properties. (For property 4 you will need to assume that the field is such that $1+1 \neq 0$.)

We will show that properties 7,8 9, and 10 follow from the first six.

Lemma 9.1

Let $A, B \in \mathcal{M}_{n,n}(\mathbb{F})$. Suppose B is obtained from A by a sequence of elementary row operations. Then there is a $k \in \mathbb{F} \setminus \{0\}$ such that $\det(B) = k \det(A)$. Moreover, if D is obtained from C using the same sequence of row operations, then $\det(D) = k \det(C)$.

Proof. (in lecture) □

Lemma 9.2: Property 7

If $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is in triangular form, then $\det(A) = A_{11}A_{22}\dots A_{nn}$.

Proof. Suppose that A is in upper triangular form. That is, we have

$$A = \begin{bmatrix} d_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & d_n \end{bmatrix} \in \mathcal{M}_{n,n}(\mathbb{F})$$

If the diagonal entries d_i are all non-zero, then we have

$$\begin{aligned} |A| = \begin{vmatrix} d_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & d_n \end{vmatrix} &= d_1 d_2 \cdots d_n \begin{vmatrix} 1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & 1 \end{vmatrix} && \text{(by properties 2, 3a)} \\ &= d_1 d_2 \cdots d_n \begin{vmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{vmatrix} = d_1 d_2 \cdots d_n && \text{(by properties 5, 1)} \end{aligned}$$

If one (or more) of the d_i is equal to 0, then the REF of A will have a row of zeros and therefore $\det(A) = 0$.

□

Lemma 9.3: Property 8

$\det(A) = 0$ if and only if A is singular

Proof. We first show that $\det(A) = 0 \implies A$ is singular (by establishing the contrapositive).

$$\begin{aligned} A \text{ invertible } &\implies A \sim I \qquad \text{(Theorem 7.5)} \\ &\implies \det(A) = k\det(I) \quad \text{(some } k \neq 0\text{, Lemma 9.1)} \\ &\implies \det(A) \neq 0 \end{aligned}$$

For the converse, suppose that A is singular. Let R be the matrix in RREF such that $R \sim A$. Because A is singular, $R \neq I$ (Theorem 7.5) and R therefore has a row of zeros. Since R has a row of zeros, $\det(R) = 0$ and therefore $\det(A) = 0$ (Lemma 9.1).

□

Lemma 9.4: Property 9

$\det(AB) = \det(A)\det(B)$

Proof. Let R be the RREF of A. Then $\det(R) = k\det(A)$ for some $k \neq 0$. We also have $\det(RB) = k\det(AB)$ with *the same* k.

Assume first that A is invertible. Then $R = I$ and we have

$$\begin{aligned} \det(A)\det(B) &= k\det(A)\det(AB) && \text{(since } \det(B) = k\det(AB)\text{)} \\ &= \det(AB) && \text{(since } k\det(A) = \det(R) = 1\text{)} \end{aligned}$$

On the other hand, if A is singular, then R has a row of zeros. Therefore RB has a row of zeros, and so $\det(RB) = 0$. We have

$$\det(AB) = \frac{1}{k}\det(RB) = 0$$

and

$$\det(A)\det(B) = 0\det(B) = 0$$

□

Lemma 9.5: Property 10

$\det(A^T) = \det(A)$

Proof. If A is singular, then A^T is also singular (Exercise 38) and therefore $\det(A^T) = 0 = \det(A)$.

We can assume therefore that A is invertible. It is enough to show that $\det(E^T) = \det(E)$ for all elementary matrices E since any invertible matrix can be written as a product of elementary matrices (Corollary 7.6) and we then have

$$A = E_1E_2\dots E_k \implies \det(A) = \det(E_1)\det(E_2)\dots\det(E_k)$$

and

$$\begin{aligned} A^T = E_k^T\dots E_2^TE_1^T \implies \det(A^T) &= \det(E_k^T)\dots\det(E_2^T)\det(E_1^T) \\ &= \det(E_k)\dots\det(E_2)\det(E_1) \\ &= \det(E_1)\det(E_2)\dots\det(E_k) \end{aligned}$$

It remains to check that $\det(E^T) = \det(E)$ for elementary matrices. Note that if E is an elementary matrix corresponding to a row swap or to multiplying a row by a constant, then $E = E^T$. If E corresponds to the third kind of row operation, then both E and E^T are triangular and have all diagonal entries equal to one. □

9.2 Cofactor expansion

The following gives a version of a (recursive) formula for $\det(A)$ that can often be useful.

Lemma 9.6: Cofactor expansion along the i-th row

Let $n \geqslant 2$, $A \in \mathcal{M}_{n,n}(\mathbb{F})$ and let $i \in \{1,\dots,n\}$. Then

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(A(i,j))$$

where $A(i,j)$ denotes the $(n-1)\times(n-1)$ matrix obtained from A by deleting the i-th row and the j-th column.

Remark. This is often referred to as 'cofactor expansion along the i-th row'.

Example 9.7.

$$\begin{aligned}\begin{vmatrix} 1 & 2 & -2 \\ 2 & 3 & -4 \\ -1 & -2 & 0 \end{vmatrix} &= (-1)^4 \times (-1) \times \begin{vmatrix} 2 & -2 \\ 3 & -4 \end{vmatrix} + (-1)^5 \times (-2) \times \begin{vmatrix} 1 & -2 \\ 2 & -4 \end{vmatrix} + (-1)^6 \times 0 \times \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} \\ &= 1 \times (-1) \times (-2) + (-1) \times (-2) \times 0 + 1 \times 0 \times (-1) \\ &= 2\end{aligned}$$

Remark. The value of $(-1)^{i+j}$ is $+1$ for $i = j = 1$ (i.e., top left of the matrix) and then alternates between -1 and $+1$ as we move one entry vertically or horizontally.

9.3 Exercises

79. Use cofactor expansion to calculate the determinant of the matrix in Example 8.5.

80. Evaluate the determinant of the following matrices using cofactor expansion (Lemma 9.6):

(a) $\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 1 & 1 \\ 3 & 0 & -1 \\ 4 & 5 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 4 & 2 \\ 1 & 5 & 1 \\ 3 & -7 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 3 & 4 & 5 \\ 0 & 3 & 4 & 5 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 5 \end{bmatrix}$

(e) $\begin{bmatrix} a & ab \\ b & a^2+b^2 \end{bmatrix}$ (where $a, b \in \mathbb{C}$)

81. Evaluate the determinants of the following matrices. For what values of the variables ($x, \lambda, k \in \mathbb{C}$) are the matrices invertible?

(a) $\begin{bmatrix} x & 2x & -3x \\ x & x-1 & -3 \\ 0 & 0 & 2x-1 \end{bmatrix}$

(b) $\begin{bmatrix} \lambda-1 & 0 & 0 & 0 \\ 2 & 0 & \lambda+1 & 0 \\ 1 & \lambda-2 & 0 & 0 \\ 2 & 3 & 9 & \lambda+2 \end{bmatrix}$

(c) $\begin{bmatrix} k & k+1 & k+2 \\ k+3 & k+4 & k+5 \\ k+6 & k+7 & k+8 \end{bmatrix}$

82. (a) Determine the values of the parameter λ for which $\det(A - \lambda I) = 0$ when

i) $A = \begin{bmatrix} 4 & 2 \\ -3 & -1 \end{bmatrix}$ ii) $A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}$

(b) Find all solutions X of the linear system

$$(A - \lambda I)X = \mathbf{0}$$

for each A and each λ you found in part (a).

Extra material for lecture 9

- ▷ Think about why the properties listed *uniquely* determine the value of $\det(A)$.
- ▷ Try to prove that the cofactor expansion satisfies properties 1–3.
- ▷ An alternative way of defining (or calculating) the determinant of a matrix $A \in \mathcal{M}_{n,n}$ uses permutations of the set $\{1, \dots, n\}$.

  $$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) A_{1,\sigma(1)} A_{2,\sigma(2)} \dots A_{n,\sigma(n)}$$

  Here S_n is the set of all bijections $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$.

  See, for example, Axler (Chapter 10).

Fields

Recall that elements of $\mathbb{R}^3$ can be added together and multiplied by scalars. For example

$$\begin{aligned}(1,2,3)+(-2,3,-7)&=(-1,5,-4)\\ -7(-2,3,-7)&=(14,-21,49)\end{aligned}$$

Thinking about the properties of these operations leads to the definition of a vector space, the central topic of this subject. A vector space is a generalisation of the vector structure of $\mathbb{R}^3$ in which we have a set of 'vectors' together with a version of vector addition and scalar multiplication.

The first thing we will look at is what we can use as 'scalars'.

10.1 Fields

Examples of fields that we alerady know are $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. They each have two operations (addition and multiplication) and the two operations satisfy certain familiar properties. For example, every non-zero element has a multiplicative inverse. Before listing the defining properties we recall the meaning of 'associative' and 'commutative' for binary operations.

Definition 10.1

A **binary operation** on a set S is a function $S \times S \to S$. We often use 'infix' notation for binary operations. For example, we write $a+b$ rather than $+(a,b)$ for the binary operation of addition. A binary operation $\diamond : S \times S \to S$ is called:

1. **associative** if $\forall a,b,c \in S,\ (a \diamond b) \diamond c = a \diamond (b \diamond c)$
2. **commutative** if $\forall a,b \in S, a \diamond b = b \diamond a$

Example 10.2. The binary operations of addition and multiplication on $\mathbb{R}$ are associative and commutative. Subtraction gives a binary operation on $\mathbb{R}$ that is *not* associative since, for example, $(1-2)-3 \neq 1-(2-3)$. It is also not commutative since, for example, $1-2 \neq 2-1$. Matrix multiplication gives a binary operation on $\mathcal{M}_{2,2}(\mathbb{R})$ that is associative but not commutative.

Definition 10.3

A **field** is a set $\mathbb{F}$ together with two binary operations, $+$ and $\times$ on F satisfying the following properties:

A1) $\forall\, a,b \in \mathbb{F},\quad a+b=b+a$ (addition is commutative)

A2) $\forall\, a,b,c \in \mathbb{F},\quad a+(b+c)=(a+b)+c$ (addition is associative)

A3) $\exists\, 0 \in \mathbb{F}\, \forall\, a \in \mathbb{F},\quad a+0=a$ (additive identity)

A4) $\forall\, a \in \mathbb{F}\, \exists\, (-a) \in \mathbb{F},\quad a+(-a)=0$ (additive inverses)

M1) $\forall\, a,b \in \mathbb{F},\quad a \times b = b \times a$ (multiplication is commutative)

M2) $\forall\, a,b,c \in \mathbb{F},\quad a \times (b \times c) = (a \times b) \times c$ (multiplication is associative)

M3) $\exists\, 1 \in \mathbb{F} \setminus \{0\}\ \forall\, a \in \mathbb{F}, \quad a \times 1 = a$ (multiplicative identity)

M4) $\forall\, a \in \mathbb{F} \setminus \{0\}\ \exists\, (a^{-1}) \in \mathbb{F}, \quad a \times a^{-1} = 1$ (multiplicative inverses)

D) $\forall\, a, b, c \in \mathbb{F}, \quad a \times (b + c) = (a \times b) + (a \times c)$ (distributivity)

Example 10.4.
1. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ (with the usual operations) are fields
2. $\mathbb{Z}$ (with the usual operations) is not a field. It fails to satisfy axiom M4.
3. $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} \subset \mathbb{R}$ is a field (using the usual operations on $\mathbb{R}$)
4. $\mathcal{M}_{2,2}(\mathbb{R})$ with the usual matrix operations is not a field. It fails to satisfy axioms M1 and M4.

Exercise 83. Let $\mathbb{F}$ be a field. Show that:

a) $\forall x \in \mathbb{F}, \quad 0 \times x = 0$

b) $\forall x \in \mathbb{F}, \quad (-1) \times x = -x$

(Justify every step of your proofs using the axioms in the definition of a field.)

Remark. As is common when writing multiplication in $\mathbb{R}$ or $\mathbb{C}$, we will often write ab in place of $a \times b$.

Example 10.5. Here are two fields that each have only a finite number of elements.

The field $\mathbb{F}_2$ is a field having two elements $\mathbb{F}_2 = \{[0], [1]\}$ with operations given in the tables on the right.

$(\mathbb{F}_2, +)$

+	**[0]**	**[1]**
[0]	[0]	[1]
[1]	[1]	[0]

$(\mathbb{F}_2, \times)$

×	**[0]**	**[1]**
[0]	[0]	[0]
[1]	[0]	[1]

The field $\mathbb{F}_3$ is defined by $\mathbb{F}_3 = \{[0], [1], [2]\}$ and operations given on the right.

$(\mathbb{F}_3, +)$

+	**[0]**	**[1]**	**[2]**
[0]	[0]	[1]	[2]
[1]	[1]	[2]	[0]
[2]	[2]	[0]	[1]

$(\mathbb{F}_3, \times)$

×	**[0]**	**[1]**	**[2]**
[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]
[2]	[0]	[2]	[1]

Given any prime $p \in \mathbb{N}$ there is a field having p elements. It can be constructed as follows. We label the elements as $[0]$, $[1]$, ..., $[p-1]$, that is, $\mathbb{F}_p = \{[0], [1], \ldots, [p-1]\}$. The operations are defined by

$$[a] + [b] = [a + b]$$
$$[a] \times [b] = [a \times b]$$

where $[a+b]$ is defined to be the the element $[k] \in \mathbb{F}_p$ given by the condition that p divides $(a+b)-k$. In other words, we add as usual in $\mathbb{Z}$, but then add a multiple of p until the result lies in $\{1, 2, \ldots, p-1\}$. Similarly, the element $[a \times b]$ is defined to be the element $[k] \in \mathbb{F}_p$ given by the condition that p divides $(a \times b) - k$.

Remark. It's common to write the elements of $\mathbb{F}_p$ simply as $\{0, 1, 2, \ldots, p-1\}$ rather than $\{[0], [1], [2], \ldots, [p-1]\}$. Another notation is $\{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{(p-1)}\}$. Be careful, $\mathbb{F}_p$ is not a 'subfield' of $\mathbb{R}$.

10.2 Exercises

84. Write down the addition and multiplication tables for $\mathbb{F}_5$ and $\mathbb{F}_7$.

85. Let $\mathbb{Z}_4 = \{[0], [1], [2], [3]\}$ equipped with the addition and multiplication given in the tables.

    Show that $\mathbb{Z}_4$ is not a field.

$(\mathbb{Z}_4, +)$				
+	**[0]**	**[1]**	**[2]**	**[3]**
[0]	[0]	[1]	[2]	[3]
[1]	[1]	[2]	[3]	[0]
[2]	[2]	[3]	[0]	[1]
[3]	[3]	[0]	[1]	[2]

$(\mathbb{Z}_4, \times)$				
×	**[0]**	**[1]**	**[2]**	**[3]**
[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]
[2]	[0]	[2]	[0]	[2]
[3]	[0]	[3]	[2]	[1]

86. Show that

$$\left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \right\} \subset \mathcal{M}_{2,2}(\mathbb{F}_2)$$

equipped with the usual matrix operations is a field. That is, check that all the axioms are satisfied. (It's important to realise that the entries in the matrices are from $\mathbb{F}_2$. For example. $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$)

87. Let $A \in \mathcal{M}_{3,3}(\mathbb{F}_3)$ be given by $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$. (Note that the entries are elements of $\mathbb{F}_3$.)

    (a) Find the reduced row echelon form of A.

    (b) Find the determinant of A.

    (c) Find the inverse of A, if it exists.

88. Let $\mathbb{F}$ be a field. Show that $\forall a, b \in \mathbb{F},\ ab = 0 \implies (a = 0 \vee b = 0)$

89. Let $\mathbb{F}$ be a field and $a \in \mathbb{F} \setminus \{0\}$. Show that the function $L : \mathbb{F} \to \mathbb{F}$ given by $L(x) = ax$ is a bijection.

90. Let $\mathbb{F} = \{0, 1, a, b\}$ be a field having four elements. Show that

    (a) $ab = 1$

    (b) $a^2 = b$

    (c) $a^3 = 1$

    (d) $1 + 1 = 0$ (This is harder and just for fun. Feel free to skip it.)

Extra material for lecture 10

- ▷ **Related algebraic structures**

  A set equipped with a single operation that satisfies A2, A3, and A4 is called a **group**. If it also satisfies A1, it is called an **abelian group**. You will see groups in the subject *MAST20022 Group Theory and Linear Algebra.*

  If we drop axiom M4 (but keep all other parts from the definition of a field) we get what is called a (commutative) **ring**. Examples of rings that are not fields are: $\mathbb{Z}$ and $\mathbb{R}[X]$ (where each is equipped with the usual addition and multiplication). Rings are covered in the subject *MAST30005 Algebra*.

- ▷ More reading about fields and related structures:

  A first course in abstract algebra by John Fraleigh

  Algebra by Michael Artin

  Concrete abstract algebra : from numbers to Gröbner bases by Niels Lauritzen

LECTURE **11**

Vector spaces

11.1 Vector spaces

Consideration of the fundamental properties of 'vectors' in $\mathbb{R}^2$ leads us to the following definition of a 'vector space'. It consists of a field (often called the scalars) and another set whose elements will be called vectors. In addition there are two binary operations: 'scalar multiplication' and 'vector addition'.[*]

Definition 11.1: Vector space

Let $\mathbb{F}$ be a field. A **vector space over** $\mathbb{F}$ (also called an $\mathbb{F}$-vector space, or just a vector space) consists of a non-empty set V together with two binary operations:

vector addition: $V \times V \to V$ (with the image of (u, v) will be denoted $u + v$)

scalar multiplication: $\mathbb{F} \times V \to V$ (with the image of (α, v) being denoted αv).

These are required to satisfy the following axioms:

1) $\forall\, u, v \in V,\quad u + v = v + u$

2) $\forall\, u, v, w \in V,\quad u+(v+w) = (u+v)+w$

3) $\exists\, \vec{0} \in V\ \forall v \in V,\quad v + \vec{0} = v$

4) $\forall\, v \in V\ \exists(-v) \in V,\ v + (-v) = \vec{0}$

5) $\forall\, \alpha \in \mathbb{F}\ \forall\, u, v \in V,\ \ \alpha(u+v) = \alpha u + \alpha v$

6) $\forall\, \alpha, \beta \in \mathbb{F}\ \forall\, v \in V,\ \ (\alpha+\beta)v = \alpha v + \beta v$

7) $\forall\, \alpha, \beta \in \mathbb{F}\ \forall\, v \in V,\ \ (\alpha\beta)v = \alpha(\beta v)$

8) $\forall\, v \in V,\quad 1v = v$

Remark. 1. It's important to note that the scalars form part of the vector space structure.

2. The elements of V are called **vectors**.

3. The symbol $\vec{0}$ is being used for the **zero vector** to distinguish it from the zero scalar $0 \in \mathbb{F}$. The vector $\vec{0}$ is uniquely determined by the property in axiom 3.

4. The **additive inverse** $-v$ of a vector v is uniquely determined by the property in axiom 4.

5. A vector space over $\mathbb{R}$ is often called a **real vector space**.
   A vector space over $\mathbb{C}$ is often called a **complex vector space**.

Exercise 91. Let V be a vector space over a field $\mathbb{F}$. Show that

(a) $\forall v \in V,\quad 0v = \vec{0}$

(b) $\forall v \in V,\quad (-1)v = -v$

(Each step in your argument should be justified by reference to one of the axioms in the definition.)

Example 11.2. Here are some important examples of vector spaces. Everything we will say about vector spaces applies to each of these.

1. $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ with addition and scalar multiplication defined by:

$$(x, y) + (a, b) = (x + a, y + b) \quad \text{and} \quad \alpha(x, y) = (\alpha x, \alpha y)$$

[*] These two operations are distinct from the two binary operations in the field of scalars.

2. $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{R}\}$. The operations are given by

$$\begin{aligned}(x_1, \dots, x_n) + (y_1, \dots, y_n) &= (x_1 + y_1, \dots, x_n + y_n) \\ \alpha(x_1, \dots, x_n) &= (\alpha x_1, \dots, \alpha x_n)\end{aligned}$$

3. $\mathbb{C}^n$ is a vector space over $\mathbb{C}$. (With operations as above for $\mathbb{R}^n$.)

4. $(\mathbb{F}_3)^2$ is a vector space over $\mathbb{F}_3$. The operations are the same as above for $\mathbb{R}^n$. Here are some examples of the operations.

$$\begin{aligned}(1, 2) + (2, 2) &= (0, 1) \\ 2(2, 1) &= (1, 2)\end{aligned}$$

5. $(\mathbb{F}_p)^n$ is a vector space over $\mathbb{F}_p$. (With operations as above.)

6. ***Vector space of matrices***
   Let $\mathbb{F} = \mathbb{R}$ and $V = \mathcal{M}_{m,n}(\mathbb{R})$ (for some fixed $m, n \in \mathbb{N}$) and take the usual operations of scalar multiplication and matrix addition. Each elements of V is both a matrix and a vector! Similarly $\mathcal{M}_{m,n}(\mathbb{F})$ can be considered a vector space over $\mathbb{F}$ for any field $\mathbb{F}$.

7. ***Sequence spaces***
   Let $\mathbb{F}$ be any field and let $V = \{(x_1, x_2, \dots) \mid x_i \in \mathbb{F}\}$.[†] The operations are given by

$$\begin{aligned}(x_i)_{i\in\mathbb{N}} + (y_i)_{i\in\mathbb{N}} &= (x_i + y_i)_{i\in\mathbb{N}} \\ \alpha(x_i)_{i\in\mathbb{N}} &= (\alpha x_i)_{i\in\mathbb{N}}\end{aligned}$$

8. ***Polynomials of degree at most n***
   Let $\mathbb{F}$ be any field and fix $n \in \mathbb{N}$. Let $V = \mathcal{P}_n(\mathbb{F}) = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid a_i \in \mathbb{F}\}$. The operations are given by the usual operations on polynomials.

$$\begin{aligned}(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) &= ((a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n) \\ \alpha(a_0 + a_1x + \cdots + a_nx^n) &= (\alpha a_0) + (\alpha a_1)x + \cdots + (\alpha a_n)x^n\end{aligned}$$

9. ***Polynomials***
   Let $\mathbb{F}$ be any field let $V = \mathbb{F}[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid n \in \mathbb{N}, a_i \in \mathbb{F}\}$.. The operations are given by the usual operations on polynomials.

10. ***Functions***
    Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. We get a vector space over $\mathbb{R}$ if we define operations by

$$\begin{aligned}(f + g)(x) &= f(x) + g(x) \\ (\alpha f)(x) &= \alpha \times f(x)\end{aligned}$$

    More generally, given any non-empty set S and any field $\mathbb{F}$ we have a vector space $V = \mathcal{F}(S, \mathbb{F})$ whose vectors are functions from S to $\mathbb{F}$ and with operations as given above.

11. ***Continuous functions***
    Let $V = \mathcal{C}(\mathbb{R}, \mathbb{R})$ be the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Equipped with the operations as defined above for $\mathcal{F}(\mathbb{R}, \mathbb{R})$, this gives a vector space.

[†] An element of this set is the same as a function from $\mathbb{N}$ to $\mathbb{F}$.

11.2 Direct sum of vector spaces

We describe a standard way of combining two vector spaces to produce a third. Fix a field $\mathbb{F}$ and let V and W be vector spaces over $\mathbb{F}$.

Definition 11.3: external direct sum

The **direct sum** of V and W, denoted $V \oplus W$, has underlying set is the Cartesian product of V and W

$$V \oplus W = \{(v, w) \mid v \in V, w \in W\}$$

with the operations defined by

$$\begin{aligned}(v_1, w_1) + (v_2, w_2) &= (v_1 + v_2, w_1 + w_2)\\ k(v, w) &= (kv, kw)\end{aligned}$$

Exercise 92. Show that with these operations $V \oplus W$ is a vector space over $\mathbb{F}$.

Remark. The vector space $\mathbb{R} \oplus \mathbb{R}$ is the same as $\mathbb{R}^2$ as defined above.

11.3 Exercises

93. Determine whether or not the given set is a real vector space when equipped with the usual operations. If it is not a vector space, list all properties that fail to hold.

    (a) The set of all 2×3 matrices whose second column consists of 0's.
    That is, $\{A \in \mathcal{M}_{2,3}(\mathbb{R}) \mid A_{i2} = 0 \,\forall\, i\}$.

    (b) The set of all (real) polynomials with positive coefficients.
    That is, $\{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid a_i \in \mathbb{R}, a_i > 0\}$.

    (c) The set of all (real valued) continuous functions with the property that the function is 0 at every integer. That is, $\{f \in \mathcal{C}(\mathbb{R}, \mathbb{R}) \mid f(x) = 0 \,\forall\, x \in \mathbb{Z}\}$

94. Let V be the set of positive real numbers, that is, $V = \{x \in \mathbb{R} \mid x > 0\}$. Define the operations of vector addition $\oplus$ and scalar multiplication $\odot$ as follows:

    $$\begin{aligned}&x \oplus y = xy && \text{for all } x, y \in V && \text{(the operation on the right is multiplication in } \mathbb{R})\\ &\alpha \odot x = x^\alpha && \text{for all } \alpha \in \mathbb{R}, \text{ and } x \in V\end{aligned}$$

    Show that, equipped with these operations, V forms a real vector space. (What is the zero vector? What is the (additive) inverse of a vector $x \in V$?)

95. Let $A \in \mathcal{M}_{2,2}(\mathbb{R})$ and let $V = \{X \in \mathcal{M}_{2,1}(\mathbb{R}) \mid AX = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]\}$. Show that V is closed under the usual operations of matrix addition and scalar multiplication. That is, show that $\forall u, v \in V, u + v \in V$ and that $\forall u \in V \;\forall \alpha \in \mathbb{R}, \alpha u \in V$. Check that V together with these operations forms a vector space. (We may use standard properties of $\mathbb{R}$.)

Further reading for lecture 11

- ▷ Some references for vector spaces:

  Elementary Linear Algebra by Anton and Rorres, chapter 4

  Linear Algebra Done Right by S. Axler, chapter 1

  Finite-Dimensional Vector Spaces by P. Halmos, chapter 1

- ▷ **Modules**

  What if we were to relax the requirement that the scalars must be a field? For example, could we allow $\mathbb{Z}$ as the scalars, but keep the rest of the definition of a vector space unchanged? This leads to what is called a **module**. Many of our considerations about vector spaces continue to work for modules, but some do not. Modules are covered in the subject *MAST30005 Algebra*.

Subspaces

Definition 12.1: Subspace

A **subspace** W of a vector space V is a subset $W \subseteq V$ that is itself a vector space using the addition and scalar multiplication operations defined on V.

Remark.

1. The operations on W are the restrictions of those on V.
2. W a subspace of V is denoted $W \leqslant V$.
3. It's important to realise that subspaces are vector spaces in their own right. Anything we can prove about vector spaces applies to subspaces.

Examples 12.2.

1. $\{(t, 2t, 3t) \mid t \in \mathbb{R}\} \leqslant \mathbb{R}^3$
2. $\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\} \leqslant \mathbb{R}^3$
3. $\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 1\}$ is *not* a subspace of $\mathbb{R}^3$
4. $\{(x, y, z) \in \mathbb{R}^3 \mid z \geqslant 0\}$ is *not* a subspace of $\mathbb{R}^3$

To show that a subset fails to be a subspace it can sometimes be useful to note the following.

Lemma 12.3

Let V be a vector space and $W \subseteq V$. If W is a subspace, then $\vec{0}_V \in W$.

(Where $\vec{0}_V$ denotes the zero vector in the vector space V.)

Exercise 96. Use Exercise 91 to prove the above lemma.

The following theorem allows us to avoid checking all the vector space axioms when showing that a subset of V is a subspace.

Theorem 12.4: Subspace theorem

Let V be a vector space over $\mathbb{F}$. A subset $W \subseteq V$ is a subspace of V if and only if

0. W is non-empty
1. $\forall u, v \in W, \quad u + v \in W$ (W is closed under vector addition)
2. $\forall a \in \mathbb{F}\ \forall u \in W, \quad au \in W$ (W is closed under scalar multiplication)

Proof. Suppose that W is a subspace of V. Then it is a vector space and therefore satisfies the conditions in Definition 11.1. Therefore it is non-empty (by axiom 3) and the the operations of vector addition and scalar multiplication give functions $W \times W \to W$ and $\mathbb{F} \times W \to W$ and therefore (1) and (2) are satisfied.

For the converse, suppose that $W \subseteq V$ satisfies (0), (1), and (2). We need to check that W satisfies Definition 11.1. By (1) and (2), we have that the binary operations of vector addition $V \times V \to V$ and scalar multiplication $\mathbb{F} \times V \to V$ restrict to give functions $W \times W \to W$ and $\mathbb{F} \times W \to W$. That axioms 1,2,5,6,7,8 of Definition 11.1 are satisfied is immediate (since they hold in V and $W \subseteq V$). Since W is non empty, there is an element $w \in W$. Then by condition (2) and Exercise 91 we have that $\vec{0}_V = 0w \in W$. Therefore axiom 3 is satisfied (and $\vec{0}_W = \vec{0}_V$). To see that axiom 4 holds, let $v \in W$. Appealing to Exercise 91 again we have that $-v = (-1)v \in W$. □

Exercise 97. Use the Subspace Theorem to show that the first two examples above (in Example 12.2) are indeed subspaces.

Exercise 98. Let H and K be subspaces of a vector space V. Prove that the intersection $H \cap K$ is a subspace of V.

Examples 12.5.

1. $\{\vec{0}\}$ is always a subspace of V
2. V is always a subspace of V
3. The set of diagonal matrices $\left\{ \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \mid a, b, c \in \mathbb{C} \right\}$ is a subspace of $\mathcal{M}_{3,3}(\mathbb{C})$.
4. The subset of continuous functions $\mathcal{C}(\mathbb{R}, \mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$[*]
5. $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid ad - bc = 0 \right\}$ is *not* a subspace of $\mathcal{M}_{2,2}(\mathbb{C})$
6. $\{f : [0,1] \to \mathbb{R} \mid f(0) = 2\}$ is *not* a subspace of $\mathcal{F}([0,1], \mathbb{R})$

Lemma 12.6: Solution space of a homogeneous linear system

Let $\mathbb{F}$ be a field and $A \in \mathcal{M}_{m,n}(\mathbb{F})$. The **solution space**

$$\{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = 0\}$$

is a subspace of $\mathcal{M}_{n,1}(\mathbb{F})$.

Proof. Let $W = \{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = 0\}$. By Theorem 12.4 it is enough to show that W is non-empty and closed under vector addition and scalar multiplication.

Note first that $0 \in W$ since $A0 = 0$. Therefore $W \neq \emptyset$.

Let $X, Y \in W$. Then $A(X + Y) = AX + AY = 0 + 0 = 0$. Therefore $X + Y \in W$.

Let $X \in W$ and $\alpha \in \mathbb{F}$. Then $A(\alpha X) = \alpha AX = \alpha \times 0 = 0$. Therefore $\alpha X \in W$.

Therefore, by the Subspace Theorem, W is a subspace of $\mathcal{M}_{n,1}(\mathbb{F})$. □

Example 12.7. Consider the matrix $H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \in \mathcal{M}_{3,5}(\mathbb{F}_2)$.

The set $W = \{X \in \mathcal{M}_{5,1}(\mathbb{F}_2) \mid HX = 0\}$ is a subspace of $\mathcal{M}_{5,1}(\mathbb{F}_2)$. Solving the linear system in the usual way, we get

$$W = \{(0,0,0,0,0), (1,1,1,0,0), (1,0,0,1,1), (0,1,1,1,1)\}$$

[*] This relies on the fact (from calculus) that the sum of two continuous functions is continuous and that a multiple of a continuous function is continuous.

12.1 Exercises

99. Sketch each of the following subsets of $\mathbb{R}^2$ and determine whether it is

    (i) closed under addition,

    (ii) closed under scalar multiplication,

    (iii) a subspace of $\mathbb{R}^2$.

    (a) $A = \{(x, y) \mid y \geqslant 0\}$

    (b) $B = \{(x, y) \mid x = y\}$

    (c) $C = \{(x, y) \mid x^2 + y^2 \leqslant 1\}$

    (d) $D = \{(x, y) \mid xy = 0\}$

100. Decide which of the following are subspaces of $\mathbb{C}^3$. Explain your answers.

    (a) $A = \{(a, b, 0) \in \mathbb{C}^3 \mid a, b \in \mathbb{C}\}$

    (b) $B = \{(a, b, c) \in \mathbb{C}^3 \mid 2a - 3b + 5c = 4\}$

    (c) $C = \{(a, b, c) \in \mathbb{C}^3 \mid 2a - 3b + 5c = 0\}$

    (d) $D = \{(a - b, a + b, 2a) \in \mathbb{C}^3 \mid a, b \in \mathbb{C}\}$

101. Show that the following sets of vectors are *not* subspaces of $\mathbb{R}^n$.

    (a) The set of all vectors whose first component is 2.

    (b) The set of all vectors *except* the zero vector.

    (c) The set of all vectors the sum of whose components is 1.

102. Use the subspace theorem to decide which of the following are real vector spaces with the usual operations.

    (a) The set of real polynomials of degree exactly n (where $n \in \mathbb{N}$ is fixed).

    (b) The set of real polynomials p with $p(0) = 0$.

    (c) The set of real polynomials p with $p(0) = 1$.

103. Determine whether or not the given set is a subspace of $\mathcal{M}_{n,n}(\mathbb{C})$.

    (a) The set of all matrices, the sum of whose entries is zero.

    (b) The set of all matrices whose determinant is zero.

    (c) The diagonal matrices.

    (d) The matrices with trace equal to 0.

104. Decide which of the following are *complex* vector spaces with the usual matrix operations.

    (a) All complex 2×2 matrices $\begin{bmatrix} z_1 & z_2 \\ z_3 & z_4 \end{bmatrix}$ with z_1 and z_2 real.

    (b) All complex 2×2 matrices with $z_1 + z_4 = 0$.

105. Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $B \in \mathcal{M}_{m,1}(\mathbb{F})$. Show that if $B \neq 0$, then the set of solutions $\{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = B\}$ is *not* a subspace of $\mathcal{M}_{n,1}(\mathbb{F})$.

106. Let $V = (\mathbb{F}_2)^3$. Show that $W = \{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\} \subseteq V$ is a subspace of V.

Further reading for lecture 12

▷ References about subspaces

Elementary Linear Algebra by Anton and Rorres, §4.2
Linear Algebra Done Right by Axler, §1.C

LECTURE 13

Linear combinations: span and linear independence

We define and investigate the important notions of linear dependence and span.

13.1 Linear combinations

Definition 13.1: Linear combination

Let V be a vector space over a field $\mathbb{F}$. An expression of the form

$$\alpha_1 u_1 + \cdots + \alpha_k u_k$$

with $u_1, \ldots u_k \in V$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ is called a **linear combination**.

Notice that if $W \leqslant V$ and $u_1, \ldots u_k \in W$, then the above linear combination gives a vector in W. We will investigate how to describe subspaces using linear combinations.

13.2 Span

Definition 13.2: The span of a set of vectors

Let V be a vector space with scalars $\mathbb{F}$, and let $S \subseteq V$ be a non-empty subset of V. The **span** of S is the set of all linear combinations of vectors from S

$$\operatorname{span}(S) = \{\alpha_1 u_1 + \cdots + \alpha_k u_k \mid k \in \mathbb{N} \text{ and } u_1, \ldots, u_k \in S \text{ and } \alpha_1, \ldots, \alpha_k \in \mathbb{F}\} \subseteq V$$

If $S = \emptyset$, we define $\operatorname{span}(S) = \{\vec{0}\} \subseteq V$.

Remark. 1. The above definition of $\operatorname{span}(S)$ does *not* assume that S is finite.

2. It's immediate from the definition that $S \subseteq \operatorname{span}(S)$.

3. An alternative notation for the span of S is $\langle S \rangle$.

Examples 13.3. 1. $S = \{(1,1), (-3,-3)\} \subseteq \mathbb{R}^2$, $\operatorname{span}(S)$ is the line $\{(x,y) \in \mathbb{R}^2 \mid y = x\}$

2. $S = \{(1,1), (2,3)\} \subseteq \mathbb{R}^2$, $\operatorname{span}(S) = \mathbb{R}^2$

3. $S = \{(1,1), (2,3), (4,5)\} \subseteq \mathbb{R}^2$, $\operatorname{span}(S) = \mathbb{R}^2$

4. $S = \{(-1,1,0), (-1,0,1)\} \subseteq \mathbb{R}^3$, $\operatorname{span}(S)$ is the plane $\{(x,y,z) \mid x + y + z = 0\} \subseteq \mathbb{R}^3$

Example 13.4. Let $A \in \mathcal{M}_{4,4}(\mathbb{C})$ be given by $A = \left[\begin{smallmatrix} -2 & 0 & 6 & 8 \\ 1 & 0 & -5 & -8 \\ 2 & 0 & -2 & 0 \\ 2 & 0 & -5 & -6 \end{smallmatrix}\right]$. Calculation gives that the reduced row echelon form of A is $R = \left[\begin{smallmatrix} 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{smallmatrix}\right]$. The solution space of the linear system $AX = 0$ is given by

$$\begin{aligned}\{\left[\begin{smallmatrix} x \\ y \\ z \\ w \end{smallmatrix}\right] \mid x, y, z, w \in \mathbb{C}, x = -2w, z = -2w\} &= \{\left[\begin{smallmatrix} -2w \\ y \\ -2w \\ w \end{smallmatrix}\right] \mid y, w \in \mathbb{C}\} = \{\left[\begin{smallmatrix} 0 \\ y \\ 0 \\ 0 \end{smallmatrix}\right] + \left[\begin{smallmatrix} -2w \\ 0 \\ -2w \\ w \end{smallmatrix}\right] \mid y, w \in \mathbb{C}\} \\ &= \{y \left[\begin{smallmatrix} 0 \\ 1 \\ 0 \\ 0 \end{smallmatrix}\right] + w \left[\begin{smallmatrix} -2 \\ 0 \\ -2 \\ 1 \end{smallmatrix}\right] \mid y, w \in \mathbb{C}\} = \operatorname{span}\{\left[\begin{smallmatrix} 0 \\ 1 \\ 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} -2 \\ 0 \\ -2 \\ 1 \end{smallmatrix}\right]\}\end{aligned}$$

Proposition 13.5

Let V be a vector space with scalars $\mathbb{F}$, and let $S \subseteq V$ be a subset.

1. The span of S, $\operatorname{span}(S)$, is a subspace of V.
2. Let $W \leqslant V$ be a subspace. If $S \subseteq W$, then $\operatorname{span}(S) \leqslant W$.

Proof. Note first that if $S = \emptyset$, then both parts are trivially true since $\operatorname{span}(S) = \{\vec{0}\}$. We will assume then that $S \neq \emptyset$.

To show that $\operatorname{span}(S)$ is a subspace we show that it satisfies the conditions of the Subspace Theorem. Firstly, $\operatorname{span}(S)$ is non-empty since $\operatorname{span}(S) \supseteq S \neq \emptyset$. Now suppose that $u, v \in \operatorname{span}(S)$. Then, from the definition of the span, we have that $u = \sum_{i=1}^{k} \alpha_i u_i$ for some $\alpha_i \in \mathbb{F}$ and $u_i \in S$ and $v = \sum_{i=1}^{k} \beta_i v_i$ for some $\beta_i \in \mathbb{F}$ and $v_i \in S$. But then

$$u + v = \alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 v_1 + \cdots \beta_l v_l \in \operatorname{span}(S)$$

Therefore $\operatorname{span}(S)$ is closed under vector addition. Similarly, for any $a \in \mathbb{F}$ we have that

$$au = a(\alpha_1 u_1 + \cdots + \alpha_k u_k) = (a\alpha_1)u_1 + \cdots + (a\alpha_k)u_k \in \operatorname{span}(S)$$

and therefore $\operatorname{span}(S)$ is closed under scalar multiplication. It follows from the Subspace Theorem that $\operatorname{span}(S)$ is a subspace of V.

For the second part, suppose that $W \leqslant V$ and that $S \subseteq W$. We have

$$\begin{aligned} u \in \operatorname{span}(S) &\implies u = \alpha_1 u_1 + \cdots + \alpha_k u_k \qquad \text{for some } \alpha_i \in \mathbb{F} \text{ and } u_i \in S \\ &\implies u \in W \qquad \text{(since } W \text{ is a subspace and } u_i \in W) \end{aligned}$$

□

Remark. The above proposition tells us that $\operatorname{span}(S)$ is the 'smallest' subspace of V that contains S. We sometimes say that $\operatorname{span}(S)$ is the **subspace spanned by S**.

Exercise 107. Let V be a vector space and let $S, T \subseteq V$ be two subsets. Show that if $S \subseteq T$, then $\operatorname{span}(S) \leqslant \operatorname{span}(T)$.

Definition 13.6

Given a subspace $W \leqslant V$ of a vector space, we say that a subset $S \subseteq V$ is a **spanning set** for W if $\operatorname{span}(S) = W$. We also say that **$S$ spans W**.

Exercise 108. Show that $S \subseteq V$ is a spanning set for a subspace $W \leqslant V$ if and only if

(a) $S \subseteq W$ and

(b) $\forall w \in W \; \exists k \in \mathbb{N} \; \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \; \exists u_1, \dots, u_k \in S, \quad w = \alpha_1 u_1 + \cdots + \alpha_k u_k$

Example 13.7. We will show that $S = \{(1,1), (2,3), (4,5)\} \subseteq \mathbb{R}^2$ is a spanning set for $\mathbb{R}^2$. By Exercise 108 we need to show that given any $(a,b) \in \mathbb{R}^2$, there exist $x, y, z \in \mathbb{R}$ such that $(a,b) = x(1,1) + y(2,3) + z(4,5)$. That is, we need to show that the following linear system is consistent:

$$\begin{aligned} x + 2y + 4z &= a \\ x + 3y + 5z &= b \end{aligned}$$

Forming the augmented matrix of the linear system and row reducing gives:

$$\begin{bmatrix} 1 & 2 & 4 & a \\ 1 & 3 & 5 & b \end{bmatrix} \xrightarrow{R_2 - R1} \begin{bmatrix} 1 & 2 & 4 & a \\ 0 & 1 & 1 & b-a \end{bmatrix}$$

Therefore the linear system is consistent (for all $a, b \in \mathbb{R}$) and we conclude that S is a spanning set for $\mathbb{R}^2$.

Exercise 109. By setting up an appropriate linear system, show that the set

$$S = \{(1,-1,1), (3,-2,3), (1,0,1), (1,1,1)\}$$

is *not* a spanning set for $\mathbb{R}^3$. Use your working to find a vector $u \in \mathbb{R}^3$ such that $u \notin \operatorname{span}(S)$.

13.3 Linear independence

Definition 13.8

Let V ba a vector space over a field $\mathbb{F}$. A subset $S \subseteq V$ is called **linearly dependent** if there are vectors $u_1, \dots, u_k \in S$ and scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = \vec{0}$$

and at least one of the α_i is non-zero.

A subset $S \subseteq V$ which is not linearly dependent is called **linearly independent**.

Note. In other words, a subset S is linearly independent iff $\forall k \in \mathbb{N} \; \forall u_1, \dots, u_k \in S \; \forall \alpha_1, \dots, \alpha_k \in \mathbb{F}$,

$$\sum_{i=1}^{k} \alpha_i u_i = \vec{0} \implies \forall i, \; \alpha_i = 0$$

In any vector space V the empty set $\emptyset \subseteq V$ is a linearly independent set.

Example 13.9. We decide whether or not the subset $S = \{(1,4,1), (2,5,1), (3,6,1)\} \subseteq \mathbb{C}^3$ is linearly dependent. From the definition, the set is linearly independent iff $x(1,4,1) + y(2,5,1) + z(3,6,1) = (0,0,0) \implies x = y = z = 0$ That is, the set is linearly independent iff the following (homogeneous) linear system has a unique solution

$$\begin{aligned} x + 2y + 3z &= 0 \\ 4x + 5y + 6z &= 0 \\ x + y + z &= 0 \end{aligned}$$

We write down the corresponding matrix $A = \left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 1 & 1 \end{smallmatrix}\right]$. Reducing to row echelon form gives

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow[R_3-R_1]{R_2-4R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -1 & -2 \end{bmatrix} \xrightarrow{R_3-\frac{1}{3}R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

Since the solution set will require one parameter (Lemma 6.8), we know that there are non-trivial solutions. Therefore the set S is linearly dependent.

Exercise 110. By solving an appropriate linear system, show that the following subset of $\mathcal{P}_3(\mathbb{C})$ is linearly independent:

$$\{2+2x-2x^2+6x^3, 1+4x-4x^2+9x^3, -1-2x+3x^2-5x^3\}$$

Lemma 13.10

A subset $S \subseteq V$ is linearly dependent if and only if $\exists u \in S, u \in \operatorname{span}(S \setminus \{u\})$.

Proof. Suppose that S is linearly dependent. Let $u_1, \ldots, u_k \in S$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ be such that $\sum_{i=1}^k \alpha_i u_i = \vec{0}$ and let $j \in \{1, \ldots, k\}$ be such that $\alpha_j \neq 0$. We have

$$\begin{aligned} \alpha_1 u_1 + \cdots + \alpha_k u_k = \vec{0} &\implies -\alpha_j u_j = \alpha_1 u_1 + \cdots + \alpha_{j-1}u_{j-1} + \alpha_{j+1}u_{j+1} + \cdots + \alpha_k u_k \\ &\implies u_j = \left(\frac{\alpha_1}{-\alpha_j}\right) u_1 + \cdots + \left(\frac{\alpha_{j-1}}{-\alpha_j}\right) u_{j-1} + \left(\frac{\alpha_j}{-\alpha_j}\right) u_{j+1} + \cdots + \left(\frac{\alpha_k}{-\alpha_j}\right) u_k \end{aligned}$$

where in the last step we have relied on the fact that $\alpha_j \neq 0$.

For the converse, suppose instead that $u \in S$, $u_1, \ldots, u_k \in S \setminus \{u\}$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ are such that $u = \alpha_1 u_1 + \cdots + \alpha_k u_k$. Rearranging gives

$$\alpha_1 u_1 + \cdots + \alpha_k u_k - u = \vec{0}$$

and therefore S is linearly dependent. □

Remark. The following are particular cases of the lemma above.

1. If $S = \{u\}$, then S is linearly dependent iff $u = \vec{0}$.
2. If If $S = \{u, v\}$ consists of exactly two vectors, then S is linearly dependent iff one of the two vectors is a multiple of the other.
3. If S contains at least two vectors, then S is linearly dependent if and only if $\exists u \in S\ \exists u_1, \ldots, u_k \in S \setminus \{u\}\ \exists \alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that $u = \alpha_1 u_1 + \cdots + \alpha_k u_k$

13.4 Exercises

111. Determine which of the following sets span $\mathbb{R}^3$.

    (a) $\{(1,2,3), (-1,0,1), (0,1,2)\}$

    (b) $\{(-1,1,2), (3,3,1), (1,2,2)\}$

112. Determine whether the given set spans the given vector space.

    (a) $V = \mathbb{C}^2$, $S = \{(i,2), (3,4)\}$

(b) $V = \mathcal{P}_2(\mathbb{R})$, $S = \{2 + x^2, 3 + x + 2x^2, 1 + x + x^2, 7 + 3x + 5x^2\}$

(c) $V = (\mathbb{F}_2)^3$, $S = \{(1,1,0), (1,0,1), (0,1,1)\}$

(d) $V = \mathcal{M}_{2,2}(\mathbb{R})$, $S = \{\left[\begin{smallmatrix} 2 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 2 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 3 & -1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 3 & 1 \end{smallmatrix}\right]\}$

113. Find (finite) spanning sets for the following subspaces of $\mathbb{R}^3$:

(a) $\{(2a, b, 0) \mid a, b \in \mathbb{R}\}$

(b) $\{(a + c, c - b, 3c) \mid a, b, c \in \mathbb{R}\}$

(c) $\{(4a + d, a + 2b, c - b) \mid a, b, c, d \in \mathbb{R}\}$

114. Find (finite) spanning sets for the given vector spaces.

(a) $\{A \in \mathcal{M}_{2,2}(\mathbb{C}) \mid A^T = A\} \leqslant \mathcal{M}_{2,2}(\mathbb{C})$

(b) $\{X \in \mathcal{M}_{7,1}(\mathbb{F}_2) \mid HX = 0\}$ where $H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \in \mathcal{M}_{3,7}(\mathbb{F}_2)$

(c) $\mathbb{C}$ as a vector space over the field $\mathbb{C}$ (i.e., $V = \mathbb{C}$ and $\mathbb{F} = \mathbb{C}$).

(d) $\mathbb{C}$ as a vector space over the field $\mathbb{R}$ (i.e., $V = \mathbb{C}$ and $\mathbb{F} = \mathbb{R}$).

115. Determine whether or not the following sets of vectors are linearly independent:

(a) $\{(1,2), (0,2), (1,0), (-1,1)\} \subseteq \mathbb{R}^2$

(b) $\{(1,2), (3,-1)\} \subseteq \mathbb{R}^2$

(c) $\{(1,0,1), (-1,1,0), (0,1,1))\} \subseteq \mathbb{R}^3$

(d) $\{(2,0,0,0), (2,1,0,0), (-1,3,-2,0), (1,-2,4,-3)\} \subseteq \mathbb{R}^4$

(e) $\{(2,-3,1,-5),\ (0,1,2,2),\ (1,-2,3,0)\} \subseteq \mathbb{R}^4$

(f) $\{(1,0,2,-3),\ (0,-4,1,1),\ (2,2,0,-1),\ (1,-2,-1,3)\} \subseteq \mathbb{C}^4$

116. By setting up and solving an appropriate linear system, decide whether the vector u is a linear combination of the vectors in the set S. If so, express u as a linear combination of the vectors in S.

(a) $u = (0,0,1) \in \mathbb{R}^3, \quad S = \{(1,2,3),\ (2,3,1)\} \subseteq \mathbb{R}^3$

(b) $u = (0,1,1,2) \in \mathbb{R}^4, \quad S = \{(1,0,-2,-1),\ (-2,-1,2,0),\ (1,1,1,1)\} \subseteq \mathbb{R}^4$

117. Which of the following subsets of $\mathbb{C}^3$ are linearly independent?

(a) $\{(1-i,1,0),\ (2,1+i,0),\ (1+i,i,0)\}$

(b) $\{(1,0,-i),\ (1+i,1,1-2i),\ (0,i,2)\}$

(c) $\{(i,0,2-i),\ (0,1,i),\ (-i,-1-4i,3)\}$

118. Let $a, b, c \in \mathbb{C}$ be such that no two of them are equal. Show that the set $\{(1,a,a^2), (1,b,b^2), (1,c,c^2)\} \subseteq \mathbb{C}^3$ is linearly independent.

119. Determine whether or not the given set is linearly dependent. If the set is linearly dependent, write one of its vectors as a linear combination of the others.

(a) $\{1, 1+x, 1+x+x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$

(b) $\{1+x^2, 1+x+2x^2, x+x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$

(c) $\left\{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\} \subseteq \mathcal{M}_{2,2}(\mathbb{R})$

(d) $\left\{\begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\right\} \subseteq \mathcal{M}_{3,3}(\mathbb{R})$

120. Show that a set $S \subseteq V$ is linearly dependent iff $\exists u \in S$ such that $\mathrm{span}(S \setminus \{u\}) = \mathrm{span}(S)$.

121. Show that if a subset $S \subseteq V$ contains the zero vector $\vec{0} \in V$, then S is linearly dependent.

122. Show that any subset of a linearly independent set is itself linearly independent.

123. Let $u, v \in V$ and let $\alpha, \beta \in \mathbb{F}$ with $\alpha \neq 0$. Show that $\{u, v\}$ is linearly independent iff $\{\alpha u + \beta v, v\}$ is linear independent.

124. Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$ be a matrix that is in row echelon form. Show that the non-zero rows of A form a linearly independent subset of $\mathcal{M}_{1,n}(\mathbb{F})$.

Further material for lecture 13

- ▷ Alternative definition of $\mathrm{span}(S)$

  As an alternative to Definition 13.6 we could have defined the span of a set of vectors $S \subseteq V$ to be the the intersection of all subspaces of V that contain S. With this version it is not necessary to treat the case $S = \emptyset$ separately. As an exercise you could check that this definition gives the same thing as Definition 13.6. This version of the definition of span can be readily adapted to other algebraic structures (e.g., groups, rings, fields, etc).

Bases and dimension

A fundamental concept when working with vector spaces is that of a basis. It is a spanning set that is as 'small' as possible. A basis gives an efficient way of describing and working with vectors. Choosing a basis allows us to use coordinates to represent vectors. The dimension of a vector space is defined using bases.

14.1 Definition of basis

Definition 14.1

Let V be a vector space over a field $\mathbb{F}$. A **basis** for V is a subset $B \subseteq V$ that satisfies:

1. B is linearly independent, and
2. B is a spanning set for V.

Examples 14.2.

1. $\{(1,0),(0,1)\} \subseteq \mathbb{R}^2$ is a basis for $\mathbb{R}^2$.
2. $\{(1,1),(3,4)\} \subseteq \mathbb{R}^2$ is also a basis for $\mathbb{R}^2$.
3. $\{(1,1),(-3,-3)\} \subseteq \mathbb{R}^2$ is *not* a basis for $\mathbb{R}^2$.
4. $\{(-1,1,0),(-1,0,1)\}$ is a basis for the plane $P = \{(x,y,z) \in \mathbb{R}^3 \mid x+y+z=0\}$
5. $\{\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -2 \\ 1 \end{bmatrix}\}$ is a basis for the solution space of the matrix $A = \left[\begin{smallmatrix} 1 & 2 & 0 & 1 \\ 1 & 2 & 1 & 3 \end{smallmatrix}\right]$,
$V = \{X \in \mathcal{M}_{4,1}(\mathbb{R}) \mid AX = 0\}$
6. Let $V \leqslant \mathcal{M}_{2,2}(\mathbb{C})$ be the subspace of all matrices having trace equal to 0.
Then $\{\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]\}$ is a basis for V.

A vector space has many subsets that are bases. The following are the **standard bases** for some common vector spaces. If no other basis if specified, these are the assumed choices.

6. The standard basis for $\mathbb{F}^n$ is $\{(1,0,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,0,\dots,0,1)\}$. These vectors are sometimes denoted as $\{e_1, e_2, \cdots, e_n\}$.
7. The standard basis for $\mathcal{M}_{m,n}(\mathbb{F})$ is $\{E_{i,j} \mid i \in \{1,\dots m\}, j \in \{1,\dots,n\}\}$ where $E_{i,j}$ is the matrix with a 1 in the (i,j)-th entry and 0 in all other entries.
8. The standard basis for $\mathcal{P}_n(\mathbb{F})$ is $\{1, x, x^2, \dots, x^n\}$.
9. The standard basis for $\mathbb{F}[x]$ is $\{1, x, x^2, \dots\}$.

The following result illustrates why bases are so useful.

Lemma 14.3: Uniqueness of coordinates

Let V be a vector space and $B = \{b_1, \dots, b_n\} \subseteq V$ a basis for V. Every vector in $u \in V$ can be written *uniquely* as a linear combination of elements from B. That is, there exist *unique* $\alpha_i \in \mathbb{F}$ such that

$$u = \alpha_1 b_1 + \cdots + \alpha_n b_n$$

Proof. Since B is a basis, it is a spanning set for V and therefore every element in $u \in V$ can be written as a linear combination $u = \sum_{i=1}^{n} \alpha_i b_i$ where $\alpha_i \in \mathbb{F}$. We will use the fact that B is linearly independent to show that the α_i are uniquely determined by u. Suppose that $u = \sum_{i=1}^{n} \beta_i b_i$ for some $\beta_i \in \mathbb{F}$. Then we have

$$\begin{aligned} \sum_{i=1}^{n} \alpha_i b_i = \sum_{i=1}^{n} \beta_i b_i &\implies \sum_{i=1}^{n} \alpha_i b_i - \sum_{i=1}^{n} \beta_i b_i = \vec{0} \\ &\implies \sum_{i=1}^{n} (\alpha_i - \beta_i) b_i = \vec{0} \\ &\implies \forall i,\ (\alpha_i - \beta_i) = 0 && \text{(since } B \text{ is linearly independent)} \\ &\implies \forall i,\ \beta_i = \alpha_i \end{aligned}$$

□

14.2 Dimension

Theorem 14.4

Let V be a vector space and let $B = \{b_1, \dots, b_n\}$ be a basis for V. Let $S \subseteq V$ be a subset.

1. If S in linearly independent, then it has at most n elements.
2. If S is a spanning set, then it has at least n elements.
3. Every basis of V has n elements.

Proof. Note first that the third part is an immediate consequence of the first two parts, which we now prove.

1) We prove the contrapositive: if S has more than n elements, then S is linearly dependent.

Suppose that $|S| > n$, and let $\{u_1, \dots, u_{n+1}\} \subseteq S$ be distinct elements of S. Let $\alpha_{ij} \in \mathbb{F}$ be such that $u_j = \alpha_{1j} b_1 + \cdots + \alpha_{nj} b_n$. Let $A \in \mathcal{M}_{n,n+1}(\mathbb{F})$ be given by $A_{ij} = \alpha_{ij}$. To show that S is linearly dependent we will show that there is a non-trivial linear combination of the u_i that gives the zero vector. For $x_1, \dots, x_{n+1} \in \mathbb{F}$ we have

$$\begin{aligned} \sum_{j=1}^{n+1} x_j u_j = \vec{0} &\iff \sum_{j=1}^{n+1} x_j \sum_{i=1}^{n} \alpha_{ij} b_i = \vec{0} \\ &\iff \sum_{i=1}^{n} \left(\sum_{j=1}^{n+1} \alpha_{ij} x_j \right) b_i = \vec{0} \\ &\iff \forall i, \quad \sum_{j=1}^{n+1} \alpha_{ij} x_j = 0 && \text{(since } B \text{ is linearly independent)} \\ &\iff AX = 0 \quad \text{where } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n+1} \end{bmatrix} \end{aligned}$$

The final expression is a homogeneous linear system. The solution space of this linear system will need $(n+1) - \operatorname{rank}(A)$ parameters (see Lemma 6.8). Since A has n rows, we have $\operatorname{rank}(A) \leqslant n$, and therefore $(n+1) - \operatorname{rank}(A) \geqslant 1$. Therefore the linear system more than one solution. In particular, there is a non-trivial solution.

2) Will will show the contrapositive: if S has fewer than n elements, then it is not a spanning set. Suppose that $S = \{u_1, \dots, u_k\}$ with $k < n$. Let $\alpha_{ij} \in \mathbb{F}$ be such that $u_j = \alpha_{1j}b_1 + \cdots + \alpha_{nj}b_n$. Let $A \in \mathcal{M}_{n,k}(\mathbb{F})$ be given by $A_{ij} = \alpha_{ij}$. We will show that there exist $\gamma_1, \dots, \gamma_n \in \mathbb{F}$ such that $v = \gamma_1 b_1 + \cdots + \gamma_n b_n$ is not in $\operatorname{span}(S)$.

$$\begin{aligned}
v \in \operatorname{span}(S) &\iff \exists x_j \in \mathbb{F}, \quad \sum_{i=1}^{n} \gamma_i b_i = \sum_{j=1}^{k} x_j u_j \\
&\iff \exists x_j \in \mathbb{F}, \quad \sum_{i=1}^{n} \gamma_i b_i = \sum_{j=1}^{k} x_j \sum_{i=1}^{n} \alpha_{ij} b_i \\
&\iff \exists x_j \in \mathbb{F}, \quad \sum_{i=1}^{n} \gamma_i b_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} \alpha_{ij} x_j \right) b_i \\
&\iff \exists x_j \in \mathbb{F}\ \forall i, \quad \gamma_i = \sum_{j=1}^{n} \alpha_{ij} x_j \qquad \text{(Lemma 14.3)} \\
&\iff \exists x_j \in \mathbb{F}, \quad A \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}
\end{aligned}$$

We need to show that there is a choice for the γ_i such that the linear system $AX = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}$ is inconsistent. Let $E \in \mathcal{M}_{n,n}(\mathbb{F})$ be an invertible matrix such that $EA = R$ is in reduced row echelon form. Since $R \in \mathcal{M}_{n,k}(\mathbb{F})$ has fewer columns than rows, the bottom row of R is all zeros. Now let $e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \in \mathcal{M}_{n,1}(\mathbb{F})$ and choose the γ_i by $C = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = E^{-1}e_n$. Then $E[A \,|\, C] = [R \,|\, e_n]$. Since the bottom row of R is all zeros, the linear system in inconsistent. □

Definition 14.5: Dimension

The **dimension** of a vector space V is size of a basis for V. It is denoted by $\dim(V)$.

We call a vector space **finite dimensional** if it has a basis with a finite number of elements, and **infinite dimensional** otherwise.

Examples 14.6. (cf. Example 14.2)

1. $\dim(\mathbb{R}^2) = 2$
2. $\dim(P) = 2$, where $P \leqslant V$ is the plane $P = \{(x,y,z) \in \mathbb{R}^3 \mid x+y+z=0\}$
3. $\dim(V) = 3$, where $V \leqslant \mathcal{M}_{2,2}(\mathbb{C})$ is $V = \{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \mid a+d=0\}$
4. $\dim(\mathbb{F}^n) = n$
5. $\dim(\mathcal{M}_{m,n}(\mathbb{F})) = mn$
6. $\dim(\mathcal{P}_n(\mathbb{F})) = n+1$
7. $\mathbb{F}[x]$ is infinite dimensional
8. $\mathcal{F}(\mathbb{R},\mathbb{R})$ is infinite dimensional

We finish with an important fact.

Theorem 14.7

Every vector space has a basis.

The full proof is a little too technical to go through here, but the idea of the proof is the following. Start with a linearly independent set $S \subseteq V$. If S is not a basis, then there exists $u \in V$ such that $S \cup \{u\}$ is linearly independent. Therefore a maximal linearly independent set must be a basis. That there is a maximal linearly independent set requires the use of 'Zorn's Lemma'.

We isolate the following consequence of the above argument.

Lemma 14.8

Let V be a vector space and $S \subseteq V$ a subset. Suppose that $|S| = \dim(V)$. Then S is linearly independent if and only if S is a spanning set (for V).

14.3 Exercises

125. Determine whether or not the given set is a basis for $\mathbb{C}^3$.

   (a) $\{(i, 0, -1), (1, 1, 1), (0, -i, i)\}$ (b) $\{(i, 1, 0), (0, 0, 1)\}$

126. Which of the following sets of vectors are bases for $\mathcal{P}_2(\mathbb{R})$?

   (a) $\{1 - 3x + 2x^2, 1 + x + 4x^2, 1 - 7x\}$

   (b) $\{1 + x + x^2, x + x^2, x^2\}$

127. Find a basis for the given vector space:

   (a) The subspace of $\mathcal{M}_{2,2}(\mathbb{C})$ consisting of all diagonal 2×2 matrices.

   (b) The subspace of $\mathcal{M}_{2,2}(\mathbb{F}_2)$ consisting of all 2×2 matrices whose diagonal entries are zero.

   (c) The subspace of $\mathcal{P}_3(\mathbb{R})$ consisting of all polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ with $a_2 = 0$.

128. (a) Show that any set of four polynomials in $\mathcal{P}_2(\mathbb{C})$ is linearly dependent.

   (b) Show that a set consisting of two polynomials cannot span $\mathcal{P}_2(\mathbb{C})$.

129. Let

$$A = \begin{bmatrix} 0 & 1 & 4 \\ 6 & 1 & -8 \\ -9 & 3 & 15 \end{bmatrix} \quad \text{and} \quad W = \left\{(x, y, z) \in \mathbb{R}^3 : A \begin{bmatrix} x & y & z \end{bmatrix}^T = 3 \begin{bmatrix} x & y & z \end{bmatrix}^T \right\}.$$

   Show that W is a subspace of $\mathbb{R}^3$, and find a basis for it.

130. Let $V = \mathbb{R}[x]$ and $W = \{p(x) \in \mathbb{R}[x] \mid p(1) = 0\}$.

   Show that W is a subspace of V, and find a basis for W.

131. Let V be a finite dimensional vector space and $W \leqslant V$ a subspace of V.

   (a) Show that $\dim(W) \leqslant \dim(V)$.

   (b) Show that if $\dim(W) = \dim(V)$, then $W = V$.

Extra material for lecture 14

existence of basis appendix inserted here (pdf cut-paste)

LECTURE **15**

Coordinates relative to a basis

15.1 Coordinates

From Lemma 14.3 we know that every vector in a vector space can be written in exactly one way as a linear combination of a given basis. The scalars that appear as the coefficients in the linear combination are called the coordinates. Once we fix a basis for a finite dimensional vector space V, there is a one-to-one correspondence[*] between vectors and their coordinate matrices.

Definition 15.1: Coordinates

Let V be a vector space over a field $\mathbb{F}$. Suppose that $\mathcal{B} = \{b_1, \dots, b_n\}$ is an *ordered* basis for V, that is, a basis arranged in order: b_1 first, b_2 second and so on. For $v \in V$ we can write

$$v = \alpha_1 b_1 + \cdots + \alpha_n b_n \quad \text{for some scalars} \quad \alpha_1, \dots, \alpha_n \in \mathbb{F}$$

The scalars $\alpha_1, \dots, \alpha_n$ are uniquely determined by v (see Lemma 14.3) and are called the **coordinates** of v relative to $\mathcal{B}$.

The column matrix

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathcal{M}_{n,1}(\mathbb{F})$$

is called the **coordinate matrix**[†]of v with respect to $\mathcal{B}$.

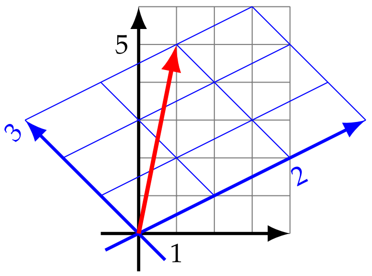
Example 15.2.

1. $V = \mathcal{P}_3(\mathbb{C})$, $\mathcal{B} = \{1, x, x^2, x^3, x^4\}$, $v = 1 + 2x^3 + 2x^4$, $[v]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 2 \end{bmatrix}$
2. $V = \mathcal{M}_{2,2}(\mathbb{Q})$, $\mathcal{B} = \{\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]\}$, $v = \left[\begin{smallmatrix} 1 & 0 \\ 2 & 2 \end{smallmatrix}\right]$, $[v]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 2 \end{bmatrix}$
3. $V = \mathbb{R}^2$, $\mathcal{B} = \{(1,2), (3,4)\}$, $v = (2,2)$, $[v]_{\mathcal{B}} = \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$
4. $V = \{(x,y,z) \in \mathbb{R}^3 \,|\, x+y+z = 0\} \leqslant \mathbb{R}^3$, $\mathcal{B} = \{(1,2,-3), (2,-1,-1)\}$, $v = (-3,4,-1)$, $[v]_{\mathcal{B}} = \left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right]$

Note. The coordinates depend on the basis chosen. If we change the basis $\mathcal{B}$, the coordinates of a vector will change.

Example 15.3. Consider the vector space $V = \mathbb{R}^2$. Then $\mathcal{S} = \{(1,0), (0,1)\}$ and $\mathcal{C} = \{(2,1), (-1,1)\}$ are bases for V. For $v = (1,5) \in V$ we have

$$[v]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \quad \text{and} \quad [v]_{\mathcal{C}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



[*] i.e., a bijection

[†] also called the **coordinate vector** of v

Lemma 15.4

Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be an ordered basis for V. Let $u, v \in V$ and $\alpha \in \mathbb{F}$. Then

$$[u+v]_{\mathcal{B}} = [u]_{\mathcal{B}} + [v]_{\mathcal{B}}$$

$$[\alpha v]_{\mathcal{B}} = \alpha[v]_{\mathcal{B}}$$

Proof. Let the ordered basis be $\mathcal{B} = \{b_1, \dots, b_n\}$ and let $\alpha_i, \beta_i \in \mathbb{F}$ be such that $u = \sum_{i=1}^n \alpha_i b_i$ and $v = \sum_{i=1}^n \beta_i b_i$. Then

$$u + v = \sum_{i=1}^n \alpha_i b_i + \sum_{i=1}^n \beta_i b_i = \sum_{i=1}^n (\alpha_i + \beta_i) b_i$$

and therefore

$$[u]_{\mathcal{B}} + [v]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} \alpha_1+\beta_1 \\ \vdots \\ \alpha_n+\beta_n \end{bmatrix} \text{ and } [u+v]_{\mathcal{B}} = \begin{bmatrix} \alpha_1+\beta_1 \\ \vdots \\ \beta_n+\alpha_n \end{bmatrix}$$

The second statement is similar, and left as an exercise. □

Exercise 132. Let V be an n-dimensional vector space with scalars $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V. Show that the map $\varphi : V \to \mathcal{M}_{n,1}(\mathbb{F})$ given by $\varphi(v) = [v]_{\mathcal{B}}$ is a bijection.

The following observation allows us to convert some questions about an n-dimensional vector space to the corresponding questions about $\mathbb{F}^n$.

Lemma 15.5

Let V be an n-dimensional vector space with scalars $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V. Let $S \subseteq V$ be a subset of V and define $T \subseteq \mathcal{M}_{n,1}(\mathbb{F})$ by $T = \{[v]_{\mathcal{B}} \mid v \in S\}$. Then

1. S is linearly independent iff T is linearly independent
2. S is a spanning set for V iff T is a spanning set for $\mathcal{M}_{n,1}(\mathbb{F})$

Proof.

$$\begin{aligned}
S \text{ linearly dependent} &\iff \exists u_1, \dots, u_k \in S\ \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \text{ (not all zero)} \\
&\qquad \sum_{i=1}^k \alpha_i u_i = \vec{0}_V \\
&\iff \exists u_1, \dots, u_k \in S\ \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \text{ (not all zero)} \\
&\qquad \left[\sum_{i=1}^k \alpha_i u_i\right]_{\mathcal{B}} = [\vec{0}_V]_{\mathcal{B}} && \text{(by Exercise 132)} \\
&\iff \exists u_1, \dots, u_k \in S\ \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \text{ (not all zero)} \\
&\qquad \sum_{i=1}^k \alpha_i [u_i]_{\mathcal{B}} = \vec{0}_{\mathcal{M}_{n,1}(F)} && \text{(by Lemma 15.4)} \\
&\iff T \text{ is linearly dependent}
\end{aligned}$$

$$\begin{aligned}
S \text{ a spanning set for } V &\iff \forall v \in V\ \exists u_1,\dots,u_k \in S\ \exists \alpha_1,\dots,\alpha_k \in \mathbb{F}, \quad v = \sum_{i=1}^{k} \alpha_i u_i \\
&\iff \forall v \in V\ \exists u_1,\dots,u_k \in S\ \exists \alpha_1,\dots,\alpha_k \in \mathbb{F}, \quad [v]_{\mathcal{B}} = \sum_{i=1}^{k} \alpha_i [u_i]_{\mathcal{B}} \\
&\iff \forall u \in \mathcal{M}_{n,1}(F)\ \exists u_1,\dots,u_k \in S\ \exists \alpha_1,\dots,\alpha_k \in \mathbb{F}, \quad u = \sum_{i=1}^{k} \alpha_i [u_i]_{\mathcal{B}} \\
&\iff T \text{ a spanning set for } \mathcal{M}_{n,1}(F)
\end{aligned}$$

□

Examples 15.6.

1. Consider $S = \{2-3x+x^2-5x^3, x+2x^2+2x^3, 1-2x+3x^2\} \subseteq \mathcal{P}_3(\mathbb{R})$. Taking coordinates with respect to the standard basis for $\mathcal{P}_3(\mathbb{C})$ we get $\{(2,-3,1,-5),\ (0,1,2,2),\ (1,-2,3,0)\} \subseteq \mathbb{R}^4$. Since this set is linearly independent (see Exercise 115(e)), the set S is linearly independent.

2. Consider the vector space $V \leqslant \mathcal{M}_{2,2}(\mathbb{C})$ of matrices having trace 0. In Example 14.2.6 we saw that $\mathcal{B} = \{\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right], \left[\begin{smallmatrix}0 & 1\\ 0 & 0\end{smallmatrix}\right], \left[\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right]\}$ is a basis for V. Let $S = \{\left[\begin{smallmatrix}1 & 0\\ 1 & -1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1 & 1\\ 0 & 1\end{smallmatrix}\right], \left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]\} \subseteq V$. Taking coordinates with respect to $\mathcal{B}$ we get $\{\left[\begin{smallmatrix}1\\ 0\\ 1\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\ 1\\ 0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\ 1\\ 1\end{smallmatrix}\right]\}$. Since this set is linearly dependent (see Exercise 115(c)), we have that S is linearly dependent.

15.2 Exercises

133. (a) Show that the set $B = \{(-2,2,2),\ (3,-2,3),\ (2,-1,1)\}$ is a basis for $\mathbb{R}^3$.

(b) Find the vectors $x, y \in \mathbb{R}^3$ whose coordinates with respect to B are

$$[x]_B = \begin{bmatrix}2\\1\\1\end{bmatrix} \quad \text{and} \quad [y]_B = \begin{bmatrix}1\\0\\-1\end{bmatrix}$$

(c) For each of the following vectors find its coordinates with respect to B:

$$a = (2,-1,1) \qquad b = (1,0,5) \qquad c = (3,-1,6)$$

134. Find the coordinate vector of v with respect to the given basis $\mathcal{B}$ for the vector space V.

(a) $v = 2 - x + 3x^2, \mathcal{B} = \{1, x, x^2, x^3\}, V = \mathcal{P}_3(\mathbb{C})$.

(b) $v = \begin{bmatrix}1 & 2 & 1\\ -1 & 1 & 2\end{bmatrix}, \mathcal{B} = \{E_{ij} \mid i = 1,2; j = 1,2,3\}, V = \mathcal{M}_{2,3}(\mathbb{R})$.
(Here E^{ij} is the matrix with (i,j) entry equal to 1 and other entries equal to 0.)

(c) $v = -2 + (5+i)x, \mathcal{B} = \{x+1, x-1\}, V = \mathcal{P}_1(\mathbb{C})$

(d) $v = \begin{bmatrix}-2 & 0\\ 0 & 3\end{bmatrix}, \mathcal{B} = \left\{\begin{bmatrix}i & 0\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right\}$, V is the vector space of all diagonal 2×2 complex matrices.

135. Use coordinate matrices to decide whether or not the given set is linearly independent. If it is linearly dependent, express one of the vectors as a linear combination of the others.

(a) $\{x^2 + x - 1, x^2 - 2x + 3, x^2 + 4x - 3\} \in \mathcal{P}_2(\mathbb{R})$

(b) $\left\{\begin{bmatrix}1 & 2\\ -1 & 0\end{bmatrix}, \begin{bmatrix}0 & -1\\ 1 & 1\end{bmatrix}, \begin{bmatrix}1 & 0\\ 1 & 2\end{bmatrix}\right\}$ in $\mathcal{M}_{2,2}(\mathbb{R})$

Extra material for lecture 15

- ▷ References about bases and coordinates

  Elementary Linear Algebra by Anton and Rorres, §4.4

Row and column space of a matrix

16.1 Row and column space of a matrix

We've already seen that, given a matrix $A \in \mathcal{M}_{m,n}(\mathbb{F})$, its solution space is a subspace of $\mathcal{M}_{n,1}(\mathbb{F})$. There are two other spaces we can associate to a matrix.

Definition 16.1

Let $A = (a_{ij}) \in \mathcal{M}_{m,n}(\mathbb{F})$ be a matrix.

1. The **solution space** of A (also called the *null space* or *kernel*) is the subspace of $\mathcal{M}_{n,1}(\mathbb{F})$ given by $\{X \in \mathcal{M}_{n,1}(\mathbb{F}) \mid AX = 0\}$.

   It will often be identified with a subspace of $\mathbb{F}^n$.
   Explicitly, $\mathrm{solspace}(A) = \{(x_1, \dots, x_n) \in \mathbb{F}^n \mid A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T = 0\}$.

2. The **row space**, $\mathrm{rowspace}(A)$, is the subspace of $\mathcal{M}_{1,n}(\mathbb{F})$ spanned by the rows of A.

   It will often be identified with a subspace of $\mathbb{F}^n$. Explicitly, we identify the i-th row with the element $r_i = (a_{i1}, \dots, a_{in}) \in \mathbb{F}^n$ and let $\mathrm{rowspace}(A) = \mathrm{span}\{r_1, \dots, r_m\} \leqslant \mathbb{F}^n$

3. The **column space**, $\mathrm{colspace}(A)$, is the subspace of $\mathcal{M}_{m,1}(\mathbb{F})$ spanned by the columns of A.

   It will often be identified with a subspace of $\mathbb{F}^m$. Explicitly, we identify the j-th column with the element $c_j = (a_{1j}, \dots, a_{mj}) \in \mathbb{F}^m$ and let $\mathrm{colspace}(A) = \mathrm{span}\{c_1, \dots, c_n\}$

Remark. Observe that

$$AX = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Therefore, AX is an element of the column space of A (for any $X \in \mathcal{M}_{n,1}(\mathbb{F})$).

Example 16.2. As a concrete example, let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 1 & 1 \end{bmatrix} \in \mathcal{M}_{3,2}(\mathbb{R})$.

Then $\mathrm{rowspace}(A) = \mathrm{span}\{(1,2), (3,4), (1,1)\} \leqslant \mathbb{R}^2$ and $\mathrm{colspace}(A) = \mathrm{span}\{(1,3,1), (2,4,1)\} \leqslant \mathbb{R}^3$.

Lemma 16.3

Let $A, R \in \mathcal{M}_{m,n}(\mathbb{F})$. Suppose that $A \sim R$ and R is in row echelon form.

1. $\mathrm{rowspace}(A) = \mathrm{rowspace}(R)$
2. The non-zero rows of R are a basis for the row space of A.
3. The pivot columns*of A form a basis for the column space of A.

4. Every non-pivot column of A can be written as a linear combination of the columns to its left

Proof. 1) We show that if two matrices are row equivalent, then they have the same row space. For that it is enough to show that if R is obtained from A by a single row operation, then the the row space of R is a subset of the row space of A. But this is clear since each row of R is a linear combination of the rows of A.

2) From the first part, we know that the non-zero rows of R form a spanning set for $\mathrm{rowspace}(A)$. That the non-zero rows are linearly independent is exercise 124.

3) Since the position of the leading entries will be the same, we can assume that R is in reduced row echelon form. The pivot columns of R are then linearly independent. Denoting the columns of A by $c_1, \dots, c_n$ and those of R by $d_1, \dots, d_n$ we have that

$$\begin{aligned}\alpha_1 c_1 + \cdots + \alpha_n c_n = 0 &\iff A\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{bmatrix} = 0 && \text{(see the remark ofter Definition 16.1)}\\ &\iff R\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{bmatrix} = 0 && \text{(A and R are row equivalent)}\\ &\iff \alpha_1 d_1 + \cdots + \alpha_n d_n = 0\end{aligned}$$

Therefore, the pivot columns of A form a linearly independent set. To see that the set of pivot columns of A forms a spanning set for $\mathrm{colspace}(A)$ it is enough to show that each of the non-pivot columns of A can be written as a linear combination of the pivot columns of A. But this is clearly true for the columns of R, and therefore for the columns of A by the above calculation.

4) The statement holds for the columns of R since R is in row echelon form. That the same holds for A then follows from the remark after Definition 16.1 (as in the previous part). □

Example 16.4. Let

$$A = \begin{bmatrix}2 & 1 & 3 & 1 & -1\\ 2 & 1 & 3 & 1 & -1\\ 2 & 1 & 4 & 3 & 1\\ 2 & 1 & 5 & 5 & 3\end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix}2 & 1 & 3 & 1 & -1\\ 0 & 0 & 1 & 2 & 2\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

Note that $A \sim R$ and that R is in row echelon form. We have that

- ▷ The first two rows of R form a basis for the row space of A
- ▷ The row space of A has dimension 2
- ▷ The first two rows of A *do not* form a basis for the row space of A
- ▷ The row space of A is equal to the row space of R
- ▷ The first and third columns of A form a basis for the column space of A
- ▷ The column space of A has dimension 2
- ▷ The first and third columns of R *do not* form a basis for the column space of A
- ▷ The column space of A is *not* equal to the column space of R

Definition 16.5

The dimension of the row space is called the **row rank** of a matrix. The dimension of the column space is called the **column rank** of a matrix.

From the above lemma we have:

[*] That is, the columns of A such that the corresponding column in R has a leading entry.

Corollary 16.6

The rank, row rank and column rank of a matrix are all equal. □

16.2 Exercises

136. In each part find a basis for, and the dimension of, the indicated subspace.

    (a) The solution space of the homogeneous linear system (over $\mathbb{R}$):

$$\begin{array}{ccccccccc} x_1 & - & 2x_2 & + & x_3 & & & = & 0 \\ & & x_2 & - & x_3 & + & x_4 & = & 0 \\ x_1 & - & x_2 & & & + & x_4 & = & 0 \end{array}$$

    (b) The solution space (over $\mathbb{C}$) of

$$\begin{array}{ccccccccccc} x_1 & - & 3x_2 & + & x_3 & & & - & x_5 & = & 0 \\ x_1 & - & 2x_2 & + & x_3 & - & x_4 & & & = & 0 \\ x_1 & - & x_2 & + & x_3 & - & 2x_4 & + & x_5 & = & 0 \end{array}$$

    (c) The subspace of $\mathbb{R}^4$ of all vectors of the form $(x, -y, x - 2y, 3y)$.

137. For each of the following real matrices find a basis for the

    (i) column space,

    (ii) row space,

    (iii) solution space.

(a) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -1 & 3 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix}$

138. Find a basis for each of the column space, row space and solution space of the matrix

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \end{bmatrix} \in \mathcal{M}_{4,4}(\mathbb{F}_3)$$

What is the rank of the matrix?

139. Let $w = [x_1 \cdots x_n]^T \in \mathcal{M}_{n,1}(\mathbb{R})$ be fixed, and let $W = \operatorname{span}\{w\}$. Show the there exists a matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$ whose solution space is W.

Extra material for lecture 16

- ▷ References about row and column spaces

  Elementary Linear Algebra by Anton and Rorres, §4.7

Some techniques for finite-dimensional vector spaces

We collect here some techniques that result from the theory that we've seen so far.

Algorithm 17.1: To decide if a set is linearly independent

Let V be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V.
Let $S = \{u_1, \dots, u_k\} \subseteq V$. To decide if S is linearly independent:

1. Form the matrix $A \in \mathcal{M}_{n,k}(\mathbb{F})$ given by $A = [\,[u_1]_{\mathcal{B}} \cdots [u_k]_{\mathcal{B}}\,]$
2. Calculate $\operatorname{rank}(A)$
3. S is linearly independent iff $\operatorname{rank}(A) = k$

Remark. It is always the case that $\operatorname{rank}(A) \leqslant k$.

Example 17.2. Let $S = \{(2+2i)+2x+2x^2-2x^3, i+(1+i)x+x^2-x^3, -1-x^2+x^3\} \subseteq \mathcal{P}_3(\mathbb{C})$. Is S linearly independent? To apply the above technique we first fix a basis for $\mathcal{P}_3(\mathbb{C})$. Let's use the standard basis $\mathcal{B} = \{1, x, x^2, x^3\}$. Letting $u_1 = (2+2i)+2x+2x^2-2x^3$, $u_2 = i+(1+i)x+x^2-x^3$, $u_3 = -1-x^2+x^3$ we get

$$A = [\,[u_1]_{\mathcal{B}} \quad [u_2]_{\mathcal{B}} \quad [u_3]_{\mathcal{B}}\,] = \begin{bmatrix} 2+2i & i & -1 \\ 2 & 1+i & 0 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} -2 & -1 & 0 \\ 0 & i & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since the rank is 2 which is strictly less the the number of elements in S, we conclude that S is linearly dependent.

Remark. Since we know that $\operatorname{rank}(A) = \operatorname{rank}(A^T)$ (Corollary 16.6), we could have instead calculated the rank of the matrix A^T

$$A^T = \begin{bmatrix} 2+2i & 2 & 2 & -2 \\ i & 1+i & 1 & -1 \\ -1 & 0 & -1 & 1 \end{bmatrix}$$

Algorithm 17.3: To decide whether a subset of V is a spanning set for V

Let V be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V.
Let $S = \{u_1, \dots, u_k\} \subseteq V$. To decide if S is a spanning set for V:

1. Form the matrix $A \in \mathcal{M}_{n,k}(\mathbb{F})$ given by $A = [\,[u_1]_{\mathcal{B}} \cdots [u_k]_{\mathcal{B}}\,]$
2. Calculate $\operatorname{rank}(A)$
3. S is a spanning set for V iff $\operatorname{rank}(A) = n$

Example 17.4. With $S \subseteq \mathcal{P}_3(\mathbb{C})$ as in Example 17.2, the calculation done there shows that S is not a spanning set for $\mathcal{P}_3(\mathbb{C})$ because $\operatorname{rank}(A) = 2 < 4 = \dim(\mathcal{P}_3(\mathbb{C}))$.

Algorithm 17.5: To find a subset of a set $S = \{u_1, \dots, u_k\}$ that is a basis for span(S)

Let V be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V.
Let $S = \{u_1, \dots, u_k\} \subseteq V$. To find a subset of S that is a basis for $\operatorname{span}(S)$:

1. Form the matrix $A \in \mathcal{M}_{n,k}(\mathbb{F})$ given by $A = [\,[u_1]_{\mathcal{B}} \cdots [u_k]_{\mathcal{B}}\,]$
2. Reduce A to row echelon form R
3. Locate the pivot columns of A
4. The corresponding elements of S form a basis for $\operatorname{span}(S)$. That is, the basis is

$$\{u_i \mid \text{the } i\text{-th column of } R \text{ is a pivot column}\}$$

Example 17.6. With $S \subseteq \mathcal{P}_3(\mathbb{C})$ as in Example 17.2, the calculation done there shows that the set $\mathcal{C} = \{(2+2i) + 2x + 2x^2 - 2x^3, i + (1+i)x + x^2 - x^3\}$ is a basis for $\operatorname{span}(S)$.

Remark. Alternatively, to find a basis for $\operatorname{span}(S)$ we could consider the matrix

$$A^T = \begin{bmatrix} 2+2i & 2 & 2 & -2 \\ i & 1+i & 1 & -1 \\ -1 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} -1 & 0 & -1 & 1 \\ 0 & 1 & -i & i \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The non-zero rows in the row echelon form give a basis $\{(-1,0,-1,1),(0,1,-i,i)\}$ for $\operatorname{rowspace}(A)$ and hence a basis $\{-1 - x^2 + x^3, x - ix^2 + ix^3\}$ for $\operatorname{span}(A)$.

Algorithm 17.7: To find a superset of a linearly independent set that is a basis

Let V be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V.
Let $W \leqslant V$ be a subspace and fix a basis $\{w_1, \dots, w_m\}$ for W.

Given a linearly independent subset $S = \{u_1, \dots, u_k\} \subseteq W$, to extend S to obtain a basis of W:

1. Form the matrix $A = [\,[u_1]_{\mathcal{B}} \;\cdots\; [u_k]_{\mathcal{B}} \;\; [w_1]_{\mathcal{B}} \;\cdots\; [w_m]_{\mathcal{B}}\,] \in \mathcal{M}_{n,k+m}(\mathbb{F})$
2. Reduce to row echelon form
3. Locate the pivot columns of A.[a]
4. The vectors correspnding to the pivot columns form a basis for W. That is

$$\{u_1, \dots, u_k\} \cup \{w_i \mid \text{the } (k+i)\text{-th column of } A \text{ is a pivot column}\}$$

is a basis for W.

[a] The first k columns will all be pivot columns since S is linearly independent.

Note. This includes the case in which $W = V$ and $\mathcal{B} = \{w_1, \dots, w_m\}$.

Example 17.8. We saw in the previous example that $\{-1 - x^2 + x^3, x - ix^2 + ix^3\} \subseteq \mathcal{P}_3(\mathbb{C})$ is a linearly independent set. We can extend it to a basis for $\mathcal{P}_3(\mathbb{C})$ as follows (letting $u_1 = -1 - x^2 + x^3, u_2 = x - ix^2 + ix^3$ and using the standard basis $\mathcal{B} = \{1, x, x^2, x^3\}$)

$$[[u_1]_{\mathcal{B}}\;[u_2]_{\mathcal{B}}\;[1]_{\mathcal{B}}\;[x]_{\mathcal{B}}\;[x^2]_{\mathcal{B}}\;[x^3]_{\mathcal{B}}] = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ -1 & -i & 0 & 0 & 1 & 0 \\ 1 & i & 0 & 0 & 0 & 1 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & i & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Therefore the set $\{-1 - x^2 + x^3, x - ix^2 + ix^3, 1, x^2\}$ is a basis for $\mathcal{P}_3(\mathbb{C})$.

Algorithm 17.9: To write a vector as a linear combination of a set S

Let V be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B}$ be a basis for V.
Let $S = \{u_1, \dots, u_k\} \subseteq V$ and let $w \in V$. To write w as a linear combination of the elements in S:

1. Form the matrix $A = [\, [u_1]_{\mathcal{B}} \, \cdots \, [u_k]_{\mathcal{B}} \;\; [w]_{\mathcal{B}}]$
2. Reduce to row echelon form
3. If the last column is a pivot column, then $w \notin \operatorname{span}(S)$
4. Otherwise, continue to reduced row echelon form R. The entries in the last column of R give the coefficients in the linear combination. (See the example below.)

Example 17.10. Consider the set $S = \{-1 - x^2 + x^3, x - ix^2 + ix^3\} \subseteq \mathcal{P}_3(\mathbb{C})$. Let's try to write the element $w = 1 + x + x^2 + x^3$ as a linear combination of these two elements:

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -i & 1 \\ 1 & i & 1 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & i \\ 0 & 0 & 0 \end{bmatrix}$$

From which we conclude that $w \notin \operatorname{span}(S)$. If we consider instead the vector $v = -1 + ix$ we get

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & 1 & i \\ -1 & -i & 0 \\ 1 & i & 0 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and we see that $v = u_1 + iu_2$, where $u_1 = -1 - x^2 + x^3$, $u_2 = x - ix^2 + ix^3$.

17.1 Exercises

140. Let $W = \operatorname{span}(S)$ where $S = \left\{ \left[\begin{smallmatrix} 1 & -2 \\ 4 & 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 & -3 \\ 9 & -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 6 & -5 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 & -5 \\ 7 & 5 \end{smallmatrix}\right] \right\} \subseteq \mathcal{M}_{2,2}(\mathbb{R})$.
    (a) Find a subset T of S that forms a basis for W.
    (b) Write each element of $S \setminus T$ as a linear combination of the elements of T.
    (c) Find a basis for $\mathcal{M}_{2,2}(\mathbb{R})$ that contains T.

141. (a) Show that $\mathcal{B} = \{(1,1,1), (1,1,0), (1,0,0)\}$ is a basis for $\mathbb{R}^3$.
    (b) Find the coordinates of $(4, -3, 2) \in \mathbb{R}^3$ relative to $\mathcal{B}$.

142. (a) Show that the set $\{(1-i, i), (2, -1+i)\}$ is linearly dependent in $\mathbb{C}^2$
    (b) Now consider the vector space V with underlying set $\mathbb{C}^2$, but with $\mathbb{R}$ as the field of scalars. Show that the set $\{(1-i, i), (2, -1+i)\}$ is linearly independent in V

143. Let $A = \begin{bmatrix} 1 & 3 & -2 & 5 & 4 \\ 1 & 4 & 1 & 3 & 5 \\ 1 & 4 & 2 & 4 & 3 \\ 2 & 7 & -3 & 6 & 13 \end{bmatrix} \in \mathcal{M}_{4,5}(\mathbb{R})$
    (a) Find a basis for the solution space of A (as a subspace of $\mathbb{R}^5$).
    (b) Extend to a basis of $\mathbb{R}^5$.
    (c) Find a basis for the column space of A (as a subspace of $\mathbb{R}^4$).

144. Which of the following are linear combinations of $(0, -2, 2)$ and $(1, 3, -1)$?

(a) $(2,2,2)$ (b) $(0,4,5)$

145. Let $u=(1,0,-1)$ and $v=(-2,1,1)$.

 (a) Write $(-1,2,-1)$ as a linear combination of u and v.
 (b) Show that $(-1,1,1)$ cannot be written as a linear combination of u and v.
 (c) For what value of c is the vector $(1,1,c)$ a linear combination of u and v?

146. Is $\begin{bmatrix} 6 & -8 \\ -1 & -8 \end{bmatrix}$ a linear combination of the matrices in $\left\{ \begin{bmatrix} 4 & 0 \\ -2 & -2 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 4 \end{bmatrix} \right\}$?

147. Express the polynomial $-9-7x-15x^2$ as a linear combination of $p_1=2+x+4x^2$, $p_2=1-x+3x^2$, and $p_3=3+2x+5x^2$.

148. Show that the vectors $(1,a,a^2),(1,b,b^2),(1,c,c^2)$ are linearly independent if a,b,c are distinct (i.e., $a\neq b$, $a\neq c$ and $b\neq c$).

149. In this question let $\mathcal{S}=\{v_1,v_2,v_3,v_4,v_5\}\subseteq\mathbb{R}^3$ and let A be the 3×5 matrix with the i-th column given by the vector v_i. Suppose that the reduced row echelon form of A is

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

Are the following sets linearly dependent or independent? If linearly dependent, express one vector as a linear combination of the others.

 (a) $\{v_1,v_2,v_3\}$ (b) $\{v_1,v_3,v_4\}$ (c) $\{v_1,v_4,v_5\}$ (d) $\{v_3,v_4,v_5\}$

150. In each part explain why the given statement is true "by inspection."

 (a) The set $\{(1,0,3),(-1,1,0),(1,2,4),(0,-1,-2)\}$ is linearly dependent.
 (b) The set $\{(1,-1,2),(0,1,1)\}$ does not span $\mathbb{R}^3$.
 (c) If the set $\{v_1,v_2,v_3,v_4\}$ of vectors in $\mathbb{R}^4$ is linearly independent, then it spans $\mathbb{R}^4$.
 (d) The set $\{(0,1,-1,0),(0,-1,2,0)\}$ is linearly independent, and so it spans the subspace of $\mathbb{R}^4$ of all vectors of the form $(0,a,b,0)$.

151. In each part determine whether or not the given set forms a basis for the indicated subspace.

 (a) $\{(1,2,3),(-1,0,1),(0,1,2)\}$ for $\mathbb{R}^3$
 (b) $\{(-1,1,2),(3,3,1),(1,2,2)\}$ for $\mathbb{R}^3$
 (c) $\{(1,-1,0),(0,1,-1)\}$ for the subspace of $\mathbb{R}^3$ consisting of all (x,y,z) such that $x+y+z=0$.
 (d) $\{(1,1,0),(1,1,1)\}$ for the subspace of $\mathbb{R}^3$ consisting of all (x,y,z) such that $y=x+z$.

152. Which of the following sets of vectors are bases for $\mathbb{R}^3$?

 (a) $\{(1,0,0),(2,2,0),(3,3,3)\}$
 (b) $\{(2,-3,1),(4,1,1),(0,-7,1)\}$

153. Find a basis for and the dimension of the subspace of $\mathbb{R}^n$ spanned by the following sets.

 (a) $\{(0,1,-2),\ (3,0,1),\ (3,2,-3)\}$ $(n=3)$
 (b) $\{(1,3),\ (-1,2),\ (7,6)\}$ $(n=2)$
 (c) $\{(-1,2,0,4),\ (3,1,-1,2),\ (-5,3,1,6),\ (7,0,-2,0)\}$ $(n=4)$

154. For each of the following sets choose a *subset* that is a basis for the subspace spanned by the set. Then express each vector that is not in the basis as a linear combination of the basis vectors.

 (a) $\{(1,2,0,-1),\ (2,-1,2,3),\ (-1,-11,6,13),\ (4,3,2,1)\}\subseteq\mathbb{Q}^4$
 (b) $\{(0,-1,-3,3),\ (-1,-1,-3,2),\ (3,1,3,0),\ (0,-1,-2,1)\}\subseteq\mathbb{Q}^4$
 (c) $\{(1,2,-1),\ (0,3,4),\ (2,1,-6),\ (0,0,2)\}\subseteq\mathbb{Q}^3$

LECTURE 18

Linear error-correcting codes

When storing or transmitting data (on a disk, over the internet etc) errors are often introduced. Errors might result, for example, from physical damage or radiation. In a memory chip, background radiation can alter the memory contents. We would like to be able to protect against this, and reduce the risk of using corrupted data.

How can we encode data in order to detect and perhaps correct errors in transmission or storage? The study of such problems is known as *Coding Theory*.

18.1 Codes

A key idea is to build in redundancy before sending or storing the data. A simple way to do this is by repetition. If we wanted to send the message '1011', we could send each bit twice and send '11001111'. If the message '11011111' were received we would know that there had been some corruption of the message. In this example the sent message is made up of combinations of 00 and 11.

By repeating each bit three times we would even be able to correct errors:

message: 1011 **sent:** 111000111111 **received:** 111010111111

We would know that there had been some interference *and* that the original message was (most probably) 101, since 010 is 'closer' to 000 than to 111.

Definition 18.1

Let $\mathcal{A}$ be a finite set. We will refer to $\mathcal{A}$ as the **alphabet**. A **code** over $\mathcal{A}$ is a non-empty subset of $\mathcal{A}^n$. The number n is called the **length** of the code. The elements of a code are called **codewords**.

Example 18.2.

1. The set $\{(c,a,t),(d,o,g),(p,i,g),(a,b,c),(l,d,r)\}$ is a code of length 3 over the alphabet $\mathcal{A}=\{a,b,c,\dots,z\}$.
2. $\{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}$ is a code of length 3 over $\mathbb{F}_2=\{0,1\}$

Remark. When considering codes it is convenient to drop the commas and parentheses. So the examples above would be written as $\{cat, dog, pig, abc, ldr\}$ and $\{000, 011, 101, 110\}$

18.2 Linear codes

Definition 18.3

A **linear code** of length n and rank k is a k-dimensional subspace of $\mathbb{F}_p^n$. (for some prime $p \in \mathbb{N}$). The code is called a **binary linear code** if $p=2$, and it is a **ternary linear code** in the case $p=3$.

Example 18.4. For the above repetition code, our codewords were 000 and 111. These two elements together form a subspace of $\mathbb{F}_2^3$ so it is a linear code.

Example 18.5. $\{(a,b,c) \mid a,b,c \in \mathbb{F}_2 \text{ and } a+b+c=0\} = \{000, 011, 101, 110\}$ is a subspace of $\mathbb{F}_2^3$, and so is a binary linear code. This code could be used as follows.

original	codeword
00	000
01	101
10	110
11	011

If we receive a word abc we can check whether it is a codeword by calculating (remembering that entries are in $\mathbb{F}_2$) whether

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$$

If it is a codeword, we know that the intended (original) message was bc.

Definition 18.6

A **check matrix** for a linear code is a matrix H such that C is the solution space of H.

Example 18.7. (The Hamming Code $\mathcal{H}_2(3)$)

Consider the matrix

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \in \mathcal{M}_{3,7}(\mathbb{F}_2)$$

We take as codewords all the elements in the solution space of H.

A basis for the solution space is given by $\left\{ \begin{bmatrix} 1\\1\\1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\0\\0\\1\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\1\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\0\\1\\0\\0\\1 \end{bmatrix} \right\}$

We then send $abcd \in \mathbb{F}_2^4$ using the codeword

$$\begin{bmatrix} x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7 \end{bmatrix} = a \begin{bmatrix} 1\\1\\1\\0\\0\\0\\0 \end{bmatrix} + b \begin{bmatrix} 1\\0\\0\\1\\1\\0\\0 \end{bmatrix} + c \begin{bmatrix} 0\\1\\0\\1\\0\\1\\0 \end{bmatrix} + d \begin{bmatrix} 1\\1\\0\\1\\0\\0\\1 \end{bmatrix}$$

$$= \begin{bmatrix} 1&1&0&1\\1&0&1&1\\1&0&0&0\\0&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1 \end{bmatrix} \begin{bmatrix} a\\b\\c\\d \end{bmatrix} = \begin{bmatrix} a+b+d\\a+c+d\\a\\b+c+d\\b\\c\\d \end{bmatrix}$$

We have 4 parameters (called ***information bits*** in this setting) which can be chosen arbitrarily. (In this case they are the 3rd, 5th, 6th and 7th bits.) The other 3 bits are called ***check bits***. (The 1st, 2nd and 4th bits.)

If we want to send the message $abcd \in \mathbb{F}_2^4$, we calculate the codeword $\begin{bmatrix} a+b+d & a+c+d & a & b+c+d & b & c & d \end{bmatrix}$ and send it. When a message is received, we check that it is a codeword by multiplication by H.

For example, suppose we wanted to send 1011 (i.e., $a = 1$, $b = 0$, $c = 1$, $d = 1$).

Given the specified encoding, the codeword for this is 0110011. We send the codeword 0110011

Suppose some interference occurs and the received word is $v = 0100011$

The receiver knows that an error has occurred because

$$Hv^T = H \begin{bmatrix} 0\\1\\0\\0\\0\\1\\1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0\\1\\0\\0\\0\\1\\1 \end{bmatrix} = \begin{bmatrix} 0\\1\\1 \end{bmatrix} \neq \begin{bmatrix} 0\\0\\0 \end{bmatrix}$$

In fact, since

$$Hv^T = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

the receiver knows that the error occurred in the third bit.

Assuming that a single error has occurred, the (column) matrix Hv^T is equal to a column (of H). If it is the nth column, the error occurred in the nth bit.

She therefore knows that to get the intended codeword the third bit should be swapped from 0 to 1 — giving 0110011

The original message is then recovered by dropping the check bits (1st, 2nd and 4th) — giving 1011

This code gives a way of correcting a single error, which we explore below.

Exercise 155. Using the above encoding, recover the original messages from the received words: 0110001, 1110000, 0000011

18.3 Hamming distance

When an error occurs, we would like to be able to correct the error, by deciding what the original message was. In the above triple repetition code, when the non codeword 010 was received we deduced that the (probable) intended codeword was 000, since this is 'closer' to 010 than is the other codeword 111. We saw another example in the previous section.

In correcting an error, we are assuming that the original codeword (before interference) is the codeword closest to the received word. This is called the **nearest neighbour principle**.

Definition 18.8

The **Hamming distance** between two strings $u = a_1 \cdots a_n, v = b_1 \cdots b_n \in \mathcal{A}^n$ is the number of places in which they differ. It is denoted $d(u, v)$.

Example 18.9. So, $d(010, 000) = 1$ and $d(010, 111) = 2$

For a code C, an important property is the minimum distance between codewords.

Definition 18.10

The **minimum distance** of a code C is $d_{min} = \min\{d(u, v) \mid u, v \in C, u \neq v\}$

Proposition 18.11: Nearest neighbour principle

Let C be a code with minimum distance d_{min} between codewords. Then C can be used to

1. ***detect*** up to $d_{min} - 1$ errors and
2. ***correct*** up to $\lfloor \frac{d_{min}-1}{2} \rfloor$ errors.

In general, to find the minimum distance between codewords one needs to calculate the distance between every pair of codewords (and then take the minimum). However, for linear codes calculating the minimum distance is simpler than for general codes.

Definition 18.12

The **weight** $w(u)$ of a word $u = a_1 \cdots a_n \in \mathbb{F}_p^n$ is the number of non-zero coordinates, that is, $w(u) = d(u, \vec{0})$.

Lemma 18.13

For a linear code, the minimum distance between codewords is equal to the smallest non-zero weight.

Exercise 156. Write out a proof of the above lemma.

Example 18.14. The linear code $\{0000, 1011, 0110, 1101\} \subset \mathbb{F}_2^4$ has minimum weight 2, and hence minimum distance 2.

Suppose a linear code is defined by a check matrix H. How can we calculate the minimum weight without having to list all the words in the code?

Lemma 18.15: Minimum distance of a linear code

1. For a binary linear code defined by a check matrix H, the minimum weight is the smallest number $r > 0$ such that r columns of H sum to zero.
2. More generally, let $H \in \mathcal{M}_{m,n}(\mathbb{F}_p)$ be the check matrix for a linear code $\mathcal{C}$. Then the minimum weight of $\mathcal{C}$ is equal to the size of the smallest linearly dependent set of columns of H.

Exercise 157. Prove the above lemma.

18.4 Exercises

158. Determine the minimum distance d_{min} between codewords for each of the following binary codes:
    (a) $\{1000, 1011, 0100\}$
    (b) $\{000000, 101010, 010101, 111001, 011110\}$
    (c) For the code in (b), find the distance of each of the following received words from the codewords, and hence decode each of the received words using the nearest neighbour principle: (i) 100010 (ii) 000101 (iii) 000110

159. What is the smallest minimum distance that a code must have in order to correct two errors? How many errors will it detect?

160. Determine whether each of the following sets of codewords forms a binary linear code.
    (a) $\{000, 110, 100\}$
    (b) $\{000, 100, 011, 111\}$
    (c) $\{00000, 01110, 10111, 11001\}$

161. Verify that each of the following sets gives a binary linear code, and find the minimum distance.

(a) $\{00000000, 10101010, 01010101, 11111111\}$

(b) $\{00000, 00111, 01011, 01100, 10011, 10100, 11000, 11111\}$

162. Show that

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

cannot be used as the check matrix for a single-error-correcting code by

(a) writing down a codeword of weight 2, and

(b) determining two non-zero codewords which are distance 2 apart.

163. Consider the binary linear code with check matrix

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

(a) Decode the received words 111001, 110111.

(b) Show that if 111111 is received then more than one error has occurred in transmission. Find two possible codewords which could have been transmitted with two errors occurring, and a codeword which could have been transmitted with three errors.

164. Let

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \in \mathcal{M}_{4,10}(\mathbb{F}_2).$$

This check matrix defines a linear code whose information bits (i.e., parameters) are the 3rd, 5th, 6th, 7th, 9th, and 10th.

(a) Encode the information messages (i) 101101, (ii) 001011.

(b) Decode the received words (i) 0001110101, (ii) 0000101100, (iii) 1011110111, giving if possible the information messages which were sent.

165. Let

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{bmatrix} \in \mathcal{M}_{3,13}(\mathbb{F}_3)$$

Each codeword has 10 information bits and 3 check bits (1st, 2nd and 5th)

(a) Prove that this code has minimum distance 3.

(b) The word $w = 0121002011001$ is received. By calculating Hw^T, show that an error has occurred.

(c) Assuming that only one error has occurred, find the intended codeword.

(d) What was the original 10-digit message?

Extra material for lecture 18

▷ References for further reading about codes

Introduction to coding and information theory, by Steven Roman
Coding Theory : A First Course, by San Ling and Chaoping Xing
Mathematics in computing, by Gerard O'Regan, Chapter 9.

LECTURE 19

Linear transformations

We next consider functions between vector spaces that preserve the vector space structure (addition and scalar multiplication). Such linear transformations arise in many applications and are a central tool in the study of vector spaces.

19.1 Definition of linear transformation

Definition 19.1: Linear transformation

Let V and W be vector spaces over the same field of scalars $\mathbb{F}$. A **linear transformation** from V to W is a function $T\colon V \to W$ that satisfies the following properties:

1. $\forall u, v \in V$, $\quad T(u+v) = T(u) + T(v) \quad$ (T preserves vector addition)
2. $\forall u \in V\ \forall \alpha \in \mathbb{F}$, $\quad T(\alpha u) = \alpha T(u) \quad$ (T preserves scalar multiplication)

Remark.

1. Properties 1 and 2 together are equivalent to the single condition:
$$\forall u, v \in V \ \ \forall \alpha \in \mathbb{F}, \quad T(u + \alpha v) = T(u) + \alpha T(v)$$
2. The condition $T(\alpha u) = \alpha T(u)$ implies that 'lines are mapped to lines (or points)'

Examples 19.2.

1. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (0, 2x + z, -y)$. Then T is a linear transformation, since
$$\begin{aligned} T((a, b, c) + (x, y, z)) &= T(a + x, b + y, c + z) \\ &= (0, 2(a + x) + (c + z), -(b + y)) = (0, 2a + c, -b) + (0, 2x + z, -y) \\ &= T(a, b, c) + T(x, y, z) \end{aligned}$$
and
$$T(\alpha(x, y, z)) = T(\alpha x, \alpha y, \alpha z) = (0, 2\alpha x + \alpha z, -\alpha y) = \alpha(0, 2x + z, -y) = \alpha T(x, y, z)$$
2. The map $D : \mathcal{P}_3(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$, $D(p(x)) = \frac{dp}{dx}$ is a linear transformation
3. Many familiar geometric transformations of the plane $\mathbb{R}^2$ are linear, such as: reflection across a line through the origin, rotation about the origin, projection onto a line through the origin.

Exercise 166. Let $T : V \to W$ be a linear transformation. Show that

(a) $T(\vec{0}_V) = \vec{0}_W$

(b) $\forall v \in V, T(-v) = -T(v)$

Example 19.3 (Translation is not a linear transformation)**.** The function $\tau : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\tau(x, y) = (x + 1, y)$ is *not* a linear transformation since $\tau(\vec{0}) \neq \vec{0}$. More generally, let $n \in \mathbb{N}$ and $t \in \mathbb{R}^n \setminus \{\vec{0}\}$. The function $\tau : \mathbb{R}^n \to \mathbb{R}^n$ given by $\tau(u) = u + t$ is not a linear transformation.

Exercise 167. Let $T : V \to W$ be a linear transformation.

(a) Suppose that $U \leqslant V$ is a subspace of V. Prove that the **image** of U, $T(U) = \{T(u) \mid u \in U\}$, is a subspace of W.

(b) Suppose that $U \leqslant W$ is a subspace of W. Prove that the **pre-image** of U, $T^{-1}(U) = \{v \in V \mid T(v) \in U\}$, is a subspace of V.
(Note that the notation $T^{-1}(U)$ for the pre-image does not mean that we are assuming that the linear transformation is invertible.)

19.2 The linear transformations determined by a matrix

There is a natural way to use matrices to define linear transformations. In fact, as we will see later, all linear transformations (between finite-dimensional vector spaces) can be represented by matrices.

Lemma 19.4: linear transformation given by a matrix

Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$. The function $T : \mathcal{M}_{n,1}(\mathbb{F}) \to \mathcal{M}_{m,1}(\mathbb{F})$ given by

$$T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

is a linear transformation.

Proof. We need to show that given any $u, v \in \mathcal{M}_{n,1}(\mathbb{F})$ and $\alpha \in \mathbb{F}$ we have $T(u + v) = T(u) + T(v)$ and $T(\alpha u) = \alpha T(u)$.

$$\begin{aligned} T(u+v) &= A(u+v) = Au + Av && \text{(matrix multiplication is distributive)} \\ &= T(u) + T(v) \\ T(\alpha u) &= A(\alpha u) = \alpha A(u) && \text{(property of matrix multiplication)} \\ &= \alpha T(u) \end{aligned}$$

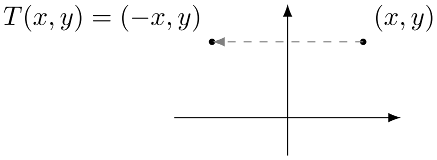
□

Examples 19.5 (Some 'geometric' transformations of the Euclidean plane)**.**

Using coordinates with respect to the standard basis, we identify $\mathcal{M}_{2,1}(\mathbb{R})$ with $\mathbb{R}^2$ in the following.

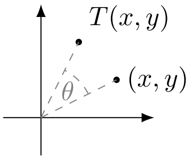
1. The linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix on the right is **reflection** across the y-axis.

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

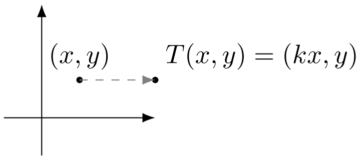


2. The linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix on the right is **rotation** by θ radians (anticlockwise) about the origin.

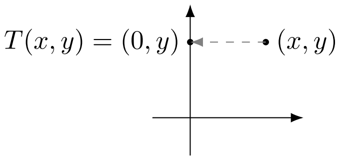
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



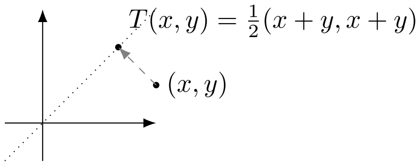
3. The linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix on the right is **dilation** by factor $k > 0$ along the x-axis. $\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$



4. The linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix on the right is **orthogonal projection** onto the y-axis. $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$



5. The linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix on the right is **orthogonal projection** onto the line $y = x$. $\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$



19.3 Linear transformations and bases

Linear transformations are completely determined by their effect on a spanning set.

Proposition 19.6

Let $T_1, T_2 : V \to W$ be two linear transformations and let $S \subseteq V$ be a spanning set for V.

If $T_1(u) = T_2(u)$ for all $u \in S$, then $T_1 = T_2$.

Proof. We need to show that $\forall v \in V$, $T_1(v) = T_2(v)$. Let $v \in V$. Then, since S is a spanning set for V, there exist $\alpha_i \in \mathbb{F}$ and $u_i \in S$ such that $v = \sum_{i=1}^{k} \alpha_i u_i$. Then we have

$$\begin{aligned} T_1(v) &= T_1\left(\sum_{i=1}^{k} \alpha_i u_i\right) \\ &= \sum_{i=1}^{k} \alpha_i T_1(u_i) && (T_1 \text{ is a linear transformation}) \\ &= \sum_{i=1}^{k} \alpha_i T_2(u_i) && (T_1(u_i) = T_2(u_i)) \\ &= T_2\left(\sum_{i=1}^{k} \alpha_i u_i\right) && (T_2 \text{ is a linear transformation}) \\ &= T_2(v) \end{aligned}$$

□

The next result says that a linear transformation can be defined by choosing images for the elements of a basis.

Theorem 19.7

Let V and W be vector spaces over the same field $\mathbb{F}$. Let $\mathcal{B} \subseteq V$ be a basis for V.

Given any function $f : \mathcal{B} \to W$, there exists a unique linear transformation $T : V \to W$ having the property that $T(b) = f(b)$ for all $b \in \mathcal{B}$.

Proof. We need to show that there exists a linear transformation with the property that $T(b_i) = f(b_i)$ for all i, and that if two linear transformations each have this property, then they are equal.

To establish the existence part of the statement we define a function $T : V \to W$ as follows. Given $u \in V$, we have $u = \alpha_1 b_1 + \cdots + \alpha_n b_n$ for uniquely determined $\alpha_i \in \mathbb{F}$ and $b_i \in \mathcal{B}$ (Lemma 14.3). We define $T(u) = \alpha_1 f(b_1) + \cdots + \alpha_n f(b_n)$. To see that this gives a linear transformation, let $u, v \in V$ and $\alpha \in F$. Then there are $b_1, \ldots, b_n \in \mathcal{B}$ and $\alpha_i, \beta_i \in \mathbb{F}$ such that $u = \alpha_1 b_1 + \cdots + \alpha_n b_n$ and $v = \beta_1 b_1 + \cdots + \beta_n b_n$. Then

$$\begin{aligned}
T(u+v) &= T(\alpha_1 b_1 + \cdots + \alpha_n b_n + \beta_1 b_1 + \cdots + \beta_n b_n) \\
&= T((\alpha_1 + \beta_1) b_1 + \cdots + (\alpha_n + \beta_n) b_n) \\
&= (\alpha_1 + \beta_1) f(b_1) + \cdots + (\alpha_n + \beta_n) f(b_n) && \text{(definition of } T\text{)} \\
&= \alpha_1 f(b_1) + \cdots + \alpha_n f(b_n) + \beta_1 f(b_1) + \cdots + \beta_n f(b_n) \\
&= T(\alpha_1 b_1 + \cdots + \alpha_n b_n) + T(\beta_1 b_1 + \cdots + \beta_n b_n) && \text{(definition of } T\text{)} \\
&= T(u) + T(v)
\end{aligned}$$

and

$$\begin{aligned}
T(\alpha u) &= T(\alpha(\alpha_1 b_1 + \cdots + \alpha_n b_n)) \\
&= T((\alpha\alpha_1) b_1 + \cdots + (\alpha\alpha_n) b_n) \\
&= (\alpha\alpha_1) f(b_1) + \cdots + (\alpha\alpha_n) f(b_n) && \text{(definition of } T\text{)} \\
&= \alpha(\alpha_1 f(b_1) + \cdots + \alpha_n f(b_n)) \\
&= \alpha T(\alpha_1 b_1 + \cdots + \alpha_n b_n) && \text{(definition of } T\text{)} \\
&= \alpha T(u)
\end{aligned}$$

The uniqueness part of the theorem is an immediate consequence of Proposition 19.6 □

19.4 Exercises

168. Show that each of the following maps is a linear transformation:

    (a) $S : \mathbb{R}^2 \to \mathbb{R}^2$, $S(x, y) = (2x - y, x + y)$

    (b) $T : \mathbb{R}^3 \to \mathcal{M}_{2,2}(\mathbb{R})$ given by $T(x, y, z) = \begin{bmatrix} y & z \\ -x & 0 \end{bmatrix}$

169. Determine whether or not the given map is a linear transformation, and justify your answer.

    (a) $F : \mathbb{R}^3 \to \mathbb{R}^2$, $F(x, y, z) = (0, 2x + y)$

    (b) $K : \mathbb{R}^2 \to \mathbb{R}^3$, $K(x, y) = (x, \sin y, 2x + y)$

170. Let v_1, v_2, and v_3 be vectors in a vector space V and $T : V \to \mathbb{R}^3$ a linear transformation for which $T(v_1) = (1, -1, 2)$, $T(v_2) = (0, 3, 2)$, and $T(v_3) = (-3, 1, 2)$. Find $T(2v_1 - 3v_2 + 4v_3)$.

171. For the linear transformations of $\mathbb{R}^2$ into $\mathbb{R}^2$ given by the following matrices:

    (i) Sketch the image of the rectangle with vertices $(0, 0), (2, 0), (0, 1), (2, 1)$.

    (ii) Describe the geometric effect of the linear transformation.

    (a) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} b & 0 \\ 0 & c \end{bmatrix}$

    (b) $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix}$ (f) $\frac{1}{5}\begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$

172. Show that there is no line through the origin in $\mathbb{R}^2$ that is invariant under the transformation determined by the matrix

$$A(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

when θ is not an integral multiple of π. Give a geometric interpretation of this observation commenting on the case when $\theta = k\pi$ for some $k \in \mathbb{Z}$.

173. Let V and W be two vector spaces over a field $\mathbb{F}$. Let $S \subseteq V$ be non-empty and linearly independent. Use Theorem 19.7 to show that for all functions $f : S \to W$, there exists a linear transformation $T : V \to W$ with the property that $T(v) = f(v)$ for all $v \in S$.

Extra material for lecture 19

▷ $\mathrm{Hom}(V, W)$

Let V and W be vector spaces over the same field $\mathbb{F}$. The set of all linear transformations from V to W is itself a vector space over $\mathbb{F}$ when given the usual 'pointwise' operations.

$$\begin{aligned}(S+T)(v) &= S(v)+T(v)\\ (kT)(v) &= kT(v)\end{aligned}$$

This vector space is denoted $\mathrm{Hom}(V, W)$.

LECTURE 20

Matrix representations of linear transformations

We saw in Lemma 19.4 that matrices can be used to define linear transformations. In fact, *any* linear transformation can be represented by a matrix. Just as the coordinate matrix of a vector depends on a choice of basis, the matrix of a linear transformation depends on a choice of basis for each of the domain and codomain.

20.1 Matrix of a linear transformation

Let V, W be finite dimensional vector spaces with the same scalars $\mathbb{F}$ and let $T\colon V \to W$ be a linear transformation. Let $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ be an ordered basis for V and $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$ be an ordered basis for W. Then $T(b_i) \in W$ for each $i = 1, \ldots, m$ and we can therefore write $T(b_i)$ uniquely as a linear combination of the basis vectors in $\mathcal{C}$.

$$\begin{array}{ccccccccc}
T(b_1) & = & \alpha_{11}c_1 & + & \alpha_{21}c_2 & + & \ldots & + & \alpha_{m1}c_m \\
T(b_2) & = & \alpha_{12}c_1 & + & \alpha_{22}c_2 & + & \ldots & + & \alpha_{m2}c_m \\
\vdots & \vdots & \vdots & & \vdots & & & & \vdots \\
T(b_n) & = & \alpha_{1n}c_1 & + & \alpha_{2n}c_2 & + & \ldots & + & \alpha_{mn}c_m
\end{array}$$

Definition 20.1: matrix of a linear transformation

We form a matrix $[T]_{\mathcal{C},\mathcal{B}} \in \mathcal{M}_{m,n}(\mathbb{F})$ by defining $[T]_{\mathcal{C},\mathcal{B}} = (\alpha_{ij})$. This matrix is called the **matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$**.

Note. The i-th column of $[T]_{\mathcal{C},\mathcal{B}}$ is given by $[T(b_i)]_{\mathcal{C}}$. That is,

$$[T]_{\mathcal{C},\mathcal{B}} = [\, [T(b_1)]_{\mathcal{C}} \cdots [T(b_n)]_{\mathcal{C}} \,]$$

Remark. In the case in which $V = W and \mathcal{B} = \mathcal{C}$, we sometimes write $[T]_{\mathcal{B}}$ in place of $[T]_{\mathcal{B},\mathcal{B}}$.

The way in which this matrix represents the linear transformation is given by the following.

Lemma 20.2

Let V and W be finite dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$ respectively. Let $T : V \to W$ be a linear transformation. Then

$$\forall u \in V, \quad [T(u)]_{\mathcal{C}} = [T]_{\mathcal{C},\mathcal{B}}\,[u]_{\mathcal{B}}$$

Proof. Let $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ and $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$. Let $u \in V$. Then $u = \alpha_1 b_1 + \cdots + \alpha_n b_n$ for some $\alpha_i \in \mathbb{F}$. We have

$$\begin{aligned}
u &= \alpha_1 b_1 + \cdots + \alpha_m b_n \\
T(u) &= \alpha_1 T(b_1) + \cdots + \alpha_n T(b_n) && (T \text{ is linear}) \\
[T(u)]_{\mathcal{C}} &= \alpha_1 [T(b_1)]_{\mathcal{C}} + \cdots + \alpha_n [T(b_n)]_{\mathcal{C}} && (\text{Lemma 15.4}) \qquad (*)
\end{aligned}$$

and

$$\begin{aligned}
[T]_{\mathcal{C},\mathcal{B}}[u]_{\mathcal{B}} &= \big[[T(b_1)]_{\mathcal{C}} \quad \cdots \quad [T(b_n)]_{\mathcal{C}}\big] \begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{bmatrix} \\
&= \alpha_1[T(b_1)]_{\mathcal{C}} + \cdots + \alpha_n[T(b_n)]_{\mathcal{C}} && \text{(see the remark on page 16-1)} \\
&= [T(u)]_{\mathcal{C}} && \text{(from } (*) \text{ above)}
\end{aligned}$$

□

In summary, we have

$$\begin{tikzcd}
u \in V \arrow[r, "\text{apply } T"] \arrow[d, "\text{take coords}"'] & T(u) \in W \arrow[d, "\text{take coords}"] \\
{[u]_{\mathcal{B}}} \arrow[r, "\text{mult by } {[T]_{\mathcal{C},\mathcal{B}}}"] & {[T(u)]_{\mathcal{C}}}
\end{tikzcd}$$

Exercise 174. Suppose that $A \in \mathcal{M}_{m,n}(\mathbb{F})$ is such that $A[u]_{\mathcal{B}} = [T(u)]_{\mathcal{C}}$ for all $u \in V$. Show that $A = [T]_{\mathcal{C},\mathcal{B}}$. (Hint: Show that the i-th column of A is equal to $[T(b_i)]_{\mathcal{C}}$.)

Example 20.3. Consider the linear transformation $T\colon \mathcal{M}_{2,2}(\mathbb{R}) \to \mathcal{M}_{2,2}(\mathbb{R})$ where T is defined by

$$T(A) = A^T$$

Let's calculate the matrix representation of T with respect to the basis $\mathcal{S} = \{\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]\}$ (for both domain and codomain).

$$[T\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]]_{\mathcal{S}} = [\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]]_{\mathcal{S}} = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \qquad [T\left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right]]_{\mathcal{S}} = [\left[\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right]]_{\mathcal{S}} = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$$

$$[T\left[\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right]]_{\mathcal{S}} = [\left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right]]_{\mathcal{S}} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \qquad [T\left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]]_{\mathcal{S}} = [\left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]]_{\mathcal{S}} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

Therefore $[T]_{\mathcal{S},\mathcal{S}} = \left[\begin{smallmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{smallmatrix}\right]$

Exercise 175. With T as above and $\mathcal{C} = \{\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]\}$, calculate $[T]_{\mathcal{C},\mathcal{C}}$.

Example 20.4. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(x, y) = (x + 4y, x + y)$. Let $\mathcal{S} = \{(1, 0), (0, 1)\}$ and $\mathcal{B} = \{(2, -1), (2, 1)\}$. Then

$$[T]_{\mathcal{S},\mathcal{S}} = \left[\begin{smallmatrix}1&4\\1&1\end{smallmatrix}\right] \quad \text{and} \quad [T]_{\mathcal{B},\mathcal{B}} = \left[\begin{smallmatrix}-1&0\\0&3\end{smallmatrix}\right] \quad \text{and} \quad [T]_{\mathcal{S},\mathcal{B}} = \left[\begin{smallmatrix}-2&6\\1&3\end{smallmatrix}\right]$$

Exercise 176. A linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ has matrix $[T] = \left[\begin{smallmatrix}5&1&0\\1&5&-2\end{smallmatrix}\right]$ with respect to the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$. What is the matrix of T with respect to the bases $\mathcal{B} = \{(1, 1, 0), (1, -1, 0), (1, -1, -2)\}$ of $\mathbb{R}^3$ and $\mathcal{C} = \{(1, 1), (1, -1)\}$ of $\mathbb{R}^2$?

Lemma 20.5

Let U, V, W be finite dimensional vector spaces with bases $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ respectively. Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then

$$[T \circ S]_{\mathcal{C},\mathcal{A}} = [T]_{\mathcal{C},\mathcal{B}}\,[S]_{\mathcal{B},\mathcal{A}}$$

Proof. Let $u \in U$.

$$\begin{aligned}
[T \circ S]_{\mathcal{C},\mathcal{A}}[u]_{\mathcal{A}} &= [T \circ S(u)]_{\mathcal{C}} && \text{Lemma 20.2} \\
&= [T(S(u))]_{\mathcal{C}} \\
&= [T]_{\mathcal{C},\mathcal{B}}[S(u)]_{\mathcal{B}} && \text{Lemma 20.2} \\
&= ([T]_{\mathcal{C},\mathcal{B}}[S]_{\mathcal{B},\mathcal{A}})[u]_{\mathcal{A}} && \text{Lemma 20.2}
\end{aligned}$$

Since $[T \circ S]_{\mathcal{C},\mathcal{A}}[u]_{\mathcal{A}} = ([T]_{\mathcal{C},\mathcal{B}}[S]_{\mathcal{B},\mathcal{A}})[u]_{\mathcal{A}}$ for all $u \in U$, we must have that $[T \circ S]_{\mathcal{C},\mathcal{A}} = [T]_{\mathcal{C},\mathcal{B}}[S]_{\mathcal{B},\mathcal{A}}$ □

Example 20.6. With the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ and bases $\mathcal{S}$ and $\mathcal{B}$ of Example 20.4, we have

$$\begin{aligned}
[T^2]_{\mathcal{S},\mathcal{S}} &= [T]_{\mathcal{S},\mathcal{S}}[T]_{\mathcal{S},\mathcal{S}} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 2 & 5 \end{bmatrix} \\
[T^2]_{\mathcal{B},\mathcal{B}} &= [T]_{\mathcal{B},\mathcal{B}}[T]_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix} \\
[T^2]_{\mathcal{S},\mathcal{B}} &= [T]_{\mathcal{S},\mathcal{S}}[T]_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -2 & 6 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 18 \\ -1 & 9 \end{bmatrix}
\end{aligned}$$

20.2 Kernel and image of a linear transformation

Definition 20.7: Kernel and Image

Let $T\colon V \to W$ be a linear transformation. The **kernel** (or **nullspace**) of T is defined to be

$$\ker(T) = \{u \in V \mid T(u) = \vec{0}\}$$

The **image** of T is defined to be

$$\operatorname{im}(T) = T(V) = \{w \in W \mid w = T(u) \text{ for some } u \in V\}$$

Note. From Exercise 167 we know that $\ker(T) \leqslant V$ and $\operatorname{im}(T) \leqslant W$.

Definition 20.8: rank and nullity of a linear transformation

The dimension of $\ker(T)$ is called the **nullity** of T and is denoted $\operatorname{nullity}(T)$. The dimension of $\operatorname{im}(T)$ is called the **rank** of T and is denoted $\operatorname{rank}(T)$.

Example 20.9. Consider the linear transformation $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ given by differentiation. Then $\ker(T) = \operatorname{span}\{1\}$ and $\operatorname{im}(T) = \mathcal{P}_2(\mathbb{R})$. Hence $\operatorname{nullity}(T) = 1$ and $\operatorname{rank}(T) = 3$.

Lemma 20.10

Let $T\colon V \to W$ be a linear transformation. Then T is injective if and only if $\ker(T) = \{\vec{0}\}$.

Proof. Suppose first that T is injective. Then we have

$$u \in \ker(T) \iff T(u) = \vec{0} \iff T(u) = T(\vec{0}) \iff u = \vec{0}$$

Therefore $\ker(T) = \{\vec{0}\}$.

Now, conversely, suppose that $\ker(T) = \{\vec{0}\}$. For $u, v \in V$ we have

$$T(u) = T(v) \implies T(u) - T(v) = \vec{0} \implies T(u-v) = \vec{0} \implies u - v \in \ker(T) \implies u - v = \vec{0} \implies u = v$$

Therefore T is injective. □

Exercise 177. Let $T\colon V \to W$ be a linear transformation.

(a) Let $X \subseteq V$ be a linearly independent subset of the domain. Show that if T is injective, then $T(X)$ is linearly independent.

(b) Let $Y \subseteq V$ be a spanning set for the domain. Show that $T(Y)$ is a spanning set for the image $\operatorname{im}(T)$.

(c) Use parts (a) and (b) to show that if $\mathcal{B}$ is a basis for V and T is injective, then $T(\mathcal{B})$ is a basis for $\operatorname{im}(T)$.

20.3 Rank-nullity theorem

If both the domain and codomain are finite dimensional, the kernel and image of a linear transformation T can calculated from a matrix representation of T:

$$\begin{aligned} u \in \ker(T) &\iff [u]_{\mathcal{B}} \in \text{Solution space of } [T]_{\mathcal{C},\mathcal{B}} \\ w \in \operatorname{im}(T) &\iff [w]_{\mathcal{C}} \in \operatorname{colspace}([T]_{\mathcal{C},\mathcal{B}}) \end{aligned}$$

Therefore,

$$\begin{aligned} \operatorname{nullity}(T) &= \dim(\operatorname{solspace}([T]_{\mathcal{C},\mathcal{B}})) \\ \operatorname{rank}(T) &= \operatorname{rank}([T]_{\mathcal{C},\mathcal{B}}) \end{aligned}$$

The following is essentially the observation that each column of a matrix is either a pivot column or a non-pivot column.

Theorem 20.11: Rank-Nullity Theorem

Let $T : V \to W$ be a linear transformation. Suppose that V is finite dimensional. Then

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V)$$

Proof. We first prove under the assumption that W is also finite dimensional. Let $n = \dim(V)$ and $m = \dim(W)$ and let $\mathcal{B}$ be a basis for V and $\mathcal{C}$ a basis for W. Let $A = [T]_{\mathcal{C},\mathcal{B}} \in \mathcal{M}_{m,n}(\mathbb{F})$.

$$\begin{aligned} \operatorname{nullity}(T) &= \dim(\text{solution space of } A) = \text{number of non-pivot columns of } A \\ \operatorname{rank}(T) &= \dim(\text{column space of } A) = \text{number of pivot columns of } A \\ \dim(V) &= n = \text{number of columns of } A \end{aligned}$$

Therefore $\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V)$.

For the general case (without assuming that W is finite dimensional) we define $W' = \operatorname{im}(T)$. Then W' is finite dimensional, and we apply the previous part to the linear transformation $T' : V \to W'$ given by $T'(u) = T(u)$ to conclude that $\operatorname{rank}(T') + \operatorname{nullity}(T') = \dim(V)$. Then note that $\ker(T') = \ker(T)$ and $\operatorname{im}(T') = \operatorname{im}(T) = W'$. □

Example 20.12. For the linear transformation $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ of Example 20.9 we have $\operatorname{rank}(T) + \operatorname{nullity}(T) = 3 + 1 = 4 = \dim(\mathcal{P}_3(\mathbb{R}))$.

20.4 Exercises

178. Find the standard matrix of the following linear transformations of $\mathbb{R}^2$.

    (a) rotation by $\frac{3\pi}{4}$
    (b) rotation by $-\frac{\pi}{2}$
    (c) reflection in the line $y = x$
    (d) reflection in the $x-$axis

179. In each part, find a single matrix that performs the indicated succession of operations.

    (a) Compresses by a factor of $\frac{1}{2}$ in the x-direction, then expands by a factor of 5 in the y-direction.
    (b) Reflects about $y = x$, then rotates about the origin through an angle of π.
    (c) Reflects about the y-axis, then expands by a factor of 5 in the x-direction, and then reflects about $y = x$.

180. Find the standard matrix (i.e., with respect to the standard basis of $\mathbb{R}^2$) of the rotation about the origin through

    (a) $\frac{\pi}{4}$ anticlockwise
    (b) π

181. Consider the following linear transformations:

    $K : \mathbb{R}^3 \to \mathbb{R}^3$ $K(x, y, z) = (x, x+y, x+y+z)$ $L : \mathbb{R}^3 \to \mathbb{R}^2$ $L(x, y, z) = (2x - y, x + 2y)$
    $S : \mathbb{R}^3 \to \mathbb{R}^3$ $S(x, y, z) = (z, y, x)$ $T : \mathbb{R}^2 \to \mathbb{R}^4$ $T(x, y) = (2x+y, x+y, x-y, x-2y)$

    Find the matrix that represents each of the following linear transformations (with respect to the standard bases).

    (a) K (b) L (c) S (d) T

182. Find the indicated linear transformation if it is defined. If it is not defined, explain why not.

    (a) LK (b) TL (c) S^2 (d) $K + S$ (e) T^2

    (Where, for example, LK denotes the composition $L \circ K$.)

183. Find the matrix which represents (with respect to the standard bases) those linear transformations in question 182 which exist.

184. Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be the function defined by multiplication by x. That is, $T(a+bx+cx^2) = ax + bx^2 + cx^3$.

    (a) Show that T is a linear transformation.
    (b) Find the matrix of T with respect to the standard bases $\mathcal{B} = \{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{R})$ and $\mathcal{C} = \{1, x, x^2, x^3\}$ for $\mathcal{P}_3(\mathbb{R})$.

185. Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the linear transformation defined by $T(p(x)) = p(2x + 1)$, that is,

    $$T(a_0 + a_1x + a_2x^2) = a_0 + a_1(2x + 1) + a_2(2x + 1)^2$$

    Find $[T]_{\mathcal{B}}$, where $\mathcal{B} = \{1, x, x^2\}$.

186. Find the matrix that represents the linear transformation T with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$.

    (a) $T : \mathbb{R}^3 \to \mathcal{M}_{2,2}(\mathbb{R})$ given by

    $$T(x, y, z) = \begin{bmatrix} y & z \\ -x & 0 \end{bmatrix}$$

    where $\mathcal{B} = \{e_1, e_2, e_3\}$ and $\mathcal{B}' = \{E^{ij} | i = 1, 2; j = 1, 2\}$ (i.e. the standard bases).

(b) $T : \mathcal{P}_3(\mathbb{C}) \to \mathcal{P}_3(\mathbb{C})$ given by

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = (a_0 + a_2) - (a_1 + 2a_3)x^2$$

where $\mathcal{B} = \mathcal{B}' = \{1, x, x^2, x^3\}$.

187. Consider the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ given by the matrix (wrt the standard bases):

$$[T] = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & -4 & 6 & 2 \end{bmatrix}$$

(a) Determine whether or not $v_1 = (-2, 0, 0, 2)$ and $v_2 = (-2, 2, 2, 0)$ are in the kernel of T.

(b) Determine whether or not $w_1 = (1, 3, 1)$ and $w_2 = (-1, -1, -2)$ are in the image of T.

(c) Find the nullity of T and give a basis for the kernel of T. Is the transformation injective?

(d) Find the rank of T and give a basis for the image of T. Is the transformation surjective?

188. For each of the linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ given below find:

(i) its standard matrix (i.e., with respect to the standard bases),

(ii) a basis for the kernel,

(iii) a basis for the image.

(a) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + y \\ 3y \end{bmatrix}$ (b) $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 - x_3 \\ 2x_1 + x_2 \end{bmatrix}$

(c) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + 2y \\ -y \\ x - y \end{bmatrix}$ (d) $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 - x_2 - 6x_3 \\ -2x_1 + x_2 + 5x_3 \\ 3x_1 + 3x_2 + 6x_3 \end{bmatrix}$

189. Let $T : \mathcal{M}_{2,2}(\mathbb{R}) \to \mathcal{M}_{1,2}(\mathbb{R})$ be the map defined by

$$T\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 2a_{11} - a_{21} & 2a_{12} - a_{22} \end{bmatrix}$$

(a) Show that T is a linear transformation.

(b) Find bases for the kernel and image of T. Deduce the rank and nullity of T.

(c) Find the matrix of T with respect to the standard bases of $\mathcal{M}_{2,2}(\mathbb{R})$ and $\mathcal{M}_{1,2}(\mathbb{R})$.

190. Let $S : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be defined as follows. For each $p(x) = a_0 + a_1x + a_2x^2$, define $S(p) = a_0x + \frac{1}{2}a_1x^2 + \frac{1}{3}a_2x^3$. The linear transformation S gives the antiderivative of $p(x)$, with the constant term equal to zero.

(a) Find the matrix A that represents S with respect to the bases $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{B}' = \{1, x, x^2, x^3\}$

(b) Use A to find the antiderivative of $p(x) = 1 - x + 2x^2$.

191. Let U, V, W be vector spaces over the same field $\mathbb{F}$ with V being fininte dimensional. Consider linear transformations $S : U \to V$ and $T : V \to W$ that satisfy $T \circ S = 0$. Show that $\operatorname{rank}(S) + \operatorname{rank}(T) \leqslant \dim(V)$.

Extra material for lecture 20

▷ Let V and W be finite dimensional vector spaces over the same field $\mathbb{F}$. Let $n = \dim(V)$ and $m = \dim(W)$. Choose bases $\mathcal{B}$ for V and $\mathcal{C}$ for W and define $f : \mathrm{Hom}(V, W) \to \mathcal{M}_{m,n}(\mathbb{F})$ by $f(T) = [T]_{\mathcal{C},\mathcal{B}}$. Then f is a bijective linear transformation.

▷ Short exact sequences

A sequence of two linear transformations of the form

$$\begin{tikzcd} U \arrow[r, "\varphi"] & V \arrow[r, "\psi"] & W \end{tikzcd}$$

is said to be **exact at V** if $\mathrm{im}(\varphi) = \ker(\psi)$. A **short exact sequence** is a sequence of linear transformations of the form

$$\begin{tikzcd} 0 \arrow[r] & U \arrow[r, "\varphi"] & V \arrow[r, "\psi"] & W \arrow[r] & 0 \end{tikzcd} \tag{$*$}$$

that is exact at each of U, V, and W.

Show that:

(a) the sequence $(*)$ is exact at U iff φ is injective;

(b) the sequence $(*)$ is exact at W iff ψ is surjective;

(c) if the sequence $(*)$ is exact, then $\dim(V) = \dim(U) + \dim(W)$.

▷ Commutative diagrams

A diagram of four linear transformations of the form

$$\begin{tikzcd} U \arrow[r, "\varphi"] \arrow[d, "f"'] & V \arrow[d, "g"] \\ U' \arrow[r, "\varphi'"] & V' \end{tikzcd}$$

is said to **commute** if $g \circ \varphi = \varphi' \circ f$.

Suppose that in the following diagram of linear transformations the rows are exact and the left square commutes.

$$\begin{tikzcd} 0 \arrow[r] & U \arrow[r, "\varphi"] \arrow[d, "f"] & V \arrow[r, "\psi"] \arrow[d, "g"] & W \arrow[r] \arrow[d, dashed, "h"] & 0 \\ 0 \arrow[r] & U' \arrow[r, "\varphi'"] & V' \arrow[r, "\psi'"] & W' \arrow[r] & 0 \end{tikzcd}$$

Show that:

(d) There exists a unique linear transformation $h : W \to W'$ that makes the right square commute;

(e) If g is surjective, then so too is h;

(f) If f is surjective and g is injective, then h is injective.

LECTURE 21

Change of basis

The matrix of linear transformation depends on the bases used for both domain and codomain. We want to develop a convenient way of relating the different matrices that represent a given linear transformation. In other words, if we change the bases used, how does the matrix representation change?

21.1 Invertible linear transformations

Definition 21.1: invertible linear transformation

A linear transformation $T : V \to W$ is called **invertible** if there exists a linear transformation $T^{-1} : W \to V$ satisfying $T \circ T^{-1} = \mathrm{Id}_W$ and $T^{-1} \circ T = \mathrm{Id}_V$.

Invertible linear transformations are also called **isomorphisms**. We say that two vector spaces V and W are **isomorphic** if there exists an isomorphism $T : V \to W$. We write $V \cong W$ to denote that V and W are isomorphic.

Exercise 192. Show that a linear transformation $T : V \to W$ is invertible if and only if T is a bijection.

Note that in the above definition and exercise we are not assuming that either V or W is finite dimensional. In the case in which they are both finite dimensional we have the following.

Proposition 21.2: Matrix of an invertible linear transformation

Let V and W be finite dimensional vector spaces and let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W respectively. Let $T : V \to W$ be a linear transformation.

1. T is inveritble if and only if $[T]_{\mathcal{C},\mathcal{B}}$ is invertible
2. If T is invertible, then $[T^{-1}]_{\mathcal{B}.\mathcal{C}} = ([T]_{\mathcal{C},\mathcal{B}})^{-1}$

Proof. Suppose first that T is invertible. Then T is a bijection and therefore $\ker(T) = \{\vec{0}\}$. Therefore, by the rank-nullity theorem, $\dim(\mathrm{im}(T)) = \dim(V)$. But since T is a bijection, $\mathrm{im}(T) = W$. Therefore $\dim(W) = \dim(V)$.

$$\begin{aligned} T \circ T^{-1} = \mathrm{Id}_W &\implies [T \circ T^{-1}]_{\mathcal{C},\mathcal{C}} = [\mathrm{Id}_W]_{\mathcal{C},\mathcal{C}} && \\ &\implies [T]_{\mathcal{C},\mathcal{B}}\,[T^{-1}]_{\mathcal{B},\mathcal{C}} = [\mathrm{Id}_W]_{\mathcal{C},\mathcal{C}} && \text{Lemma 20.5} \\ &\implies [T]_{\mathcal{C},\mathcal{B}}\,[T^{-1}]_{\mathcal{B},\mathcal{C}} = I_n && \text{where } n = \dim(W) \end{aligned}$$

Similarly,

$$[T^{-1}]_{\mathcal{B},\mathcal{C}}\,[T]_{\mathcal{C},\mathcal{B}} = I_n \qquad \text{where } n = \dim(V)$$

It follows that $[T^{-1}]_{\mathcal{B},\mathcal{C}}$ is the inverse of $[T]_{\mathcal{C},\mathcal{B}}$.

Suppose instead that $A = [T]_{\mathcal{C},\mathcal{B}}$ is invertible. Since it is invertible, it is square and therefore $\dim(V) = \dim(W)$. Let $n = \dim(V)$ and $A = [T]_{\mathcal{C},\mathcal{B}} \in \mathcal{M}_{n,n}(\mathbb{F})$. Let $S : W \to V$ be the linear transformation defined by $[S]_{\mathcal{B},\mathcal{C}} = A^{-1}$. Then $[T \circ S]_{\mathcal{C},\mathcal{C}} = [T]_{\mathcal{C},\mathcal{B}}\,[S]_{\mathcal{B},\mathcal{C}} = A\,A^{-1} = I_n$. Therefore $T \circ S = \mathrm{Id}_W$.

Similarly, $[S \circ T]_{\mathcal{B}.\mathcal{B}} = [S]_{\mathcal{B},\mathcal{C}}\,[T]_{\mathcal{C},\mathcal{B}} = A^{-1}A = I_n$, and therefore $S \circ T = \mathrm{Id}_V$. Therefore T is invertible, and $T^{-1} = S$. □

Exercise 193. Let V be an n-dimensional $\mathbb{F}$-vector space and $\mathcal{B}$ a basis for V. Show that the map $\varphi : V \to \mathcal{M}_{n,1}(\mathbb{F})$, $\varphi(u) = [u]_{\mathcal{B}}$ is an isomorphism. Conclude that $V \cong \mathbb{F}^n$.

21.2 Change of basis for vectors

How can we convert coordinates with respect to one basis to coordinates with respect to another?

Definition 21.3

Let V be a finite dimensional vector space and let $\mathcal{B}$ and $\mathcal{C}$ be two bases for V. The **transition matrix** from $\mathcal{B}$ to $\mathcal{C}$, denoted $P_{\mathcal{C},\mathcal{B}}$, is defined to be

$$P_{\mathcal{C},\mathcal{B}} = [\mathrm{Id}_V]_{\mathcal{C},\mathcal{B}}$$

Remark. If $\mathcal{B} = \{b_1, \ldots, b_n\}$, then $P_{\mathcal{C},\mathcal{B}} = [\,[b_1]_{\mathcal{C}} \cdots [b_n]_{\mathcal{C}}\,]$

Exercise 194. Let V be a finite dimensional vector space and let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be bases for V. Show that

(a) $P_{\mathcal{C},\mathcal{B}}$ is invertible and $(P_{\mathcal{C},\mathcal{B}})^{-1} = P_{\mathcal{B},\mathcal{C}}$

(b) $P_{\mathcal{C},\mathcal{A}} = P_{\mathcal{C},\mathcal{B}}\,P_{\mathcal{B},\mathcal{A}}$

Proposition 21.4

Let V be a finite dimensional vector space and let $\mathcal{B}$ and $\mathcal{C}$ be two bases for V. Then

$$\forall u \in V, \quad [u]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{B}}\,[u]_{\mathcal{B}}$$

Proof. Follows immediately from Lemma 20.2. □

Example 21.5. Consider the following bases for $\mathbb{R}^2$: $\mathcal{B} = \{(1,1),(1,-1)\}$ and $\mathcal{S} = \{(1,0),(0,1)\}$. Then we have $P_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $P_{\mathcal{B},\mathcal{S}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

If $u = (2,-4) \in \mathbb{R}^2$, then $[u]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{S}}[u]_{\mathcal{S}} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

Example 21.6. Consider the following bases for $\mathbb{C}^2$: $\mathcal{B} = \{(1,1),(1,-1)\}$ and $\mathcal{C} = \{(1+i,1),(1,1-i)\}$. To find $P_{\mathcal{C},\mathcal{B}}$, we could calculate $[(1,1)]_{\mathcal{C}}$ and $[(1,-1)]_{\mathcal{C}}$. Alternatively, we can proceed as follows:

$$P_{\mathcal{C},\mathcal{B}} = P_{\mathcal{C},\mathcal{S}}\,P_{\mathcal{S},\mathcal{B}} = (P_{\mathcal{S},\mathcal{C}})^{-1}\,P_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} 1+i & 1 \\ 1 & 1-i \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -i & 2-i \\ i & -2-i \end{bmatrix}$$

21.3 Change of basis for linear transformations

We can also use transition matrices to relate two different matrix representations of the same linear transformation.

Proposition 21.7

Let V and W be finite dimensional vector spaces. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases for V and let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two bases for W. Let $T : V \to W$ be a linear transformation. Then

$$[T]_{\mathcal{C}_2,\mathcal{B}_2} = P_{\mathcal{C}_2,\mathcal{C}_1}\,[T]_{\mathcal{C}_1,\mathcal{B}_1}\,P_{\mathcal{B}_1,\mathcal{B}_2}$$

Proof. Let $u \in V$. We have

$$\begin{aligned}
(P_{\mathcal{C}_2,\mathcal{C}_1}\,[T]_{\mathcal{C}_1,\mathcal{B}_1}\,P_{\mathcal{B}_1,\mathcal{B}_2})\,[u]_{\mathcal{B}_2} &= P_{\mathcal{C}_2,\mathcal{C}_1}\,[T]_{\mathcal{C}_1,\mathcal{B}_1}\,(P_{\mathcal{B}_1,\mathcal{B}_2}\,[u]_{\mathcal{B}_2}) \\
&= P_{\mathcal{C}_2,\mathcal{C}_1}\,[T]_{\mathcal{C}_1,\mathcal{B}_1}\,[u]_{\mathcal{B}_1} && \text{(Proposition 21.4)} \\
&= P_{\mathcal{C}_2,\mathcal{C}_1}\,[T(u)]_{\mathcal{C}_1} && \text{(Lemma 20.2)} \\
&= [T(u)]_{\mathcal{C}_2} && \text{(Proposition 21.4)}
\end{aligned}$$

It follows that $P_{\mathcal{C}_2,\mathcal{C}_1}\,[T]_{\mathcal{C}_1,\mathcal{B}_1}\,P_{\mathcal{B}_1,\mathcal{B}_2} = [T]_{\mathcal{C}_2,\mathcal{B}_2}$ (see Exercise 174). □

Corollary 21.8

Let V be a finite dimensional vector space and let $\mathcal{B}$ and $\mathcal{C}$ be two bases for V. Denote $P = P_{\mathcal{C},\mathcal{B}}$. Let $T : V \to V$ be a linear transformation. Then

$$[T]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{B}}\,[T]_{\mathcal{B}}\,P_{\mathcal{B},\mathcal{C}} = P\,[T]_{\mathcal{B}}\,P^{-1}$$

Proof. Apply the Proposition with $\mathcal{B}_2 = \mathcal{C}_2 = \mathcal{C}$ and $\mathcal{B}_1 = \mathcal{C}_1 = \mathcal{B}$. Note that $P_{\mathcal{B},\mathcal{C}} = (P_{\mathcal{B},\mathcal{C}})^{-1}$. □

Example 21.9. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by $T(x, y) = (3x - y, -x + 3y)$. Using the standard basis $\mathcal{B} = \{(1, 0), (0, 1)\}$ we find the matrix of f is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

Now let's calculate the matrix with respect to the basis $\mathcal{C} = \{(1, 1), (-1, 1)\}$.

$$[T]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{B}}\,[T]_{\mathcal{B}}\,P_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

(Alternatively, we could calculate $[T]_{\mathcal{C}}$ by finding $[(1, 1)]_{\mathcal{C}}$ and $[(-1, 1)]_{\mathcal{C}}$.)

The above corollary motivates the following definition.

Definition 21.10

Let $A, B \in \mathcal{M}_{n,n}(F)$. We say that A and B are **similar** if there exists an invertible matrix $P \in \mathcal{M}_{n,n}(F)$ such that $A = PBP^{-1}$. It is denoted $A \sim B$.[a]

[a] Beware! This is not the same as saying that A and B are row-equivalent.

From Corollary 21.8 we know that $[T]_{\mathcal{C}} \sim [T]_{\mathcal{B}}$

Exercise 195. Let $A, B \in \mathcal{M}_{n,n}(\mathbb{F})$ and suppose that A and B are similar. Show that there exists a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$ and a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that $A = [T]_{\mathcal{S}}$ and $B = [T]_{\mathcal{B}}$ (where $\mathcal{S}$ is the standard basis for $\mathbb{F}^n$).

21.4 Exercises

196. Determine whether or not the given linear transformation is invertible. If it is invertible, compute its inverse.

   (a) $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(x, y, z) = (x + z, x - y + z, y + 2z)$

   (b) $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (3x + 2y, -6x - 4y)$

   (c) $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ an anticlockwise rotation around the origin through an angle of θ.

   (d) $T^\theta : \mathbb{R}^2 \to \mathbb{R}^2$ a reflection in the line through the origin which forms an angle θ with the x-axis.

197. Show that the transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y+z \\ z+x \end{bmatrix}$$

   is invertible and find its inverse

198. (a) Find the transition matrix P from B to C, where B, C are the following bases of $\mathbb{R}^3$

$$B = \{(1, -2, 1),\ (0, 3, 2),\ (1, 0, -1)\} \text{ and } C = \{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)\}$$

   (b) Use P to find $[x]_B$ if

   i) $x = (3, -2, 5)$ ii) $x = (-2, 7, 4)$

199. Verify that the given set $\mathcal{B}$ is a basis for $\mathbb{R}^n$. Compute the change of basis matrix for each of the bases, and use it to find the coordinate matrix of v with respect to $\mathcal{B}$.

   (a) $\mathcal{B} = \{(1, 2), (1, -2)\}, v = (-1, 3)$

   (b) $\mathcal{B} = \{(1, 1, 1), (1, 0, 1), (-1, 1, 0)\}, v = (3, -1, 1)$

200. (a) Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $[T]_\mathcal{S} = \left[\begin{smallmatrix} 1 & 1 \\ 0 & -1 \end{smallmatrix}\right]$. Find the matrix $[T]_\mathcal{B}$ that represents T with respect to the basis $\mathcal{B}$ of question 199 (a).

   (b) Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $[T]_\mathcal{S} = \left[\begin{smallmatrix} 2 & -1 & 0 \\ -2 & 1 & 2 \\ -1 & -1 & 3 \end{smallmatrix}\right]$. Find the matrix $[T]_\mathcal{B}$ that represents T with respect to the basis $\mathcal{B}$ of question 199 (b).

201. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (4x + y - 4z, -3x - y + 5z, x)$. Find the matrix $[T]_\mathcal{B}$ that represents T with respect to the basis $\mathcal{B}$ of question 199 (b).

202. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x_1, x_2) = (x_1 - 2x_2, -x_2)$, and let $B = \{u_1, u_2\}$, $B' = \{v_1, v_2\}$, $u_1 = (1, 0)$, $u_2 = (0, 1)$, where $v_1 = (2, 1)$, $v_2 = (-3, 4)$.

   (a) Write down the matrix $[T]_B$ of T with respect to B.

   (b) Compute the matrix $[T]_{B'}$ of T with respect to B'.

203. A linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ has matrix

$$[T]_\mathcal{S} = \begin{bmatrix} 13 & -4 & -5 \\ 15 & -4 & -6 \\ 18 & -6 & = 7 \end{bmatrix}$$

   with respect to the standard basis for $\mathbb{R}^3$. Find the matrix $[T]_\mathcal{B}$ of T with respect to the basis $\mathcal{B} = \{(1, 2, 1), (0, 1, -1), (2, 3, 2)\}$.

204. Let $\mathcal{B} = \{b_1, b_2, b_3\}$ be a basis for $\mathbb{C}^3$. Calculate the nullity and rank of the linear transformation $T : \mathbb{C}^3 \to \mathbb{C}^3$ determined by

$$\begin{aligned} T(b_1) &= b_1 - b_2 \\ T(b_2) &= b_2 - b_3 \\ T(b_3) &= b_1 - b_3 \end{aligned}$$

205. Calculate the nullity and rank of the linear transformation $T : (\mathbb{F}_7)^3 \to (\mathbb{F}_7)^3$ determined by

$$\begin{aligned} T\,(1,0,0) &= (1,2,3) \\ T\,(0,1,0) &= (3,4,5) \\ T\,(0,0,1) &= (5,1,4) \end{aligned}$$

Extra material for lecture 21

- ▷ Show that the relation of similarity is an equivalence relation on $\mathcal{M}_{n,n}(\mathbb{F})$. That is, show that the relation is reflexive, symmetric and transitive.

- ▷ Let V be an n-dimensional $\mathbb{F}$-vector space. Show that every invertible $n \times n$ matrix is a change of basis matrix for V. That is, show that for all invertible $P \in \mathcal{M}_{n,n}(\mathbb{F})$ there exist bases $\mathcal{B}$ and $\mathcal{B}'$ for V such that $P = [\mathrm{Id}_V]_{\mathcal{B}',\mathcal{B}}$.

- ▷ Given a linear transformation $T : V \to V$, show that there exist bases $\mathcal{B}$ and $\mathcal{B}'$ for V such that $[T]_{\mathcal{B}',\mathcal{B}}$ is diagonal and all entries are either 0 or 1. (Hint: start with a basis for the kernel of T.) It is important here that $\mathcal{B}$ and $\mathcal{B}'$ do not have to be the same. We will be investigating later the special case in which there exists a basis $\mathcal{B}$ such that $[T]_{\mathcal{B},\mathcal{B}}$ is diagonal.

LECTURE 22

Dual space

22.1 Linear functionals and the dual space

Let $\mathbb{F}$ be a field and V a vector space over $\mathbb{F}$. We consider linear transformations from V to $\mathbb{F}$.

Definition 22.1: Linear functional

A **linear functional** is a linear transformation $\varphi : V \to \mathbb{F}$. That is, $\forall u, v \in V$ and $\forall k \in \mathbb{F}$

$$\begin{aligned} \varphi(u+v) &= \varphi(u) + \varphi(v) \\ \varphi(ku) &= k\varphi(u) \end{aligned}$$

Exercise 206. Let $\varphi : V \to \mathbb{F}$ be a linear functional. Show that if $\varphi \neq 0$, then φ is surjective.

Exercise 207. Let $u \in V \setminus \{\vec{0}\}$. Show that there exists $\varphi \in V^*$ such that $\varphi(u) \neq 0$. (You may assume that V is finite dimensional.)

Examples 22.2.

1. $V = \mathbb{R}^3$, $\varphi(x, y, z) = z$
2. $V = \mathbb{F}[x]$, $\varphi(p(x)) = \int_0^1 p(x)\,dx$
3. $V = \mathcal{M}_{n,n}(\mathbb{F})$, $\varphi(A) = A_{11} + \cdots + A_{nn}$ (i.e., the trace of A)

Definition 22.3: Dual space

Let V be a vector space over $\mathbb{F}$. The **dual space**, V^*, is the vector space made of all linear functionals with operations given by, for $\varphi, \psi \in V^*$ and $k \in \mathbb{F}$,

$$\begin{aligned} (\varphi+\psi)(v) &= \varphi(v) + \psi(v) \\ (k\varphi)(v) &= k\varphi(v) \end{aligned}$$

22.2 Dual basis

Definition 22.4

Suppose that $\mathcal{B} = \{b_1, \ldots, b_n\}$ is a basis for V. Define $b_1^*, \ldots, b_n^* \in V^*$ by[a]

$$b_i^*(b_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$\mathcal{B}^* = \{b_1^*, \ldots, b_n^*\} \subset V^*$ is called the **dual basis** (or the basis dual to $\mathcal{B}$).

[a] That this uniquely determines the b_i^* follows from Theorem 19.7

Theorem 22.5

Let $\{b_1, \dots, b_n\}$ be a basis for V. Then $\{b_1^*, \dots, b_n^*\}$ is a basis for V^*, and

$$\begin{aligned} \forall u \in V, \qquad & u = b_1^*(u)b_1 + \cdots + b_n^*(u)b_n \\ \forall \varphi \in V^*, \qquad & \varphi = \varphi(b_1)b_1^* + \cdots + \varphi(b_n)b_n^* \end{aligned}$$

Proof. For all $u \in V$ we have $u = \sum_{j=1}^n \beta_j b_j$ for some $\beta_j \in \mathbb{F}$. Then

$$b_i^*(u) = b_i^*\left(\sum_{j=1}^n \beta_j b_j\right) = \sum_{j=1}^n \beta_j b_i^*(b_j) = \beta_i$$

Having shown that $\beta_i = b_i^*(u)$, we have that $u = \sum_{i=1}^n b_i^*(u)b_i$.

For the second statement, let $\varphi \in V^*$. For all $u \in V$ we have $u = \sum_{j=1}^n \beta_j b_j$ for some $\beta_j \in \mathbb{F}$. Then

$$\begin{aligned} \left(\sum_{i=1}^n \varphi(b_i)b_i^*\right)(u) &= \left(\sum_{i=1}^n \varphi(b_i)b_i^*\right)\left(\sum_{j=1}^n \beta_j b_j\right) \\ &= \sum_{i=1}^n \varphi(b_i) \sum_{j=1}^n \beta_j b_i^*(b_j) \\ &= \sum_{i=1}^n \varphi(b_i)\beta_i = \varphi\left(\sum_{i=1}^n \beta_i b_i\right) = \varphi(u) \end{aligned}$$

Therefore $\varphi = \sum_{i=1}^n \varphi(b_i)b_i^*$. To see that $\{b_1^*, \dots, b_n^*\}$ is linearly independent we have

$$\begin{aligned} \sum_{i=1}^n \beta_i b_i^* = 0 &\implies \forall j, \quad \sum_{i=1}^n \beta_i b_i^*(b_j) = 0 \\ &\implies \forall j, \quad \beta_j = 0 \end{aligned}$$

□

Corollary 22.6

It follows that if V is finite dimensional, then $V^* \cong V$. In particular, $\dim(V^*) = \dim(V)$. □

22.3 Second dual space

Given $u \in V$ we define an element $\widehat{u} \in V^{**}$ by

$$\widehat{u}(\varphi) = \varphi(u)$$

Exercise 208. Show that $\widehat{u}$ is a linear transformation $V^* \to \mathbb{F}$.

Theorem 22.7: natural isomorphism

Let V be a finite dimensional vector space. The map $V \to V^{**}$ given by $u \mapsto \widehat{u}$ is an isomorphism.

Proof. To see that the map is a linear transformation, for $u, v \in V$ and $\varphi \in V^*$ we have:

$$\begin{aligned}\widehat{(u+v)}(\varphi) &= \varphi(u+v) = \varphi(u) + \varphi(v) = \widehat{u}(\varphi) + \widehat{v}(\varphi)\\ \widehat{(ku)}(\varphi) &= \varphi(ku) = k\varphi(u) = k\widehat{u}(\varphi)\end{aligned}$$

To see that the map is injective, let $u \in V$ and suppose that $\widehat{u} = 0$. We have

$$\begin{aligned}\widehat{u} = 0 &\implies \forall \varphi \in V^*, \quad \widehat{u}(\varphi) = 0\\ &\implies \forall \varphi \in V^*, \quad \varphi(u) = 0\\ &\implies u = \vec{0} && \text{(by Exercise 207)}\end{aligned}$$

Therefore the map is injective.

From Corollary 22.6, $\dim(V^{**}) = \dim(V^*) = \dim(V)$. Since the linear transformation is injective and the domain and codomain have the same dimension, we conclude that the map is an isomorphism.

□

Remark. The above isomorphism does not require any choice of basis.

22.4 Transpose linear transformation

Definition 22.8

Given a linear transformation $T : V \to W$ we define the **transpose** of T to be the linear transformation $T^* : W^* \to V^*$ given by

$$T^*(\psi) = \psi \circ T$$

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[dr, "T^*(\psi)"'] & W \arrow[d, "\psi"] \\ & \mathbb{F} \end{tikzcd}$$

The matrix of the transpose linear transformation T^* is the transpose of the matrix of T.

Theorem 22.9

Let V and W be finite dimensional vector spaces. Let $\mathcal{B}$ be a basis for V and $\mathcal{B}^*$ the dual basis for V^*. Let $\mathcal{C}$ be a basis for W and $\mathcal{C}^*$ the dual basis for W^*. Let $T : V \to W$ be a linear transformation. Then

$$[T^*]_{\mathcal{B}^*,\mathcal{C}^*} = ([T]_{\mathcal{C},\mathcal{B}})^T$$

Proof. Let $\mathcal{B} = \{b_1, \dots, b_n\}$, $\mathcal{B}^* = \{b_1^*, \dots, b_n^*\}$, $\mathcal{C} = \{c_1, \dots, c_n\}$, and $\mathcal{C}^* = \{c_1^*, \dots, c_n^*\}$. Recall that

$$b_i^*(b_j) = c_i^*(c_j) = \begin{cases} 1 & \text{if } i = j\\ 0 & \text{if } i \neq j\end{cases}$$

Let $a_{ij} \in \mathbb{F}$ be such that $T(b_j) = \sum_{i=1}^n a_{ij}c_i$. That is, a_{ij} is the (i,j)-th entry of $[T]_{\mathcal{C},\mathcal{B}}$. The i-th row of $[T]_{\mathcal{C},\mathcal{B}}$ is $[a_{i1} \cdots a_{in}]$.

We will show that i-th column of $[T^*]_{\mathcal{B}^*,\mathcal{C}^*}$ is the transpose of the i-th row of $[T]_{\mathcal{C},\mathcal{B}}$.

The i-th column of $[T^*]_{\mathcal{B}^*,\mathcal{C}^*}$ is given by $[T^*(c_i^*)]_{\mathcal{B}^*}$. We have

$$\begin{aligned}T^*(c_i^*) &= c_i^* \circ T && \text{(definition of transpose linear transformation)}\\ &= \sum_{j=1}^{n} c_i^* \circ T(b_j) b_j^* && \text{(Theorem 22.5)}\end{aligned}$$

The i-th column of $[T^*]_{\mathcal{B}^*,\mathcal{C}^*}$ is therefore $[c_i^* \circ T(b_1) \ \cdots \ c_i^* \circ T(b_n)]^T$. We will be done if we can show that $c_i^* \circ T(b_j) = a_{ij}$. We have

$$c_i^* \circ T(b_j) = c_i^*(\sum_{k=1}^{n} a_{kj} c_k) = \sum_{k=1}^{n} a_{kj} c_i^*(c_k) = a_{ij}$$

□

22.5 Exercises

209. Let $\varphi \in (\mathbb{R}^2)^*$ satisfy $\varphi(1,-1) = 2$ and $\varphi(-4,5) = -3$. Find $\varphi(x,y)$.

210. For each of the following bases of $\mathbb{R}^3$, find the dual basis $\mathcal{B}^* = \{u^*, v^*, w^*\}$ of $(\mathbb{R}^3)^*$.

    (a) $\mathcal{B} = \{u = (-2,-2,3), v = (1,2,-1), w = (-1,-1,2)\}$

    (b) $\mathcal{B} = \{u = (-1,1,1), v = (-2,1,2), w = (-1,1,2)\}$

211. Let V be a finite dimensional vector space with basis $\mathcal{B}$. Show that $\forall u \in V$ and $\forall \varphi \in V^*$,

$$\varphi(u) = [\varphi]_{\mathcal{B}^*}^T [u]_{\mathcal{B}}$$

212. For each of the following linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^3$ and linear functionals $\varphi \in (\mathbb{R}^3)^*$, calculate $T^*(\varphi)(x,y)$.

    (a) $T(x,y) = (x+y, 0, x-y)$, $\varphi(x,y,z) = x+y+z$

    (b) $T(x,y) = (y, x, x+y)$, $\varphi(x,y,z) = z$

213. Consider linear transformations $S : U \to V$ and $T : V \to W$. Show that $(T \circ S)^* = S^* \circ T^*$.

214. Let $V = \mathcal{P}_2(\mathbb{F})$. Given a scalar $a \in \mathbb{F}$, define $\varphi_a : V \to \mathbb{F}$ by

$$\varphi_a(a_0 + a_1 x + a_2 x^2) = a_0 + a_1 a + a_2 a^2$$

    Show

    (a) $\varphi_a \in V^*$

    (b) If $a \neq b$, then $\varphi_a \neq \varphi_b$

    Suppose that $a, b, c \in \mathbb{F}$ are distinct.

    (c) Show that $\{\varphi_a, \varphi_b, \varphi_c\}$ is a basis for V^*.

    (d) Find the basis $\mathcal{B}$ for V, such that $\mathcal{B}^*$ is the above basis for V^*.

Extra material for lecture 22

▷ If V is infinite dimensional, then V^* is not isomorphic to V. Given a basis $\mathcal{B} = \{b_i \mid i \in I\}$ we can define $\{b_i^* \mid i \in I\}$ as in Definition 22.4. This set is linearly dependent, but is not a spanning set (when I is infinite).

For example, consider $V = \mathbb{F}_2^{<\mathbb{N}}$, the vector space of infinite sequences $(a_i)_{i\in\mathbb{N}}$ having the property that $a_i \neq 0$ for only finitely many i. A basis for V is $\mathcal{B} = \{b_i \mid i \in \mathbb{N}\}$ where b_i has a 1 in the i-th position and zeros elsewhere. The linear functional $\varphi \in V^*$ defined by $\varphi(b_i) = 1$ (for all $i \in \mathbb{N}$) is not in the span of the set $\{b_i^* \mid i \in \mathbb{N}\}$. The dual space is $V^* = \mathbb{F}_2^{\mathbb{N}}$, the vector space of infinite sequences $(a_i)_{i\in\mathbb{N}}$ and has dimension strictly larger than that of V.

▷ The following connects transition matrices for V and transition matrices for V^*.

Theorem. *Let V be a (finite dimensional) vector space. Let $\mathcal{B}$ and $\mathcal{C}$ be two bases for V and let $\mathcal{B}^*$ and $\mathcal{C}^*$ be the (respective) dual bases for V^*. Then*

$$(P_{\mathcal{C}^*,\mathcal{B}^*})^T = (P_{\mathcal{C},\mathcal{B}})^{-1}$$

Proof. We will show that $(P_{\mathcal{C}^*,\mathcal{B}^*})^T P_{\mathcal{C},\mathcal{B}} = I$. The i-th row of $(P_{\mathcal{C}^*,\mathcal{B}^*})^T$ is $[b_i^*]_{\mathcal{C}^*}^T = [\alpha_1 \cdots \alpha_n]$ where $b_i^* = \alpha_1 c_1^* + \cdots + \alpha_n c_n^*$. The j-th column of $P_{\mathcal{C},\mathcal{B}}$ is $[b_j]_{\mathcal{C}} = [\beta_1 \cdots \beta_n]^T$ where $b_j = \beta_1 c_1 + \cdots + \beta_n c_n$. Therefore

$$((P_{\mathcal{C}^*,\mathcal{B}^*})^T P_{\mathcal{C},\mathcal{B}})_{ij} = [\alpha_1 \cdots \alpha_n][\beta_1 \cdots \beta_n]^T = \alpha_1\beta_1 + \cdots + \alpha_n\beta_n$$

Now observe that

$$b_i^*(b_j) = (\alpha_1 c_1^* + \cdots + \alpha_n c_n^*)(\beta_1 c_1 + \cdots + \beta_n c_n) = \alpha_1\beta_1 + \cdots + \alpha_n\beta_n$$

We have shown that

$$((P_{\mathcal{C}^*,\mathcal{B}^*})^T P_{\mathcal{C},\mathcal{B}})_{ij} = b_i^*(b_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

□

▷ Let $S \subseteq V$ be a subset (not necessarily a subspace) of a vector space V. The **annihilator** of S is

$$S^0 = \{\varphi \in V^* \mid \forall u \in S, \varphi(u) = 0\}$$

The annihilator is a subspace of V^* (even if S is not a subspace of V).

In the case in which V is finite dimensional and $W \leqslant V$ is a subspace, we have the following.

Theorem. *Let V be a finite dimensional vector space and W a subspace of V. Then*

1. $\dim(W) + \dim(W^0) = \dim(V)$
2. $(W^0)^0 = W$ (where we are identifying V and V^{**} via the natural isomorphism)

Proof. Exercise! Hint for the first part: Let $\{w_1, \ldots, w_k\}$ be a basis for W. Extend to a basis for V, $\{w_1, \ldots, w_k, v_1, \ldots, v_m\}$. Show that $\{v_1^*, \ldots, v_m^*\}$ is a basis for W^0. □

Eigenvalues and eigenvectors

23.1 Invariant subspaces

To help with understanding and analysing linear transformations, it is useful to identify subspaces that are mapped to themselves in the following sense.

Definition 23.1

Given a linear transformation $T : V \to V$, a subspace $W \leqslant V$ is called an **invariant subspace** if $T(W) \subseteq W$.

Example 23.2. Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be the linear transformation having standard matrix representation

$$[T]_{\mathcal{S}} = \begin{bmatrix} 2 & 0 & 1 & 2 \\ 1 & 2 & 0 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & -2 & 2 & 1 \end{bmatrix}$$

Let $u_1 = (-1, 1, 1, 0)$ and $u_2 = (0, 0, 1, -1)$. The subspace $U = \operatorname{span}\{u_1, u_2\}$ is invariant. To see this it is enough to note that

$$[T(u_1)]_{\mathcal{S}} = [T]_{\mathcal{S}}[u_1]_{\mathcal{S}} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ -1 \end{bmatrix} = [u_1]_{\mathcal{S}} + [u_2]_{\mathcal{S}} \qquad \text{therefore} \qquad T(u_1) = u_1 + u_2 \in U$$

$$[T(u_2)]_{\mathcal{S}} = [T]_{\mathcal{S}}[u_2]_{\mathcal{S}} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = [u_1]_{\mathcal{S}} - [u_2]_{\mathcal{S}} \qquad \text{therefore} \qquad T(u_2) = u_1 - u_2 \in U$$

Exercise 215. Let V be a finite-dimensional $\mathbb{F}$-vector space and let $T : V \to V$ be a linear transformation. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis for V and $k \in \{1, 2, \dots, n-1\}$. Define $W = \operatorname{span}\{b_1, \dots, b_k\}$ and suppose that W is invariant (i.e. $T(W) \subseteq W$). Show that

$$[T]_{\mathcal{B}} = \left[\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right] \text{ for some } A \in \mathcal{M}_{k,k}(\mathbb{F}), C \in \mathcal{M}_{k,(n-k)} \text{ and } B \in \mathcal{M}_{(n-k),(n-k)}(\mathbb{F}).$$

23.2 Definition of eigenvalues and eigenvectors

Consideration of 1-dimensional invariant subspaces leads to the idea of eigenvectors and eigenvalues.

Exercise 216. Let V be a vector space and let $T : V \to V$ be a linear transformation. Suppose $u \in V$ and $\lambda \in \mathbb{F}$ are such that $T(u) = \lambda u$. Show that the subspace $W = \operatorname{span}\{u\} \leqslant V$ is invariant.

Definition 23.3: Eigenvalues and eigenvectors of linear transformations

Let V be a vector space with field of scalars $\mathbb{F}$, and let $T\colon V \to V$ be a linear transformation. A scalar $\lambda \in \mathbb{F}$ is called an **eigenvalue** of T if there is a *non-zero* vector $v \in V \setminus \{\vec{0}\}$ such that

$$T(v) = \lambda v$$

Such a vector $v \in V \setminus \{\vec{0}\}$ is called an **eigenvector** of T corresponding to the eigenvalue λ.

The set $\{v \in V \mid T(v) = \lambda v\}$ is a subspace of V, called the **eigenspace** of λ.

Exercise 217. Let $T : V \to V$ be a linear transformation and suppose that λ is an eigenvalue of T. Let $W_\lambda = \{v \in V \mid T(v) = \lambda v\}$ (i.e., the corresponding eigenspace). Show that

(a) W_λ is a subspace of V. (b) $W_\lambda \neq \{\vec{0}\}$ (c) $T(W_\lambda) \subseteq W_\lambda$

Example 23.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation whose standard matrix is

$$[T]_{\mathcal{S}} = \begin{bmatrix} -7 & -18 & -12 \\ 9 & 20 & 12 \\ -9 & -18 & -10 \end{bmatrix}$$

$T(2,-1,0) = (4,-2,0) = 2(2,-1,0)$ and $T(6,-5,3) = (12,-10,6) = 2(6,-5,3)$ since

$$[T(2,-1,0)]_{\mathcal{S}} = [T]_{\mathcal{S}}[(2,-1,0)]_{\mathcal{S}} = \begin{bmatrix} -7 & -18 & -12 \\ 9 & 20 & 12 \\ -9 & -18 & -10 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix}$$

$$[T(6,-5,3)]_{\mathcal{S}} = [T]_{\mathcal{S}}[(6,-5,3)]_{\mathcal{S}} = \begin{bmatrix} -7 & -18 & -12 \\ 9 & 20 & 12 \\ -9 & -18 & -10 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} 12 \\ -10 \\ 6 \end{bmatrix}$$

So both $(2,-1,0)$ and $(6,-5,3)$ are eigenvectors with eigenvalue 2.

As we've seen above, the matrix of a linear transformation is, of course, useful for working with eigenvectors. We adapt the definition of eigenvalue and eigenvector to matrices in a natural way.

Definition 23.5: Eigenvalues and eigenvectors of a matrix

Let $\mathbb{F}$ be a field, and let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. A scalar $\lambda \in \mathbb{F}$ is an **eigenvalue** of A if there is a *non-zero* column matrix $v \in \mathcal{M}_{n,1}(\mathbb{F})$ such that

$$Av = \lambda v$$

Then v is called an **eigenvector** of A corresponding to eigenvalue λ.

Lemma 23.6

Let V be a finite-dimensional vector space over a field $\mathbb{F}$ and let $T : V \to V$ be a linear transformation. Let $\mathcal{B}$ be a basis of V and let $\lambda \in \mathbb{F}$ and $v \in V$. Then

v is an eigenvector of T with eigenvalue $\lambda \iff [v]_{\mathcal{B}}$ is an eigenvector of $[T]_{\mathcal{B}}$ with eigenvalue λ

Proof. This follows from Lemma 20.2 (and Lemma 15.4):

$$T(v) = \lambda v \iff [T(v)]_{\mathcal{B}} = [\lambda v]_{\mathcal{B}} \iff [T]_{\mathcal{B}}[v]_{\mathcal{B}} = \lambda [v]_{\mathcal{B}}$$

Note also that $v = \vec{0} \iff [v]_{\mathcal{B}} = \vec{0}$. □

23.3 Calculating eigenvalues

We can calculate the eigenvalues of a matrix (and hence a linear transformation) using the following.

Proposition 23.7

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. Let $\lambda \in \mathbb{F}$ and $v \in V$.

$$Av = \lambda v \iff Av - \lambda v = \vec{0} \iff Av - \lambda I_n v = \vec{0} \iff (A - \lambda I_n)v = \vec{0}$$

Therefore,

$$\begin{aligned} \lambda \text{ is an eigenvalue } &\iff (A - \lambda I_n) \text{ has a non-trivial solution space} \\ &\iff \operatorname{nullity}(A - \lambda I_n) > 0 \\ &\iff \operatorname{rank}(A - \lambda I_n) < n && \text{Rank-nullity theorem} \\ &\iff \det(A - \lambda I_n) = 0 && \text{Noting that } A - \lambda I_n \text{ is square} \end{aligned}$$

□

Example 23.8. Let's find the eigenvalues of $A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$.

$$\det(A - \lambda I_2) = \det(\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}) = \det \begin{bmatrix} 1-\lambda & 4 \\ 1 & 1-\lambda \end{bmatrix} = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$$

The eigenvalues of A are -1 and 3.

Exercise 218. Show that if $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is in upper triangular form, then the eigenvalues of A are exactly the entries on the diagonal of A.

Definition 23.9: Characteristic polynomial

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. The determinant $\det(xI_n - A)$ is a polynomial in x called the **characteristic polynomial of** A. We will denote it by c_A.

$$c_A(x) = \det(xI_n - A)$$

The **characteristic equation** of A is the equation $c_A(x) = 0$

Remark.

1. The characteristic polynomial of $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is always monic and of degree exactly n. That is,
$$c_A(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{n-1}x^{n-1} + x^n$$
for some $\alpha_0, \ldots, \alpha_{n-1} \in \mathbb{F}$.

2. For $A \in \mathcal{M}_{2,2}(\mathbb{F})$ we have $c_A(x) = \det(A) - \operatorname{tr}(A)x + x^2$

3. From Proposition 23.7 we know that the eigenvalues of A are exactly the roots of the characteristic polynomial. If $\lambda \in \mathbb{F}$ is an eigenvalue of A, then $(x - \lambda)$ divides the characteristic polynomial of A.

Definition 23.10: Algebraic multiplicity of eigenvalues

The **algebraic multiplicity** of an eigenvalue λ is the largest $k \in \mathbb{N}$ such that $(x-\lambda)^k$ divides the characteristic polynomial.

Remark. 1. In general, the sum of the algebraic multiplicities is at most n.

2. In the case in which $\mathbb{F} = \mathbb{C}$, there is always at least one eigenvalue.[*] Further, the sum of the algebraic multiplicities equals n.

Example 23.11. Let $A = \begin{bmatrix} 8 & -9 & -9 \\ 9 & -10 & -9 \\ -1 & 2 & 5 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{R})$. The characteristic polynomial of A is given by

$$\begin{aligned}
|xI_3 - A| &= \begin{vmatrix} x-8 & 9 & 9 \\ -9 & x+10 & 9 \\ 1 & -2 & x-5 \end{vmatrix} \xlongequal{R_2-R_1} \begin{vmatrix} x-8 & 9 & 9 \\ -x-1 & x+1 & 0 \\ 1 & -2 & x-5 \end{vmatrix} = (x+1)\begin{vmatrix} x-8 & 9 & 9 \\ -1 & 1 & 0 \\ 1 & -2 & x-5 \end{vmatrix} \\
&\xlongequal{C_1+C_2} (x+1)\begin{vmatrix} x+1 & 9 & 9 \\ 0 & 1 & 0 \\ -1 & -2 & x-5 \end{vmatrix} = (x+1)\begin{vmatrix} x+1 & 9 \\ -1 & x-5 \end{vmatrix} \\
&= (x+1)((x+1)(x-5)+9) = (x+1)(x-2)^2
\end{aligned}$$

The eigenvalues of A are therefore -1 and 2. The eignevalue -1 has algebraic multiplicity 1 and the eigenvalue 2 has algebraic multiplicity 2.

Example 23.12. Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$. The characteristic polynomial is:

$$\det \begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix} = x^2 + 1$$

This polynomial has no roots in $\mathbb{R}$. Therefore A has no eigenvalues.

Now, use the same matrix, but considered as an element of $\mathcal{M}_{2,2}(\mathbb{C})$. The characteristic polynomial is the same, but since it does have roots in $\mathbb{C}$, the matrix has eigenvalues. The eigenvalues are i and $-i$. Each has algebraic multiplicity 1.

23.4 Exercises

219. Find the eigenvalues (over $\mathbb{C}$) of the following matrices:

(a) $\begin{bmatrix} 7 & -2 \\ 15 & -4 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & -2 \\ 17 & -7 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & 5 \\ -2 & -3 \end{bmatrix}$

220. Find, by inspection, the eigenvalues of the given matrix.

(a) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & -1 \\ 0 & 0 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{bmatrix}$

221. Find the eigenvalues (over $\mathbb{R}$) of the following matrices.

[*] By the Fundamental Theorem of Algebra.

(a) $\begin{bmatrix} 2 & -3 & 6 \\ 0 & 5 & -6 \\ 0 & 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} -5 & -8 & -12 \\ -6 & -10 & -12 \\ 6 & 10 & 13 \end{bmatrix}$ (e) $\begin{bmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 2 & 2 \\ -1 & -1 & -2 \\ 1 & 2 & 3 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

222. Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$.

    (a) Prove that if $\lambda \in \mathbb{F}$ is an eigenvalue of A, then λ^2 is an eigenvalue of A^2.

    (b) Suppose that A is invertible. Prove that λ is an eigenvalue of A if and only if $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}. What relationship holds between the eigenvectors of A and A^{-1}?

223. Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$ and let $p \in \mathbb{F}[x]$. Show that

$$p(A) = 0 \implies p(\lambda) = 0 \quad \text{for all eigenvalues } \lambda \text{ of } A$$

224. Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$. Show that $\det(A)$ is equal to the product of the eigenvalues (with multiplicity) of A .

Further material for lecture 23

- ▷ References for eigenvalues and eigenvectors

  Elementary Linear Algebra by Anton and Rorres, §5

  The Art of Proof by Beck and Geoghegan, §5

- ▷ Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$. The sum of the eigenvalues (with multiplicity) of A is equal to the trace of A. The product of the eigenvalues (with multiplicity) is equal to the determinant. Neither of these is obvious!

Eigenspaces

24.1 Calculating the eigenspaces of a matrix

We saw last lecture the definitions of the eigenvalues and eigenspaces for a linear transformation and for a matrix. If $\lambda \in \mathbb{F}$ is an eigenvalue value of $A \in \mathcal{M}_{n,n}(\mathbb{F})$, the corresponding eigenspace is the solution space of the matrix $A - \lambda I_n$.

Example 24.1. Let $A = \begin{bmatrix} 2 & -3 & 6 \\ 0 & 5 & -6 \\ 0 & 1 & 0 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{R})$.

We find the eigenvalues of A and a basis for each eigenspace.

The characteristic polynomial is given by

$$\begin{aligned} \left|xI_3 - A\right| &= \begin{vmatrix} x-2 & 3 & -6 \\ 0 & x-5 & 6 \\ 0 & -1 & x \end{vmatrix} = (x-2)\begin{vmatrix} x-5 & 6 \\ -1 & x \end{vmatrix} = (x-2)(x(x-5)+6) \\ &= (x-2)(x^2-5x+6) = (x-2)^2(x-3) \end{aligned}$$

Therefore the eigenvalues of A are 2 and 3. The eigenvalue 2 has algebraic multiplicity 2 and the eigenvalue 3 has algebraic multiplicity 1.

To find the eigenspace for eigenvalue 2 we solve for the solution space of $A - 2I_3$.[*]

$$A - 2I_3 = \begin{bmatrix} 0 & -3 & 6 \\ 0 & 3 & -6 \\ 0 & 1 & -2 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 0 & 1 & -2 \\ 0 & 3 & -6 \\ 0 & -3 & 6 \end{bmatrix} \xrightarrow[R_3+3R_1]{R_2-3R_1} \begin{bmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, the eigenspace (for eigenvalue 2) is

$$\{(x,y,z) \in \mathbb{R}^3 \mid y = 2z\} = \{x(1,0,0) + z(0,2,1) \mid x,z \in \mathbb{R}\} = \operatorname{span}\{(1,0,0),(0,2,1)\}$$

The set $\{(1,0,0),(0,2,1)\}$ is a basis for the eigenspace.

For eigenvalue 3, the eigenspace is the solution space of $A - 3I_3$.

$$\begin{aligned} A - 3I_3 = \begin{bmatrix} -1 & -3 & 6 \\ 0 & 2 & -6 \\ 0 & 1 & -3 \end{bmatrix} &\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} -1 & -3 & 6 \\ 0 & 1 & -3 \\ 0 & 2 & -6 \end{bmatrix} \xrightarrow{R_3-2R_2} \begin{bmatrix} -1 & -3 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \\ &\xrightarrow{R_1+3R_2} \begin{bmatrix} -1 & 0 & -3 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{-1\times R_1} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Therefore the set $\{(-3,3,1)\}$ is a basis for the eigenspace.

Definition 24.2: Geometric multiplicity of a eigenvalue

The **geometric multiplicity** of an eigenvalue is the dimension of the corresponding eigenspace.

[*] Which is equal to the solution space of $2I_3 - A$.

Example 24.3. We saw last lecture (Example 23.11) that the matrix $A = \left[\begin{smallmatrix} 8 & -9 & -9 \\ 9 & -10 & -9 \\ -1 & 2 & 5 \end{smallmatrix}\right] \in \mathcal{M}_{3,3}(\mathbb{R})$ has eigenvalues -1 (with algebraic multiplicity 1) and 2 (with algebraic multiplicity 2). Let's find the corresponding eigenspaces.

For eigenvalue -1 we have

$$A - (-1)I_3 = \begin{bmatrix} 9 & -9 & -9 \\ 9 & -9 & -9 \\ -1 & 2 & 6 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenspace has basis $\{(-4,-5,1)\}$ and geometric multiplicity 1.

For eigenvalue 2 we have

$$A - 2I_3 = \begin{bmatrix} 6 & -9 & -9 \\ 9 & -12 & -9 \\ -1 & 2 & 3 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenspace has basis $\{(-3,-3,1)\}$ and geometric multiplicity 1. Notice that, for this matrix, the geometric multiplicity of the eigenvalue 2 is strictly less than its algebraic multiplicity.

24.2 Characteristic polynomial of a linear transformation

We saw in Lemma 23.6 that the eigenvalues of a linear transforamtion T are the same as the eigenvalues of *any* matrix representation $[T]_{\mathcal{B}}$, and that there is a correspondence between the eigenvectors of T and the eigenvectors of $[T]_{\mathcal{B}}$. We now define the characteristic polynomial of T to be the characteristic polynomial of $[T]_{\mathcal{B}}$. To see that this is independent of the choice of basis $\mathcal{B}$, we note the following lemma. Recall that if $\mathcal{B}$ and $\mathcal{C}$ are two bases, then although $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$ are not equal, they are similar in the sense that $[T]_{\mathcal{B}} = P[T]_{\mathcal{C}}P^{-1}$ for some invertible matrix P (see Definition 21.10).

Lemma 24.4

Let $A, B \in \mathcal{M}_{n,n}(\mathbb{F})$. If A and B are similar, then they have the same characteristic polynomial.

Proof. Let $P \in \mathcal{M}_{n,n}$ be an invertible matrix such that $A = PBP^{-1}$. Then

$$\begin{aligned}\det(A - \lambda I_n) &= \det(PBP^{-1} - \lambda I_n) = \det(P(BP^{-1} - \lambda P^{-1}I_n)) = \det(P(B - \lambda P^{-1}I_nP)P^{-1}) \\ &= \det(P(B-\lambda I_n)P^{-1}) = \det(P)\det(B-\lambda I_n)\det(P^{-1}) = \det(P)\det(P^{-1})\det(B-\lambda I_n) \\ &= \det(B - \lambda I_n)\end{aligned}$$

□

Definition 24.5: Characteristic polynomial of a linear transformation

Let V be an n-dimensional vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear transformation. The **characteristic polynomial** of T is defined to be the the characteristic polynomial of some (hence any) matrix representation $[T]_{\mathcal{B}} \in \mathcal{M}_{n,n}(\mathbb{F})$ of T.

Example 24.6. Consider the linear transformation $T : \mathcal{P}_2(\mathbb{F}_3) \to \mathcal{P}_2(\mathbb{F}_3)$ given by

$$T(a + bx + cx^2) = (2a + b + c) + (b + 2c)x + bx^2$$

Let's find the characteristic polynomial of T, then its eigenvalues and eigenspaces. We choose a basis $\mathcal{B} = \{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{F}_3)$.

$$[T]_{\mathcal{B}} = \big[\, [T(1)]_{\mathcal{B}}\, [T(x)]_{\mathcal{B}}\, [T(x^2)]_{\mathcal{B}} \,\big] = \big[\, [2]_{\mathcal{B}}\, [1 + x + x^2]_{\mathcal{B}}\, [1 + 2x]_{\mathcal{B}} \,\big] = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 0 \end{bmatrix}$$

Now we calculate the characteristic polynomial of $[T]_{\mathcal{B}}$.

$$\begin{aligned}\begin{vmatrix} x-2 & -1 & -1 \\ 0 & x-1 & -2 \\ 0 & -1 & x \end{vmatrix} &= (x-2)\begin{vmatrix} x-1 & -2 \\ -1 & x \end{vmatrix} = (x-2)(x^2-x-2) = (x-2)(x-2)(x+1) \\ &= (x+1)^3\end{aligned}$$

The characteristic polynomial of T is $(x+1)^3$.

There is only one eigenvalue, $2 \in \mathbb{F}_3$, and it has algebraic multiplicity $3 \in \mathbb{N}$.

Now to calculate the corresponding eigenspace.

$$[T]_{\mathcal{B}} - 2I_3 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow[R_3-R_1]{R_2+R_1} \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We see that the solution space has dimension 2 and has a basis $\{(1,0,0),(0,2,1)\} \subseteq \mathbb{F}_3^3$.
The eigenspace of T is therefore the subspace of $\mathcal{P}_2(\mathbb{F}_3)$ that has basis $\{1, 2x+x^2\}$.
The eigenvalue $2 \in \mathbb{F}_3$ has geometric multiplicity $2 \in \mathbb{N}$.

24.3 Algebraic versus geometric multiplicity

In the above examples we saw that the geometric multiplicity was always less than or equal to the algebraic multiplicity. We prove that this is always the case.

Proposition 24.7

Let V be an n-dimensional vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear transformation. Suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of T. The geometric multiplicity of λ is less than or equal to its algebraic multiplicity.

Proof. Let $W \leqslant V$ be the eigenspace corresponding to λ. Let $k = \dim(W)$ be the geometric multiplicity of λ. Let $\{w_1, \dots, w_k\}$ be a basis for W. Extend to a basis for V, $\mathcal{B} = \{w_1, \dots, w_k, b_{k+1}, \dots, b_n\}$. We have

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda I_k & M \\ 0 & N \end{bmatrix} \qquad \text{for some } M \in \mathcal{M}_{k,(n-k)}(\mathbb{F}) \text{ and } N \in \mathcal{M}_{(n-k),(n-k)}(\mathbb{F})$$

The characteristic polynomial of T is therefore $(x-\lambda)^k c_N(x)$ where $c_N(x)$ is the characteristic polynomial of the matrix N. Therefore the algebraic multiplicity is greater than or equal to k. □

24.4 Exercises

225. Find bases for the eigenspaces of matrices in Exercise 219. Give the algebraic and geometric multiplicity of each eigenvalue.

226. Find bases for the eigenspaces of matrices in Exercise 221. Give the algebraic and geometric multiplicity of each eigenvalue.

227. Find 1-dimensional subspaces of $\mathbb{R}^2$ that are invariant under the linear transformations given by the following matrices:

(a) $\begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 \\ -3 & -6 \end{bmatrix}$

Further material for lecture 24

- ▷ Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. Prove that the characteristic polynomial $c_A(x) = \det(xI_n - A)$ has degree n and is monic.

Diagonalisation

We have seen in several examples that, given a linear transformation, there is sometimes a basis with respect to which the matrix of the linear transformation is diagonal. Having such a basis is extremely useful when working with linear transformations. We will give conditions for the existence of such a basis and see that it is not always possible.

25.1 Diagonalisability

Definition 25.1: diagonalisable linear transformation

Let V be a finite dimensional vector space. A linear transformation $T : V \to V$ is **diagonalisable** if there is a basis $\mathcal{B}$ for V such that the matrix $[T]_\mathcal{B}$ is diagonal.

Example 25.2. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with standard matrix given by

$$[T]_\mathcal{S} = \begin{bmatrix} 2 & -3 & 6 \\ 0 & 5 & -6 \\ 0 & 1 & 0 \end{bmatrix}$$

In Example 24.1 we saw that this matrix has eigenvalues 2 and 3. Further we calculated a basis $\{(1,0,0),(0,2,1)\}$ for the $\lambda = 2$ eigenspace and a basis $\{(-3,3,1)\}$ for the $\lambda = 3$ eigenspace. In particular, letting $b_1 = (1,0,0)$, $b_2 = (0,2,1)$ and $b_3 = (-3,3,1)$ we have

$$T(b_1) = 2b_1 \quad T(b_2) = 2b_2 \quad T(b_3) = 3b_3$$

Moreover, $\mathcal{B} = \{b_1, b_2, b_3\}$ is a basis for $\mathbb{R}^3$. To see this note

$$P = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{R_3 - \frac{1}{2}R_2} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 2 & 3 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Then, with respect to the basis $\mathcal{B}$ the matrix of T is diagonal:

$$[T]_\mathcal{B} = [\ [T(b_1)]_\mathcal{B}\ [T(b_2)]_\mathcal{B}\ [T(b_3)]_\mathcal{B}\ ] = [\ [2b_1]_\mathcal{B}\ [2b_2]_\mathcal{B}\ [3b_3]_\mathcal{B}\ ] = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Therefore T is diagonalisable.

Remark. Notice that, in the above example, if we let $A = [T]_\mathcal{S}$, and $D = [T]_\mathcal{B}$, then we have that $A = PDP^{-1}$ since

$$A = [T]_\mathcal{S} = P_{\mathcal{S},\mathcal{B}}[T]_\mathcal{B}P_{\mathcal{B},\mathcal{S}} = PDP^{-1}$$

Definition 25.3: diagonalisable matrix

A matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is **diagonalisable** if there is an invertible matrix $P \in \mathcal{M}_{n,n}(\mathbb{F})$ and a diagonal matrix $D \in \mathcal{M}_{n,n}(\mathbb{F})$ such that

$$A = PDP^{-1}$$

Exercise 228. Let $T : V \to V$ be a linear transformation of a finite-dimensional vector space V, and let $\mathcal{B}$ be any basis of V. Show that T is diagonalisable if and only if $[T]_\mathcal{B}$ is diagonalisable.

Theorem 25.4

Let V be a finite-dimensional vector space. A linear transformation $T : V \to V$ is diagonalisable if and only if there is a basis $\mathcal{B}$ for V with the property that all elements of $\mathcal{B}$ are eigenvectors of T

Proof. Suppose first that $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis for V and that $T(b_i) = \lambda_i b_i$. Then

$$[T]_\mathcal{B} = [\ [T(b_1)]_\mathcal{B}\ \cdots\ [T(b_n)]_\mathcal{B}] = [\ [\lambda_1 b_1]_\mathcal{B}\ \cdots\ [\lambda_n b_n]_\mathcal{B}] = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Conversely, suppose that $\mathcal{C} = \{c_1, \dots, c_n\}$ is a basis of V such that $[T]_\mathcal{C}$ is diagonal and let the entries on the diagonal be $\mu_1, \dots, \mu_n$. Then considering the j-th column of $[T]_\mathcal{C}$ we have that

$$[T(c_j)]_\mathcal{C} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mu_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} = [\mu_j c_j]_\mathcal{C}$$

and therefore $T(c_j) = \mu_j c_j$ and c_j is an eigenvector of T.

□

25.2 How to diagonalise a matrix

Algorithm 25.5: Diagonalise a matrix

Given a matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$, to find an invertible $P \in \mathcal{M}_{n,n}(\mathbb{F})$ and a diagonal $D \in \mathcal{M}_{n,n}(\mathbb{F})$ such that $A = PDP^{-1}$ (or conclude that no such exist):

1. Calculate the eigenvalues of A
2. For each eigenvalue, find a basis for the corresponding eigenspace.
3. Let $\mathcal{B}$ be the union of the eigenspace bases.
4. ▷ If $|\mathcal{B}| < n$, then A if *not* diagonalisable.

   ▷ Otherwise, A is diagonalisable and $A = PDP^{-1}$ with the following $P, D \in \mathcal{M}_{n,n}(\mathbb{F})$: Label the elements of $\mathcal{B}$ as $\mathcal{B} = \{b_1, \dots, b_n\}$ and let $\lambda_i \in \mathbb{F}$ be such that $T(b_i) = \lambda_i b_i$. Take

$$P = [\ [b_1]_\mathcal{S}\ \cdots\ [b_n]_\mathcal{S}] \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Example 25.6. Let $A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$. The characteristic polynomial of A is given by

$$c_A(x) = \det \begin{bmatrix} 1-x & 4 \\ 1 & 1-x \end{bmatrix} = x^2 - 2x - 3 = (x-3)(x+1)$$

The eigenvalues are: -1 and 3.

For eigenvalue -1 we have

$$A-(-1)I_2=\begin{bmatrix}2&4\\1&2\end{bmatrix}\xrightarrow{R_1\leftrightarrow R_2}\begin{bmatrix}1&2\\2&4\end{bmatrix}\xrightarrow{R_2-2R_1}\begin{bmatrix}1&2\\0&0\end{bmatrix}$$

A basis for the $\lambda=-1$ eigenspace is $\{(-2,1)\}$.

For eigenvalue 3 we have

$$A-3I_2=\begin{bmatrix}-2&4\\1&-2\end{bmatrix}\xrightarrow{R_1\leftrightarrow R_2}\begin{bmatrix}1&-2\\-2&4\end{bmatrix}\xrightarrow{R_2+2R_1}\begin{bmatrix}1&-2\\0&0\end{bmatrix}$$

A basis for the $\lambda=3$ eigenspace is $\{(2,1)\}$.

$\mathcal{B}=\{(-2,1),(2,1)\}$ is a basis for $\mathbb{R}^2$ and if we define

$$P=\begin{bmatrix}-2&2\\1&1\end{bmatrix}\quad\text{and}\quad D=\begin{bmatrix}-1&0\\0&3\end{bmatrix}$$

we then have $A=PDP^{-1}$.

Example 25.7. Let $B=\begin{bmatrix}1&4\\0&1\end{bmatrix}\in\mathcal{M}_{2,2}(\mathbb{R})$. The characteristic polynomial of B is given by

$$c_B(x)=\det\begin{bmatrix}1-x&4\\0&1-x\end{bmatrix}=(x-1)(x-1)$$

The only eigenvalue is: 1

We have

$$B-I_2=\begin{bmatrix}0&4\\0&0\end{bmatrix}\xrightarrow{\frac{1}{4}\times R_1}\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

A basis for the $\lambda=-1$ eigenspace is $\{(1,0)\}$.

The matrix B is not diagonalisable.

Example 25.8. Let $M=\begin{bmatrix}6&-5\\8&-6\end{bmatrix}\in\mathcal{M}_{2,2}(\mathbb{R})$.

The characteristic polynomial of M is given by

$$c_M(x)=\det\begin{bmatrix}6-x&-5\\8&-6-x\end{bmatrix}=x^2+4$$

The matrix M has no eigenvalues (in $\mathbb{R}$). The matrix M is not diagonalisable (over $\mathbb{R}$).

Now let $N=\begin{bmatrix}6&-5\\8&-6\end{bmatrix}\in\mathcal{M}_{2,2}(\mathbb{C})$. The characteristic polynomial of N is given by

$$c_N(x)=\det\begin{bmatrix}6-x&-5\\8&-6-x\end{bmatrix}=x^2+4$$

The eigenvalues of N are: $-2i$ and $2i$.

For eigenvalue $\lambda=-2i$ we have

$$N-(-2i)I_2=\begin{bmatrix}6+2i&-5\\8&-6+2i\end{bmatrix}\xrightarrow{R_2-\frac{8}{6+2i}R_1}\begin{bmatrix}6+2i&-5\\0&0\end{bmatrix}\xrightarrow{\frac{1}{6+2i}\times R_1}\begin{bmatrix}1&\frac{1}{4}(-3+i)\\0&0\end{bmatrix}$$

A basis for the $\lambda=-2i$ eigenspace is $\{(3-i,4)\}$.

For eigenvalue $\lambda=2i$ we have

$$N-2iI_2=\begin{bmatrix}6-2i&-5\\8&-6-2i\end{bmatrix}\xrightarrow{R_2-\frac{8}{6-2i}R_1}\begin{bmatrix}6-2i&-5\\0&0\end{bmatrix}\xrightarrow{\frac{1}{6-2i}\times R_1}\begin{bmatrix}1&\frac{1}{4}(-3-i)\\0&0\end{bmatrix}$$

A basis for the $\lambda=2i$ eigenspace is $\{(3+i,4)\}$.

The matrix N is diagonalisable (over $\mathbb{C}$) and we have

$$N=PDP^{-1}\quad\text{with}\quad P=\begin{bmatrix}3-i&3+i\\4&4\end{bmatrix}\quad\text{and}\quad D=\begin{bmatrix}-2i&0\\0&2i\end{bmatrix}$$

25.3 Exercises

229. Decide which of the matrices A in questions 219 above are diagonalisable and, if possible, find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

230. Decide which of the matrices A in questions 221 above are diagonalisable and, if possible, find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

Further material for lecture 25

▷ Let V be a finite dimensional vector space over a field $\mathbb{F}$. Although not all linear transformations $T : V \to V$ are diagonalisable, if the field is $\mathbb{C}$ then there does always exist a basis such that $[T]_{\mathcal{B}}$ is upper triangular. Here's an outline of a proof.

- Use induction on $\dim(V)$. Suppose $\dim(V) = n + 1$.
- Since $\mathbb{F} = \mathbb{C}$, T has an eigenvalue λ and corresponding eigenvector $u \in V \setminus \{\vec{0}\}$.
- Extend to a basis $\{u, b_1, \dots, b_n\}$ for V and let $W = \operatorname{span}\{b_1, \dots, b_n\}$.
- Define $P : V \to W$ by $P(u) = \vec{0}$ and $P(b_i) = b_i$.
- Let $\varphi : W \to W$ be given by $\varphi(w) = P \circ T(w)$.
- By the induction hypothesis, there is a basis $\mathcal{C} = \{w_1, \dots, w_n\}$ for W such that $[\varphi]_{\mathcal{C}}$ is upper triangular.
- $\mathcal{A} = \{u, w_1, \dots, w_n\}$ is a basis for V.
- $T(u) \in \operatorname{span}\{u\}$ and $T(w_i) \in \operatorname{span}\{u, w_1, \dots, w_i\}$ (because $\varphi(w_i) \in \operatorname{span}\{w_1, \dots, w_i\}$)
- Therefore $[T]_{\mathcal{A}}$ is upper triangular.

Diagonalisation II

We look at sufficient conditions for diagonalisability and see why the diagonalisation method given in the last lecture works.

26.1 Distinct eigenvalues

Lemma 26.1

Let $T : V \to V$ be a linear transformation of a vector space V. Suppose that $v_1, \dots, v_k \in V$ are eigenvectors of T and let $\lambda_i \in \mathbb{F}$ be the corresponding eigenvalues. If the λ_i are distinct, then the set $\{v_1, \dots, v_k\}$ is linearly independent.

Proof. We use induction on k.

The base case is $k = 1$. We want to show that $\{v_1\}$ is linearly independent. Since v_1 is an eigenvector, $v_1 \neq \vec{0}$ and it follows that $\{v_1\}$ is linearly independent.

For the induction step, assume that the statement is true for $k = n \in \mathbb{N}$. We want to show that it holds for $k = n + 1$. So suppose that we have eigenvectors $v_1, \dots, v_{n+1} \in V \setminus \{\vec{0}\}$ with distinct eigenvalues $\lambda_1, \dots, \lambda_{n+1} \in \mathbb{F}$. Suppose that $\alpha_1, \dots, \alpha_{n+1} \in \mathbb{F}$ are such that $\sum_{i=1}^{n+1} \alpha_i v_i = \vec{0}$. We want to show that $\alpha_i = 0$ for all $i \in \{1, \dots, n+1\}$. We have

$$\begin{aligned} & \sum_{i=1}^{n+1} \alpha_i v_i = \vec{0} && (1) \\ \implies & T\Big(\sum_{i=1}^{n+1} \alpha_i v_i\Big) = T(\vec{0}) \implies \sum_{i=1}^{n+1} \alpha_i T(v_i) = \vec{0} \\ \implies & \sum_{i=1}^{n+1} \alpha_i \lambda_i v_i = \vec{0} && (2) \end{aligned}$$

Multiplying (1) by λ_{n+1} and then subtracting (2) yields

$$\begin{aligned} & \lambda_{n+1}\left(\sum_{i=1}^{n+1} \alpha_i v_i\right) - \sum_{i=1}^{n+1} \alpha_i \lambda_i v_i = \vec{0} \\ \implies & \sum_{i=1}^{n+1} \alpha_i(\lambda_{n+1} - \lambda_i) v_i = \vec{0} \implies \sum_{i=1}^{n} \alpha_i(\lambda_{n+1} - \lambda_i) v_i = \vec{0} \\ \implies & \forall i \in \{1, \dots, n\}, \quad \alpha_i(\lambda_{n+1} - \lambda_i) = 0 \qquad (\{v_1, \dots, v_n\} \text{ is linearly independent}) \\ \implies & \forall i \in \{1, \dots, n\}, \quad \alpha_i = 0 \qquad (\text{since } \lambda_{n+1} \neq \lambda_i) \end{aligned}$$

Finally, note that we now have, from (1), that $\alpha_{n+1} v_{n+1} = \vec{0}$. Since $v_{n+1} \neq \vec{0}$ this implies that $\alpha_{n+1} = 0$ also. Therefore the result holds for $k = n + 1$.

Therefore, by mathematical induction, the result holds for all $k \in \mathbb{N}$.

□

Example 26.2. For the matrix of Example 24.1, $(1,2,1)$ is an eigenvector with eigenvalue 2 and $(3,-3,-1)$ is an eigenvector with eigenvalue 3. The set $\{(1,2,1),(3,-3,-1)\}$ is linearly independent.

Exercise 231. Let $T : V \to V$ be a linear transformation of a finite dimensional vector space V. Suppose that $\lambda_1 \neq \lambda_2$ are two eigenvalues of T and let the corresponding eigenspaces be W_1 and W_2.

(a) Show that $W_1 \cap W_2 = \{\vec{0}\}$

(b) Let $\mathcal{B}_1$ be a basis for W_1 and let $\mathcal{B}_2$ be a basis for W_2. Show that $\mathcal{B}_1 \cup \mathcal{B}_2$ is linearly independent.

The following result justifies the technique given in the last lecture for diagonalising a matrix.

Theorem 26.3

Let $T : V \to V$ be a linear transformation of an n-dimensional vector space V. Then T is diagonalisable if and only if the geometric multiplicities sum to n.

Proof. Let the eigenvalues of T be $\lambda_1,\dots,\lambda_k$. Denote by g_i and a_i the geometric and algebraic multiplicities of the eigenvalue λ_i. We know from Proposition 24.7 that $g_i \leqslant a_i$. We also know that $a_1+\cdots+a_k \leqslant n$.

Suppose that T is diagonalisable. Then there exists a basis $\mathcal{B}$ for V with the property that each element of $\mathcal{B}$ is an eigenvector of T. For $i \in \{1,\dots,k\}$ let $\mathcal{B}_i = \{b \in \mathcal{B} \mid b \text{ has eigenvalue } \lambda_i\}$. Note that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ and that $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ if $i \neq i$. We have

$$n = |\mathcal{B}| = |\mathcal{B}_1| + |\mathcal{B}_2| + \cdots + |\mathcal{B}_k| \leqslant g_1 + g_2 + \cdots + g_k \leqslant a_1 + a_2 + \cdots + a_k \leqslant n$$

Therefore $g_1 + \cdots + g_k = n$.

For the converse, suppose now that $g_1+\cdots+g_k = n$. Let $\mathcal{C}_i$ be a basis for the λ_i eigenspace. Then $|\mathcal{C}_i| = g_i$. From Exercise 231 we know that $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ if $i \neq j$. Therefore, if we define $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_k$, we have

$$|\mathcal{C}| = |\mathcal{C}_1| + |\mathcal{C}_1| + \cdots + |\mathcal{C}_k| = g_1 + g_2 + \cdots + g_k = n$$

All elements of $\mathcal{C}$ are eigenvectors. If we show that $\mathcal{C}$ is linearly independent we will be done since it would follow that $\mathcal{C}$ is a basis.

Denote the elements of $\mathcal{C}_i = \{c_{i1},\dots,c_{i,g_i}\}$.

$$\begin{aligned}
\sum_{i=1}^{k}\sum_{j=1}^{g_i} \alpha_{i,j} c_{i,j} = \vec{0} &\implies \sum_{i=1}^{k} u_i = \vec{0} && \text{where we define } u_i = \sum_{j=1}^{g_i} \alpha_{i,j} c_{i,j} \\
&\implies \forall i,\ u_i = \vec{0} && \text{by Lemma 26.1} \\
&\implies \forall i\, \forall j,\ \alpha_{i,j} = 0 && \text{since } \mathcal{C}_i \text{ is linearly independent}
\end{aligned}$$

□

Corollary 26.4

Let $T : V \to V$ be a linear transformation of an n-dimensional vector space V and let $\lambda \in \mathbb{F}$ be an eigenvalue of T. If the geometric multiplicity of λ is strictly less than its algebraic multiplicity, then T is not diagonalisable. □

26.2 A sufficient condition for diagonalisability

Proposition 26.5

Let V be an n-dimensional vector space over $\mathbb{F}$. Let $T : V \to V$ be a linear transformation. If T has n distinct eigenvalues in $\mathbb{F}$, then T is diagonalisable.

Note. The converse is false! Linear transformations that have repeated eigenvalues can still be diagonalisable.

Proof. Suppose that $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ are eigenvalues of T and that $\lambda_i \neq \lambda_j$ when $i \neq j$. Let m_i be the geometric multiplicity of the eigenvalue λ_i. Since $m_i \geqslant 1$, we must must that $m_1 + \cdots + m_n = n$. Diagonalisability then follows from Theorem 26.3. □

Example 26.6. 1. $\left[\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{smallmatrix}\right] \in \mathcal{M}_{4,4}(\mathbb{F}_5)$ is diagonalisable (and we can see that without the need for any calculations).

2. $\left[\begin{smallmatrix} 1-3i & 4i \\ -2i & 1+3i \end{smallmatrix}\right] \in \mathcal{M}_{2,2}(\mathbb{C})$ has eigenvalues $1+i, 1-i$ and is therefore diagonalisable.

26.3 Exercises

232. Give an example of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that is diagonalisable and has only one eigenvalue.

233. Let $E = \begin{bmatrix} 4 & -2 & -1 \\ -2 & 4 & -1 \\ -1 & -1 & 1 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{C})$

    (a) Calculate the characteristic polynomial of E.

    (b) Find the eigenvalues of E.

    (c) Without finding any eigenspaces, explain why E is diagonalisable and write down a diagonal matrix D that is similar to E.

234. (a) Let $T : \mathbb{Q}^2 \to \mathbb{Q}^2$ be the linear transformation given by $T(x, y) = (x + y, x - y)$. Find the eigenvalues of T. Is T diagonalisable? If it is diagonalisable, give a diagonal matrix D such that $[T]_\mathcal{B} = D$ with respect to some basis $\mathcal{B}$. (You are not being asked to find $\mathcal{B}$.)

    (b) Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $S(x, y) = (x + y, x - y)$. Find the eigenvalues of S. Is S diagonalisable? If it is diagonalisable, give a diagonal matrix D such that $[S]_\mathcal{B} = D$ with respect to some basis $\mathcal{B}$. (You are not being asked to find $\mathcal{B}$.)

Further material for lecture 26

- ▷ References for diagonalisation

  Elementary Linear Algebra by Anton and Rorres, §5.2, p302
  The Art of Proof by Beck and Geoghegan, §5.C, p155

Powers of a matrix and the Cayley-Hamilton theorem

27.1 Matrix powers

In applications one often comes across the need to apply a transformation many times. If the transformation can be represented by a diagonalisable matrix A, it's much easier to compute A^k and thus the effect of the k-th application of the transformation. The first point to appreciate is that computing powers of a diagonal matrix D is easy.

Example 27.1. Let $D = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right]$. Then $D^2 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{smallmatrix}\right]$, $D^3 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & -27 & 0 \\ 0 & 0 & 8 \end{smallmatrix}\right]$, $D^k = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & (-3)^k & 0 \\ 0 & 0 & 2^k \end{smallmatrix}\right]$

Example 27.2. The matrix $A = \left[\begin{smallmatrix} 1 & 4 \\ 1 & 1 \end{smallmatrix}\right] \in \mathcal{M}_{2,2}(\mathbb{R})$ is diagonalisable and $A = PDP^{-1}$ with $P = \left[\begin{smallmatrix} -2 & 2 \\ 1 & 1 \end{smallmatrix}\right]$ and $D = \left[\begin{smallmatrix} -1 & 0 \\ 0 & 3 \end{smallmatrix}\right]$. Therefore,

$$A^k = PD^kP^{-1} = \frac{1}{4}\begin{bmatrix} -2 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} (-1)^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} (-1)^k + 3^k & 2(-1)^{k+1} + 3^{k+1} \\ \frac{1}{2}((-1)^{k+1} + 3^k) & (-1)^k + 3^k \end{bmatrix}$$

Example 27.3. As an application we will investigate a simple model of population movement between Victoria and Queensland. Assume that 2% of Victorians move to Queensland each year, 1% of Queenslanders move to Victoria each year and everybody else stays put. This is an example of a (discrete-time) 'Markov process'. We investigate what happens in the long term under these assumptions.

Let v_i be the Victorian population (in millions) after i years and q_i be the Queensland population (in millions) after i years.

$$\begin{aligned} v_{i+1} &= 0.98v_i + 0.01q_i \\ q_{i+1} &= 0.02v_i + 0.99q_i \end{aligned}$$

Which we can write as $\begin{bmatrix} v_{i+1} \\ q_{i+1} \end{bmatrix} = A\begin{bmatrix} v_i \\ q_i \end{bmatrix}$ where $A = \dfrac{1}{100}\begin{bmatrix} 98 & 1 \\ 2 & 99 \end{bmatrix}$

Then $\begin{bmatrix} v_k \\ q_k \end{bmatrix} = A^k\begin{bmatrix} v_0 \\ q_0 \end{bmatrix}$.

Diagonalisation gives $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & 0.97 \end{bmatrix}$ so

$$\begin{aligned} A^k &= \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & (0.97)^k \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} \\ &= \frac{1}{3}\begin{bmatrix} 1 + 2(0.97)^k & 1 - (0.97)^k \\ 2 - 2(0.97)^k & 2 + (0.97)^k \end{bmatrix} \end{aligned}$$

$$\begin{aligned} v_k &= \frac{1}{3}(1 + 2(0.97)^k)v_0 + \frac{1}{3}(1 - (0.97)^k)q_0 \\ q_k &= \frac{1}{3}(2 - 2(0.97)^k)v_0 + \frac{1}{3}(2 + (0.97)^k)q_0 \end{aligned}$$

Therefore $v_k \to \frac{1}{3}(v_o + q_0)$ and $q_k \to \frac{2}{3}(v_0 + q_0)$.

27.2 Cayley-Hamilton Theorem

Theorem 27.4: Cayley-Hamilton Theorem

Given a matrix $A \in \mathcal{M}_{n,n}(\mathbb{F})$, let its characteristc polynomial be

$$c_A(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$$

Then

$$a_0 I_n + a_1 A + \cdots + a_{n-1}A^{n-1} + A^n = 0$$

That is, every square matrix satisfies its own characteristic equation.

Example 27.5. The characteristic polynomial of $A = \left[\begin{smallmatrix} 3 & 2 \\ 1 & 4 \end{smallmatrix}\right]$ is $c_A(x) = x^2 - 7x + 10$ and

$$A^2 - 7A + 10I_2 = \begin{bmatrix} 11 & 14 \\ 7 & 18 \end{bmatrix} - 7\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} + 10\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Proof of Cayley-Hamilton for diagonalisable matrices. First suppose that $D \in \mathcal{M}_{n,n}(\mathbb{F})$ is diagonal and let the entries on the diagonal be $\lambda_1 \ldots, \lambda_n \in \mathbb{F}$. Then $c_D(x) = (x - \lambda_1)(x - \lambda_2)\ldots(x - \lambda_n)$. To see that $(D - \lambda_1 I)(D - \lambda_2 I)\cdots(D - \lambda_n I) = 0$ note that the entry in the i-th row and i-th column of $D - \lambda_i I$ is equal to zero.

Now suppose that $A = PDP^{-1}$ for some invertible matrix D and diagonal matrix D. Then $c_A(x) = c_D(x)$ by Lemma 24.4. Writing $c_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ we have

$$\begin{aligned} A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n &= PA^nP^{-1} + a_{n-1}PA^{n-1}P^{-1} + \cdots + a_1 PAP^{-1} + a_0 PI_nP^{-1} \\ &= P(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 I_n)P^{-1} \\ &= P0P^{-1} = 0 \end{aligned}$$

□

Remark. A slightly more involved argument can be used to show that upper triangular matrices satisfy the statement of the theorem. The full proof (at least for $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{R}$) is then completed by showing that all matrices in $\mathcal{M}_{n,n}(\mathbb{C})$ are similar to a matrix in upper triangular form.

The Cayley-Hamilton Theorem can be used to show the following.

Corollary 27.6

Let $A \in \mathcal{M}_{n,n}(\mathbb{F})$. For all $m \geqslant 0$, $\quad A^m$ can be expressed as a linear combination of $I, A, \ldots, A^{n-1}$.

If A is invertible, then for all $m \geqslant 0$,
A^{-m} can be expressed as a linear combination of $I, A, \ldots, A^{n-1}$

Example 27.7. Given that the matrix

$$A = \begin{bmatrix} 9 & 18 & -24 \\ 7 & 20 & -24 \\ 7 & 21 & -25 \end{bmatrix} \quad \text{has characteristic polynomial} \quad c_A(x) = x^3 - 4x^2 + x + 6$$

We know that

$$\begin{aligned}
A^3 &= 4A^2 - A - 6I_3\\
A^4 &= 4A^3 - A^2 - 6A = 4(4A^2 - A - 6I_3) - A^2 - 6A = 15A^2 - 10A - 24I_3\\
A^5 &= 4A^4 - A^3 - 6A^2 = 4(15A^2 - 10A - 24I_3) - (4A^2 - A - 6I_3) - 6A^2\\
&= 50A^2 - 39A - 90I_3\\
A^{-1} &= \frac{-1}{6}A^2 + \frac{2}{3}A - \frac{1}{6}I_3
\end{aligned}$$

27.3 Exercises

235. By first diagonalising the matrix, find A^5, where A is

(a) $\begin{bmatrix} 3 & -2 \\ 2 & -2 \end{bmatrix}$ (b) $\begin{bmatrix} 9 & 18 & -24 \\ 7 & 20 & -24 \\ 7 & 21 & -25 \end{bmatrix}$ (c) $\frac{1}{8}\begin{bmatrix} 8 & 1 & 27 & 5 \\ 0 & 18 & 14 & -6 \\ 0 & 2 & -18 & -6 \\ 0 & 8 & 8 & -8 \end{bmatrix}$

[*Note: The matrix has eigenvalues* $3, 2, -1$ *in (b), and* $1, 2, -1, -2$ *in (c).*]

236. Suppose the nth pass through a manufacturing process is modelled by the linear equations $x_n = A^n x_0$, where x_0 is the initial state of the system and

$$A = \frac{1}{5}\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

Show that

$$A^n = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + (\frac{1}{5})^n \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Then, with the initial state $x_0 = \begin{bmatrix} p \\ 1-p \end{bmatrix}$, calculate $\lim_{n\to\infty} x_n$.

237. The Fibonacci sequence

$$0, 1, 1, 2, 3, 5, 8, 13, \ldots$$

is given by the difference equation $F_{k+2} = F_{k+1} + F_k$ and the initial conditions $F_0 = 0, F_1 = 1$.

(a) Letting $u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$, show that $u_k = A^k u_0$ where $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.

(b) Show that A is diagonalisable, and find P and D such that $A = PDP^{-1}$.

(c) Use your answer to (b) to calculate A^k.

(d) Use your answer to (c) to show that

$$F_k = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k\right]$$

(e) (Bonus question!) Show that

$$\lim_{k\to\infty} \frac{F_{k+1}}{F_k} = \frac{1+\sqrt{5}}{2} \qquad \text{(the 'golden ratio')}$$

238. Two companies, Lemon and LIME, introduce a new type of computer. At the start, their shares of the market are 60% and 40%. After a year, Lemon kept 85% of its customers and gained 25% of LIME's customers; LIME gained 15% of Lemon's customers and kept 75% of its customers. Assume that the total market is constant and that the same fractions shift among the firms every year.

(a) Write down the market share shift as a system of linear equations.

(b) Express the shift in matrix form.

(c) Find the market shares after 5 and 10 years.

(d) Show that the market eventually reaches a steady state, and give the limit market shares.

239. Verify the Cayley-Hamilton Theorem for the following matrices.

(a) $\begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$

240. Use the Cayley-Hamilton Theorem to calculate the inverse of the matrix $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$.

241. For each matrix, find a non-zero polynomial satisfied by the matrix.

(a) $\begin{bmatrix} 2 & 5 \\ 1 & -3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 4 & -3 \\ 0 & 3 & 1 \\ 0 & 2 & -1 \end{bmatrix}$

242. (a) Give an example of a matrix $A \in \mathcal{M}_{3,3}(\mathbb{R})$ and a quadratic polynomial $p \in \mathbb{R}[x]$ such that $p(A) = 0$.

(b) Give an example of a matrix $B \in \mathcal{M}_{3,3}(\mathbb{R})$ and two *different* cubic polynomials $q, r \in \mathbb{R}[x]$ such that $q(A) = r(A) = 0$.

Further material for lecture 27

▷ Sketch of a proof of the Cayley-Hamilton theorem for upper-triangular matrices

Claim. *If $A \in \mathcal{M}_{n,n}(\mathbb{F})$ is upper-triangular, then A satisfies its own characteristic polynomial*

Sketch of proof.

$$A = \begin{bmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n \end{bmatrix} \qquad c_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$$

Let $e_i \in \mathcal{M}_{n,1}$ be the column matrix with a 1 in the i-th row and zeros everywhere else. Define subspaces $V_0, V_1, \dots, V_n \leqslant \mathcal{M}_{n,1}(\mathbb{F})$ by $V_0 = \{\vec{0}\}$ and $V_k = \operatorname{span}\{e_1, \dots, e_k\}$for $k \in \{1, 2, \dots, n\}$. Note that

$$\forall u \in V_k, \qquad (A - \lambda_k I_n)u \subseteq V_{k-1}$$

It follows that

$$\forall u \in \mathcal{M}_{n,1}(\mathbb{F}), \qquad (A-\lambda_1 I_n)\cdots(A-\lambda_n I_n)u \subseteq V_0$$

Since this holds for all $u \in \mathcal{M}_{n,1}(\mathbb{F})$, we conclude that $(A-\lambda_1 I_n)\cdots(A-\lambda_n I_n) = 0$. □

▷ Combining this with the result that all elements of $\mathcal{M}_{n,n}(\mathbb{C})$ are similar to an upper triangular matrix gives a proof of Cayley-Hamilton in the case $\mathbb{F} = \mathbb{C}$. In fact, the same argument works for any 'algebraically closed' field, and it's then not much work to extend to an arbitrary $\mathbb{F}$.

Geometry in Euclidean space

We want to be able to consider geometric notions such as length and angle in a vector space. Before defining the general notion of an inner product space, we revise the familiar context of $\mathbb{R}^n$.

28.1 Dot product in $\mathbb{R}^n$

To work with length, distance, and angle we need more than just the vector space properties of $\mathbb{R}^n$.

Definition 28.1: Dot product

Let $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. We define the **dot product** of u and v to be

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \in \mathbb{R}$$

Remark. Denoting the standard basis of $\mathbb{R}^n$ by $\mathcal{S}$, we have $u \cdot v = [u]_{\mathcal{S}}^T [v]_{\mathcal{S}}$

Exercise 243. Use the definition of the dot product to verify the following properties. For all $u, v \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

(a) $u \cdot v = v \cdot u$

(b) $(\alpha u) \cdot v = \alpha(u.v)$

(c) $u \cdot (v + w) = u \cdot v + u \cdot w$

(d) $u \cdot u \geqslant 0$

(e) $u \cdot u = 0 \iff u = \vec{0}$

Definition 28.2: Length and distance

The **length** (or magnitude or norm) of a vector $u = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ is given by

$$\|u\| = \sqrt{u \cdot u} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

A vector $u \in \mathbb{R}^n$ is called a **unit vector** if $\|u\| = 1$.

The **distance** between two vectors $u, v \in \mathbb{R}^n$ is given by $d(u, v) = \|v - u\|$

Example 28.3. $\|(1, -2, 2)\| = 3$, $\quad \|\frac{1}{3}(1, -2, 2)\| = 1$, $\frac{1}{3}(1, -2, 2)$ is a unit vector.

Example 28.4. The distance between two points $P(1, 3, -1)$ and $Q(2, 1, -1)$ is the distance between their position vectors: $d(P, Q) = d((1, 3, -1), (2, 1, -1)) = \|(1, 3, -1) - (2, 1, -1)\| = \|(-1, 2, 2)\| = 3$

Definition 28.5: Angle

The **angle** θ between two non-zero vectors $u, v \in \mathbb{R}^n$ is given by the expression

$$u \cdot v = \|u\|\|v\| \cos\theta \qquad \text{where } \ 0 \leq \theta \leq \pi$$

We say that u and v are **orthogonal** (or **perpendicular**) if $u \cdot v = 0$.
We say that u and v are **parallel** if one is a scalar multiple of the other.

28.2 Cross product in $\mathbb{R}^3$

Definition 28.6: Cross product

Let $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in \mathbb{R}^3$. The **cross product** (or **vector product**) of u and v is the vector given by

$$u \times v = (u_2v_3 - u_3v_2)\mathbf{i} + (u_3v_1 - u_1v_3)\,\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}$$

Remark. A convenient way to remember this is as a "determinant"

$$u \times v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k}$$

using cofactor expansion along the first row.

Exercise 244 (Properties of the cross product)**.** Show (directly from the definitions) that for any vectors u, v and $w \in \mathbb{R}^3$, and scalar $\alpha \in \mathbb{R}$:

(a) $u \times v = -(v \times u)$

(b) $u \times (v + w) = (u \times v) + (u \times w)$

(c) $(\alpha u) \times v = \alpha(u \times v)$

(d) $u \times u = \vec{0}$

(e) $u \cdot (u \times v) = 0$

Example 28.7. We can use the cross product to find a vector perpendicular to both $(2, 3, 1)$ and $(1, 1, 1)$.

$$(2, 3, 1) \times (1, 1, 1) = (2, -1, -1)$$

Note that $(2, -1, -1) \cdot (2, 3, 1) = 0$ and $(2, -1, -1) \cdot (1, 1, 1) = 0$

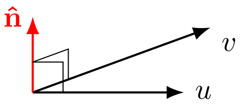
Lemma 28.8

The cross product satisfies

$$u \times v = \|u\|\|v\| \sin(\theta)\, \hat{\mathbf{n}}$$

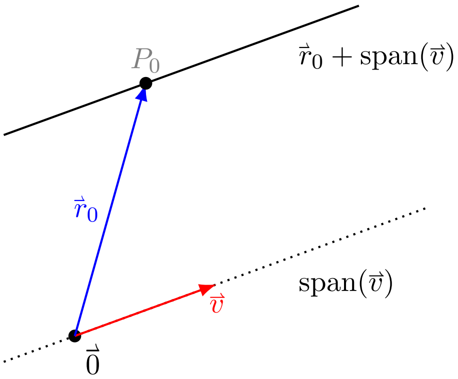
where

- ▷ $\hat{\mathbf{n}}$ is a unit vector perpendicular to both u and v and points in the direction given by the 'right-hand rule'
- ▷ $\theta \in [0, \pi]$ is the angle between u and v



28.3 Lines in $\mathbb{R}^3$

Lines through the origin are exactly the 1-dimensional subspaces of $\mathbb{R}^3$. Every line in $\mathbb{R}^3$ is a translate of a 1-dimensional subspace of $\mathbb{R}^3$ in the following way.



All lines in $\mathbb{R}^3$ are of the form

$$\vec{r}_0 + \text{span}(\vec{v}) = \vec{r}_0 + \{t\vec{v} \mid t \in \mathbb{R}\}$$

for some (not unique) $\vec{r}_0, \vec{v} \in \mathbb{R}^3$

The **vector equation** of a line through a point P_0 in the direction determined by a vector $\vec{v}$ is (where $\vec{r}_0 = \overrightarrow{OP_0}$):

$$\vec{r} = \vec{r}_0 + t\vec{v} \qquad t \in \mathbb{R}$$

Letting $\vec{r} = (x, y, z)$, $\vec{r}_0 = (x_0, y_0, z_0)$ and $\vec{v} = (a, b, c)$ and then equating coordinates gives the **parametric equations** for the line:

$$\begin{aligned} x &= x_0 + ta \\ y &= y_0 + tb \qquad t \in \mathbb{R} \\ z &= z_0 + tc \end{aligned}$$

If $a \neq 0$, $b \neq 0$ and $c \neq 0$, we can solve the parametric equations for t and equate. This gives the **cartesian form** of the straight line:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

Example 28.9. Consider the line passing through the points $P(-1, 2, 3)$ and $Q(4, -2, 5)$. It has

vector equation: $$(x, y, z) = (-1, 2, 3) + t(5, -4, 2), \qquad t \in \mathbb{R}$$

parametric equations: $$\begin{aligned} x &= -1 + 5t \\ y &= 2 - 4t \qquad t \in \mathbb{R} \\ z &= 3 + 2t \end{aligned}$$

cartesian form: $$\frac{x + 1}{5} = \frac{y - 2}{-4} = \frac{z - 3}{2}$$

Definition 28.10

Two lines are said to:

- ▷ **intersect** if there is a point lying in both
- ▷ be **parallel** if their direction vectors are parallel
- ▷ be **skew** if they do not intersect and are not parallel

The **angle** between two lines is the angle between their direction vectors

Example 28.11. Consider the three lines having parametric equations

$$L_1 : \begin{aligned} x &= 1 + t \\ y &= 2 - 4t \quad t \in \mathbb{R} \\ z &= 3 + 2t \end{aligned} \qquad L_2 : \begin{aligned} x &= -4 + 3t \\ y &= -6 + 2t \quad t \in \mathbb{R} \\ z &= 3 + t \end{aligned} \qquad L_3 : \begin{aligned} x &= \tfrac{1}{2}t \\ y &= 1 - 2t \quad t \in \mathbb{R} \\ z &= 2 + t \end{aligned}$$

Then L_1 and L_2 intersect, L_1 and L_3 are parallel, L_2 and L_3 are skew.

Distance between a point and a line

Given a point with position vector $\vec{p}$ and a line with vector equation $\vec{r} = \vec{r}_0 + t\vec{u} \quad t \in \mathbb{R}$, the distance from the point to the line is given by

$$d = \frac{\|\vec{u} \times (\vec{p} - \vec{r}_0)\|}{\|\vec{u}\|}$$

Exercise 245. Use Lemma 28.8 to derive the above expression for the distance between a point and a line.

Distance between two skew lines

Given two skew lines having vector equations

$$\vec{r} = \vec{r}_1 + t\vec{u} \quad t \in \mathbb{R} \qquad \text{and} \qquad \vec{r} = \vec{r}_2 + t\vec{v} \quad t \in \mathbb{R}$$

The distance between them is given by

$$d = \frac{|(\vec{u} \times \vec{v}) \cdot (\vec{r}_2 - \vec{r}_1)|}{\|\vec{u} \times \vec{v}\|}$$

28.4 Planes in $\mathbb{R}^3$

Planes through the origin are exactly the 2-dimensional subspaces of $\mathbb{R}^3$. All planes in $\mathbb{R}^3$ are of the form $\vec{r}_0 + W$ where W is a 2-dimensional subspace of $\mathbb{R}^3$.

The **vector equation** of a plane through a point P_0 and parallel to both $\vec{u}, \vec{v} \in \mathbb{R}^3$ is (where $\vec{r}_0 = \overrightarrow{OP_0}$):

$$\vec{r} = \vec{r}_0 + s\vec{u} + t\vec{v} \qquad s, t \in \mathbb{R}$$

The **cartesian form** of a plane is:

$$ax + by + cz = d$$

Where $a, b, c, d \in \mathbb{R}$ and the vector (a, b, c) is perpendicular to both $\vec{u}$ and $\vec{v}$. Such a vector is called a **normal** to the plane.

Examples 28.12.

1. The plane perpendicular to the direction $(1, 2, 3)$ and through the point $(4, 5, 6)$ is given by $x + 2y + 3z = d$ where $d = 1 \times 4 + 2 \times 5 + 3 \times 6$. That is, $x + 2y + 3z = 32$

2. Consider the plane perpendicular to $(1, 0, -2)$ and containing the point $(1, -1, -3)$.

   The cartesian equation is: $x - 2z = 7$
   A vector equation is: $(x, y, x) = (1, -1, -3) + s(0, 1, 0) + t(2, 0, 1) \quad s, t \in \mathbb{R}$

3. Consider the plane with vector equation

   $$(x, y, z) = (1, 2, 3) + s(2, 3, 1) + t(1, 1, 1), \quad s, t \in \mathbb{R}$$

   To find a catesian equation for the plane, we need a normal to the plane. For this we can the fact the $(2, 3, 1) \times (1, 1, 1) = (2, -1, -1)$ is orthogonal to both $(2, 3, 1)$ and $(1, 1, 1)$. The cartesian equation is of the form $2x - y - z = d$. Since $(1, 2, 3)$ lies in the plane, we have that $d = 2 - 2 - 3 = -3$. The catesian equation is

   $$2x - y - z = -3$$

Distance between a point and a plane

Given a point $\vec{p}$ and a plane that has normal vector $\vec{n}$ and contains a point $\vec{r}_0$, the distance from the point to the plane is given by

$$d = \frac{|(\vec{p} - \vec{r}_0) \cdot \vec{n}|}{\|\vec{n}\|}$$

28.5 Exercises

246. Find the following dot products:

(a) $(1,1,1)\cdot(2,1,-3)$ (b) $(2,1,1)\cdot(1,-3,7)$ (c) $(\sqrt{2},\pi,1)\cdot(\sqrt{2},-2,3)$

247. Find the angle between the following pairs of vectors in $\mathbb{R}^3$:

(a) $(1,0,0),\ (0,0,4)$ (b) $(1,-1,0),\ (0,1,1)$ (c) $(2,-2,2),\ (-1,0,2)$

248. Let $\mathbf{u}=(3,-1,4),\quad \mathbf{v}=-\mathbf{i}-3\mathbf{j}+\mathbf{k},\quad \mathbf{w}=(-1,1,2)$.
Find (if they exist):

(a) $\mathbf{u}\cdot\mathbf{v}$ (d) $(\mathbf{v}\times\mathbf{u})\cdot(-\mathbf{w})$ (g) $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$
(b) $(3\mathbf{u})\cdot(-2\mathbf{v})$ (e) $\mathbf{v}\times 2\mathbf{u}$ (h) $\mathbf{u}\cdot(\mathbf{v}\cdot\mathbf{w})$
(c) $\mathbf{u}\times\mathbf{v}$ (f) $\mathbf{u}\times(\mathbf{v}\cdot\mathbf{w})$ (i) $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$

249. Find the values of x such that the following pairs of vectors are (i) orthogonal and (ii) parallel.

(a) $(x,1-2x,3)$ and $(1,-x,3x)$ (b) $(x,x,-1)$ and $(1,x,6)$

250. Write down the equation for the following lines in both vector and cartesian form.

(a) the line passing through $P(2,1,-3)$ and parallel to $\mathbf{v}=(1,2,2)$
(b) the line through $P(2,-3,1)$ and parallel to the x-axis
(c) the line passing through the points $P(2,0,-2)$ and $Q(1,4,2)$
(d) the line through $P(2,4,5)$ and perpendicular to the plane $5x-5y-10z=2$

251. Determine whether the lines L_1 and L_2 are parallel, intersecting or skew (not parallel or intersecting). If they intersect, find the point of intersection. Let the parameters $s,t\in\mathbb{R}$.

(a) $L_1: x=-6t,\ y=1+9t,\ z=-3t$ and $L_2: x=1+2s,\ y=4-3s,\ z=s$
(b) $L_1: x=1+t,\ y=1-t,\ z=2t$ and $L_2: x=2-s,\ y=s,\ z=2$
(c) $L_1: \dfrac{x-4}{2}=\dfrac{y+5}{4}=\dfrac{z-1}{-3}$ and $L_2: x-2=\dfrac{y+1}{3}=\dfrac{z}{2}$

252. Find the equations of the following planes in both cartesian and (vector) parametric form:

(a) the plane through the point $(1,4,5)$ and perpendicular to the vector $(7,1,4)$
(b) the plane through the point $(6,5,-2)$ and parallel to the plane $x+y-z+1=0$
(c) the plane through the origin and the points $(1,1,1)$ and $(1,2,3)$
(d) the plane that passes through the point $(1,6,-4)$ and contains the line

$$x=1+2t,\ y=2-3t,\ z=3-t \text{ where } t\in\mathbb{R}$$

253. (a) Show that three points A, B and C are collinear if and only if $\overrightarrow{AB}\times\overrightarrow{AC}=\mathbf{0}$. Are the points $A(1,2,3)$, $B(3,1,0)$ and $C(9,-2,-9)$ collinear? If yes, find the equation of the line containing these points.

(b) Show that four points A, B, C and D are coplanar if and only if

$$\overrightarrow{AB}\cdot(\overrightarrow{AC}\times\overrightarrow{AD})=0.$$

Are the points $A(1,1,1)$, $B(2,1,3)$, $C(3,2,1)$ and $D(4,2,3)$ coplanar? If yes, find the equation of the plane containing these points.

254. (a) Find the point of intersection of the line $\mathbf{r}(t) = (2,1,1) + t(-1,0,4);\ t \in \mathbb{R}$ with the plane $x - 3y - z = 1$.

(b) Find the point of intersection of the line $x = 1 + t, y = 2t, z = 3t;\ t \in \mathbb{R}$ with the plane $x + y + z = 1$.

255. Find the angle between:

(a) the lines $x - 3 = 2 - y,\ z = 1$ and $x = 7,\ y - 2 = z - 5$

(b) the planes $2x + y + 3z = 0$ and $3x - 2y + 4z - 4 = 0$

(c) the line $x = 2t - 7, y = 4t - 6, z = t - 5;\ t \in \mathbb{R}$ and a vector normal to the plane $x + 2y - 4z = 0$

256. Given the line ℓ determined by the equations $2x - y + z = 0$, $x + z - 1 = 0$, and M the point $(1, 3, -2)$, find a cartesian equation of the plane:

(a) passing through M and ℓ

(b) passing through M and orthogonal to ℓ

Extra material for lecture 28

- The cross product is not associative. For example

$$((1,0,0)\times(0,1,0))\times(1,1,0)=(-1,1,0)\neq(0,1,0)=(1,0,0)\times((0,1,0)\times(1,1,0))$$

Inner products

An inner product on a vector space is a generalisation of the dot product on $\mathbb{R}^n$ seen in the last lecture. It will be used to define geometric notions such as length and angle.

When dealing with inner products, the field $\mathbb{F}$ will always be either $\mathbb{R}$ or $\mathbb{C}$. Given an element $\alpha \in \mathbb{F}$, we denote by $\overline{\alpha}$ its complex conjugate and by $|\alpha|$ its absolute value.

29.1 Definition of inner product

Definition 29.1: Inner product

Let $\mathbb{F}$ be one of $\mathbb{R}$ or $\mathbb{C}$. Let V be a vector space over $\mathbb{F}$.
An **inner product** on V is a function $V \times V \to \mathbb{F}$ (with the image of (u, v) being denoted $\langle u, v\rangle$) that satisfies the following conditions. For all $u, v, w \in V$ and $\alpha \in \mathbb{F}$:

1) $\langle u, v\rangle = \overline{\langle v, u\rangle}$
2) $\alpha\langle u, v\rangle = \langle \alpha u, v\rangle$
3) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
4) (a) $\langle u, u\rangle \geqslant 0$
   (b) $\langle u, u\rangle = 0 \Rightarrow u = \vec{0}$

The case in which $\mathbb{F} = \mathbb{R}$ is sometimes referred to as a **real inner product**.

The case in which $\mathbb{F} = \mathbb{C}$ is sometimes referred to as an **Hermitian inner product**.

A vector space V together with a fixed inner product is called an **inner product space**.

Note.

- The first condition implies that $\forall u \in V, \langle u, u\rangle \in \mathbb{R}$ and therefore the inequality in 4(a) makes sense.
- The first and second conditions together imply that $\forall u, v \in V\ \forall \alpha \in \mathbb{F}, \langle u, \alpha v\rangle = \overline{\alpha}\langle u, v\rangle$.
- It's possible to have many different inner products on the same vector space.

Exercise 257. Show that $\forall v \in V$, $\langle \vec{0}, v\rangle = 0$.

Examples 29.2.

1. The dot product on $\mathbb{R}^n$ is an inner product

$$\langle (u_1, \dots, u_n), (v_1, \dots, v_n)\rangle = u_1v_1 + \cdots + u_nv_n$$

2. The **Hermitian dot product** on $\mathbb{C}^n$ is

$$\langle (u_1, \dots, u_n), (v_1, \dots, v_n)\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n}$$

3. The following is an inner product on $\mathbb{R}^2$,

$$\langle (u_1, u_2), (v_1, v_2)\rangle = u_1v_1 - u_1v_2 - u_2v_1 + 5u_2v_2$$

4. $V = \mathcal{M}_{n,n}(\mathbb{C})$, $\langle A, B\rangle = \operatorname{tr}(A(\overline{B})^t)$

5. $V = \mathcal{P}_n(\mathbb{R})$, $\langle p, q\rangle = \int_0^1 p(x)q(x)\,dx$

Examples 29.3.

1. The following is *not* an inner product on $\mathbb{R}^2$,

$$\langle (u_1,u_2),(v_1,v_2)\rangle = u_1v_1 - 2u_1v_2 - 2u_2v_1 + 3u_2v_2$$

2. The following is *not* an inner product on $\mathbb{C}^2$,

$$\langle (u_1,u_2),(v_1,v_2)\rangle = u_1v_1 + u_2v_2$$

29.2 Length, distance and orthogonality

Definition 29.4: Length and distance

For a vector space with an inner product $\langle \cdot\,, \cdot\rangle$ we define: the **length** (or **norm**) of a vector $v \in V$ by

$$\|v\| = \sqrt{\langle v, v\rangle}$$

The **distance** between two vectors $u, v \in V$ is defined to be

$$d(v,u) = \|v-u\|$$

A vector $u \in V$ with $\|u\| = 1$ is called a **unit vector**.

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v\rangle = 0$

Exercise 258. Let V be an inner product space. Show that

$$\forall u \in V\ \forall \alpha \in \mathbb{F}, \qquad \|\alpha u\| = |\alpha|\|u\|$$

Exercise 259. Let u, v be orthogonal vectors in an inner product space V. Show that

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

Example 29.5. $\langle (u_1,u_2),(v_1,v_2)\rangle = u_1v_1 + 2u_2v_2$ defines an inner product on $\mathbb{R}^2$

Letting $u = (3,1)$ and $v = (-2,3)$, we have

$$\begin{aligned}
\|v\|^2 &= \langle(-2,3),(-2,3)\rangle = (-2)^2 + 2\times 3^2 = 22\\
d(u,v) &= \|u-v\| = \|(5,-2)\| = \sqrt{\langle(5,-2),(5,-2)\rangle} = \sqrt{25+8} = \sqrt{33}\\
\langle u,v\rangle &= \langle(3,1),(-2,3)\rangle = 3\times(-2) + 2\times 1\times 3 = 0
\end{aligned}$$

The vectors u and v are orthogonal (with respect to this inner product).

Example 29.6. Consider the real vector space $V = C[0,2\pi]$ of all continuous functions $f : [0,2\pi] \to \mathbb{R}$. We equip V with the following inner product

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)dx$$

The norms of the functions $s, c \in V$ given by $s(x) = \sin(x)$ and $c(x) = \cos(x)$ are:

$$\|s\|^2 = \langle s,s\rangle = \int_0^{2\pi} \sin^2(x)dx = \int_0^{2\pi} \frac{1}{2}(1-\cos(2x))dx = \left[\frac{x}{2} - \frac{1}{4}\sin(2x)\right]_0^{2\pi} = \pi$$

So $\|s\| = \sqrt{\pi}$. Similarly, $\|c\| = \sqrt{\pi}$.

The vectors s and c are othogonal since

$$\langle s,c\rangle = \int_0^{2\pi} \sin(x)\cos(x)dx = \int_0^{2\pi} \frac{1}{2}\sin(2x)dx = \left[-\frac{1}{4}\cos(2x)\right]_0^{2\pi} = 0$$

29.3 The Cauchy-Schwartz inequality

We would like to define what should be meant by the angle between two vectors in an inner product space by using the same expression as in Definition 28.5. To be sure that it makes sense we need the following result.

Theorem 29.7: Cauchy-Schwarz inequality

Let V be an inner product space. Then for all $u, v \in V$

$$|\langle u,v\rangle| \leqslant \|u\|\|v\|$$

Further, equality holds if and only if one vector is a multiple of the other.

Proof. Let $u, v \in V$. If $v = \vec{0}$, the result holds because $\langle u, \vec{0}\rangle = 0$ and $\|\vec{0}\| = 0$ (see Exercise 257).

We now cinsider the case in which v is non-zero. Let $p = \frac{1}{\|v\|^2}\langle u,v\rangle v$. Note that

$$\|p\|^2 = \langle \frac{1}{\|v\|^2}\langle u,v\rangle v, \frac{1}{\|v\|^2}\langle u,v\rangle v\rangle = \frac{1}{\|v\|^4}|\langle u,v\rangle|^2\langle v,v\rangle = \frac{1}{\|v\|^2}|\langle u,v\rangle|^2$$

and that $w = u - p$ is orthogonal to p since

$$\langle p,w\rangle = \langle p, u-p\rangle = \langle p,u\rangle - \langle p,p\rangle = \frac{1}{\|v\|^2}\langle u,v\rangle\langle v,u\rangle - \frac{1}{\|v\|^2}|\langle u,v\rangle|^2 = 0$$

Then we have

$$\begin{aligned}\|u\|^2 &= \|w+p\|^2 \\ &= \|w\|^2 + \|p\|^2 && \text{(by Exercise 259)} \\ &\geqslant \|p\|^2 \\ &= \frac{1}{\|v\|^2}|\langle u,v\rangle|^2\end{aligned}$$

This inequality gives $\|u\|^2\|v\|^2 \geqslant |\langle u,v\rangle|^2$ and hence $\|u\|\|v\| \geqslant |\langle u,v\rangle|$.

The above inequality is an equality iff $\|w\|^2 = 0$, that is, $w = \vec{0}$. We have

$$w = \vec{0} \iff u = p \iff u \text{ is a multiple of } v$$

□

Example 29.8. Consider $V = \mathcal{P}_2(\mathbb{R})$ with inner product given by $\langle p,q\rangle = \int_0^1 p(x)q(x)\,dx$. Let $u = -x$ and $v = x^2$. Then

$$\langle u,v\rangle = -\int_0^1 x^3\,dx = -\frac{1}{4} \qquad \|u\|^2 = \langle u,u\rangle = \int_0^1 x^2\,dx = \frac{1}{3} \qquad \|v\|^2 = \langle v,v\rangle = \int_0^1 x^4\,dx = \frac{1}{5}$$

$$|\langle u,v\rangle| = \frac{1}{4} \leqslant \frac{1}{\sqrt{15}} = \|u\|\|v\|$$

Definition 29.9

Let V be a real inner product space. The **angle** between two vectors $u, v \in V$ is defined to be $\theta \in [0, \pi]$ given by

$$\theta = \arccos \frac{\langle u, v\rangle}{\|u\|\,\|v\|}$$

Example 29.10. With $u, v \in V$ as in Example 29.8 we have that the angle between x and x^2 is

$$\theta = \arccos\left(\frac{-1/4}{1/\sqrt{15}}\right) = \arccos\left(\frac{-\sqrt{15}}{4}\right)$$

Lemma 29.11: Triangle inequality for inner product spaces

Let V be an inner product space. Then $\forall u, v \in V$,

$$\|u + v\| \leqslant \|u\| + \|v\|$$

Proof.

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= \|u\|^2 + 2\Re(\langle u,v\rangle) + \|v\|^2 && (\Re(z) \text{ denotes the real part of } z \in \mathbb{C}) \\
&\leqslant \|u\|^2 + 2|\langle u,v\rangle| + \|v\|^2 \\
&\leqslant \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 && \text{(Cauchy-Schwartz)} \\
&= (\|u\| + \|v\|)^2
\end{aligned}$$

□

29.4 Exercises

260. Given $U = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$ are two 2×2 matrices, then

$$\langle U, V\rangle = u_1v_1 + u_2v_2 + u_3v_3 + u_4v_4$$

defines an inner product on $\mathcal{M}_{2,2}(\mathbb{R})$.

(a) Compute $\langle U, V\rangle$ when $U = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$ and $V = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$.

(b) Given $A = \begin{bmatrix} -2 & 5 \\ 3 & 6 \end{bmatrix}$, find $\|A\|$.

(c) Given $A = \begin{bmatrix} 2 & 6 \\ 9 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} -4 & 7 \\ 1 & 6 \end{bmatrix}$, find the distance between them $d(A, B)$.

(d) Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$. Which of the following matrices are orthogonal to A?

i) $\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$ ii) $\begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}$

261. Given $p = a_0 + a_1x + a_2x^2$ and $q = b_0 + b_1x + b_2x^2$ are two vectors in $\mathcal{P}_2(\mathbb{R})$, then

$$\langle p, q\rangle = a_0b_0 + a_1b_1 + a_2b_2$$

is an inner product on $\mathcal{P}_2(\mathbb{R})$.

(a) Compute $\langle p,q\rangle$ if $p=-2+x+3x^2$ and $q=4-7x^2$.

(b) If $p=-2+3x+2x^2$, find $\|p\|$.

(c) Given $p=3-x+x^2$, $q=2+5x^2$ find the distance between them $d(p,q)$.

(d) Show that $p=1-x+2x^2$ and $q=2x+x^2$ are orthogonal.

262. For $\mathbf{x}=(x_1,x_2).\mathbf{y}=(y_1,y_2)\in\mathbb{R}^2$, define $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+3x_2y_2$. Show that $\langle\mathbf{x},\mathbf{y}\rangle$ is an inner product on $\mathbb{R}^2$.

263. In $\mathbb{R}^2$, let $\langle(x_1,x_2),(y_1,y_2)\rangle=x_1y_1-x_2y_2$. Is this an inner product? If not, why not?

264. Verify that the operation

$$\langle(x_1,x_2),(y_1,y_2)\rangle=x_1y_1-x_1y_2-x_2y_1+3x_2y_2$$

is an inner product in $\mathbb{R}^2$.

265. Decide which of the suggested operations on $\mathbf{x}=(x_1,x_2,x_3)$ and $\mathbf{y}=(y_1,y_2,y_3)$ in $\mathbb{R}^3$ define an inner product:

(a) $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+2x_2y_2+x_3y_3$

(b) $\langle\mathbf{x},\mathbf{y}\rangle=x_1^2y_1^2+x_2^2y_2^2+x_3^2y_3^2$

(c) $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1-x_2y_2+x_3y_3$

(d) $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+x_2y_2$

266. Decide which of the following functions $\langle p,q\rangle$ on real polynomials $p(x)=a_0+a_1x+a_2x^2$ and $q(x)=b_0+b_1x+b_2x^2$ define inner products on $\mathcal{P}_2(\mathbb{R})$:

(a) $\langle p,q\rangle=a_0b_0+a_1b_1+a_2b_2$

(b) $\langle p,q\rangle=a_0b_0$

267. For the vectors $\mathbf{x}=(1,1,0)$, $\mathbf{y}=(0,1,0)$ in $\mathbb{R}^3$ compute the norms $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ using the following inner products.

(a) $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+x_2y_2+x_3y_3$

(b) $\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+3x_2y_2+x_3y_3$

268. In each part determine whether the given vectors are orthogonal with respect to the Euclidean inner product (i.e., the usual dot product).

(a) $u=(-1,3,2)$, $v=(4,2,-1)$

(b) $u=(0,3,-2,1)$, $v=(5,2,-1,0)$

269. Endow $\mathbb{R}^4$ have the Euclidean inner product (i.e. the dot product), and let $\mathbf{u}=(-1,1,0,2)$. Determine whether the vector u is orthogonal to the following vectors:

(a) $w_1=(0,0,0,0)$

(b) $w_2=(1,-1,3,0)$

(c) $w_3=(4,0,9,2)$

270. Let $u=(1+i,3i)$ and $v=(4,2-i)$. Use the complex dot product on $\mathbb{C}^2$ to compute:

(a) $u\cdot v$

(b) $v\cdot u$

(c) $\|u\|$

(d) $\|v\|$

271. Let $\mathbb{C}^3$ have the complex dot product. If $u=(2i,i,3i)$ and $v=(i,6i,k)$, for what values of $k\in\mathbb{C}$ are u and v orthogonal?

272. Show that in every real inner product space: $v+w$ is orthogonal to $v-w$ if and only if $\|v\|=\|w\|$.

273. Prove that the following holds for all vectors $\mathbf{x},\mathbf{y}$ in a *real* inner product space:

$$\|\mathbf{x}+\mathbf{y}\|^2+\|\mathbf{x}-\mathbf{y}\|^2=2\|\mathbf{x}\|^2+2\|\mathbf{y}\|^2$$

274. Let A be a real invertible $n \times n$ matrix. Show that

$$\langle \mathbf{x}, \mathbf{y} \rangle \;\equiv\; [\mathbf{x}]^T A^T A[\mathbf{y}] = (A[\mathbf{x}])^T A[\mathbf{y}]$$

defines an inner product in $\mathbb{R}^n$, where $[\mathbf{x}]$ and $[\mathbf{y}]$ are coordinate matrices with respect to the standard basis.

Show that it fails to be an inner product if A is not invertible.
(Hint: If A is not invertible, then its kernel is non-trivial.)

275. Verify that the Cauchy-Schwartz inequality holds for the given vectors using the Euclidean inner product.

(a) $u = (-3, 1, 0), v = (2, -1, 3)$

(b) $u = (-4, 2, 1), v = (8, -4, -2)$

276. Use the Cauchy-Schwartz inequality (applied to the Euclidean inner product on $\mathbb{R}^n$) to show that given any $a_1, a_2, \ldots, a_n \in \mathbb{R}$ we have that

$$\frac{a_1 + \cdots + a_n}{n} \leqslant \sqrt{\frac{a_1^2 + \cdots + a_n^2}{n}}$$

277. Consider $\mathbb{R}^2$ and $\mathbb{R}^3$, each with the Euclidean inner product. In each part find the cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$.

(a) $u = (1, -3), v = (2, 4)$

(b) $u = (-1, 5, 2), v = (2, 4, -9)$

278. For the vectors $\mathbf{x} = (1, 1, 0),\ \mathbf{y} = (0, 1, 0)$ in $\mathbb{R}^3$ compute the angle between $\mathbf{x}$ and $\mathbf{y}$ using the following inner products.

(a) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 + x_3y_3$

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 3x_2y_2 + x_3y_3$

Extra material for lecture 29

▷ References about inner products

Elementary Linear Algebra by Anton and Rorres, §6, p345
Linear Algebra Done Right by Axler, §6.A, p164

LECTURE **30**

Orthonormal bases

Some bases for an inner product space fit nicely with the inner product. Before defining the notion of an orthonormal basis we note that, after choosing a basis, inner products (on finite-dimensional vector spaces) can be represented by matrices. If the basis is orthonormal, then the corresponding matrix is particularly simple.

In this lecture, as in the previous, the field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

30.1 Matrix representation of an inner product

Let V be an n-dimensional vector space and let $\mathcal{B}$ be a basis for V. Fix a matrix $M \in \mathcal{M}_{n,n}(\mathbb{F})$. For $u, v \in V$ define

$$\langle u, v\rangle = [u]^T\, M\, \overline{[v]} \tag{†}$$

Where, for legibility, we have written $[u]$ in place of $[u]_{\mathcal{B}}$ and $[v]$ in place of $[v]_{\mathcal{B}}$. The right hand side is a 1×1 matrix which we identify with an element of $\mathbb{F}$.

What conditions of M ensure that this gives an inner product on V?

Exercise 279. Show that the function $V \times V \to \mathbb{F}$ defined by (†) satisfies axioms 2 and 3 in the definition of an inner product.

We need to add a condition on M in order that axiom 1 be satisfied.

$$\begin{aligned}
\overline{\langle v, u\rangle} &= \overline{[v]^T M \overline{[u]}} = \overline{[v]}^T\, \overline{M}\, [u] \\
&= (\overline{[v]}^T\, \overline{M}\, [u])^T && \text{(since a 1×1 matrix is equal to its own transpose)} \\
&= [u]^T M^* \overline{[v]} && \text{(where M^* denotes $\overline{M}^T$)}
\end{aligned}$$

Therefore axiom 1 is satisfied if and only if $M^* = M$.

We fix some terminology.

Definition 30.1

A matrix $M \in \mathcal{M}_{n,n}(\mathbb{R})$ is called a **real symmetric matrix** if $M^T = M$.

A matrix $M \in \mathcal{M}_{n,n}(\mathbb{C})$ is called a **Hermitian matrix** if $M^* = M$.

A matrix $M \in \mathcal{M}_{n,n}(\mathbb{F})$ is called **positive definite** if $M^* = M$ and the following condition is satisfied

$$\forall X \in \mathcal{M}_{n,1}(\mathbb{F}) \setminus \{\vec{0}\}, \qquad X^T M \overline{X} > 0$$

Exercise 280. Show that if M is positive definite, then the function $V \times V \to \mathbb{F}$ defined by (†) is an inner product on V. (Hint: The only thing left to show is that axiom 4 in the definition of inner product is satisfied.)

Given an inner product on V, there always exists a matrix $M \in \mathcal{M}_{n,n}(\mathbb{F})$ such that the inner product is given by the expression (†).

Proposition 30.2: matrix representation of an inner product

Let V be a finite-dimensional inner product space and $\mathcal{B}$ a basis for V. There exists a matrix $M \in \mathcal{M}_{n,n}(\mathbb{F})$ such that

$$\forall u, v \in V, \qquad \langle u, v\rangle = [u]_B^T \, M \, \overline{[v]}_{\mathcal{B}}$$

Proof. Suppose $\mathcal{B} = \{b_1, \dots, b_n\}$ and that we have an inner product $\langle ., .\rangle$ on V. Define $M \in \mathcal{M}_{n,n}(\mathbb{F})$ to be the matrix whose (i, j)-th entry is given by $M_{ij} = \langle b_i, b_j\rangle$. That $\langle u, v\rangle = [u]^T \, M \, \overline{[v]}$ for all $u, v \in V$ can be readily verified.

Exercise 281. Show that, with M defined as above, we have $\langle u, v\rangle = [u]^T \, M \, \overline{[v]}$ for all $u, v \in V$.

□

Example 30.3. We saw in a previous example that

$$\langle (u_1, u_2), (v_1, v_2))\rangle = u_1v_1 - u_1v_2 - u_2v_1 + 5u_2v_2$$

defines an inner product on $\mathbb{R}^2$. What is the matrix representation of this inner product (with respect to the standard basis)? Noting that $\langle (1, 0), (1, 0)\rangle = 1$, $\langle (1, 0), (0, 1)\rangle = -1$, $\langle (0, 1), (1, 0)\rangle = -1$ and $\langle (0, 1), (0, 1)\rangle = 5$ we have that

$$\langle (u_1, u_2), (v_1, v_2)\rangle = [u_1u_2] \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

Exercise 282. Let $M, N \in \mathcal{M}_{n,n}(\mathbb{F})$. Show that if

$$\forall X, Y \in \mathcal{M}_{n.1}(\mathbb{F}), \qquad X^T M \overline{Y} = X^T N \overline{Y}$$

then $M = N$.

(It follows that the matrix representation (with respect to a fixed basis) of an inner product is unique.)

30.2 Orthogonal sets of vectors

Definition 30.4: Orthogonal set of vectors

Let V be an inner product space. A set of vectors $\{v_1, \dots, v_k\} \subseteq V$ is called **orthogonal** if $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$.

Examples 30.5.

1. $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is orthogonal in $\mathbb{R}^3$ with the dot product as inner product.
2. So is $\{(1, 1, 1), (1, -1, 0), (1, 1, -2)\} \subseteq \mathbb{R}^3$ using the dot product.
3. Consider the inner product space of Example 29.6: $V = C[0, 2\pi]$ and $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)dx$. The set $\{1, \sin(x), \cos(x)\} \subseteq V$ is orthogonal.

Lemma 30.6

Let V be an inner product space and $S \subseteq V \setminus \{\vec{0}\}$. If S is orthogonal, then S is linearly independent.

Proof. Suppose that $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ and $v_1, \dots, v_k \in S$ are such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = \vec{0}$. Then for $i \in \{1, \dots, k\}$ we have

$$\begin{aligned}
\alpha_1 v_1 + \cdots + \alpha_k v_k = \vec{0} &\implies \langle \alpha_1 v_1 + \cdots + \alpha_k v_k, v_i \rangle = 0 \\
&\implies \alpha_1 \langle v_1, v_i \rangle + \cdots + \alpha_k \langle v_k, v_i \rangle = 0 \\
&\implies \alpha_i \langle v_i, v_i \rangle = 0 && \text{(since } \langle v_j, v_i \rangle = 0 \text{ if } j \neq i\text{)} \\
&\implies \alpha_i = 0 && \text{(since } v_i \neq \vec{0}\text{)}
\end{aligned}$$

Having shown that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = \vec{0} \implies \forall i,\ \alpha_i = 0$$

we conclude that the set S is linearly independent. □

Example 30.7. 1. The set $\{(1,1,1), (1,-1,0), (1,1,-2)\} \subseteq \mathbb{R}^3$ is linearly independent.

2. The set $\{1, \sin(x), \cos(x)\} \subseteq C[0, 2\pi]$ is linearly independent.

30.3 Orthonormal bases

Definition 30.8

A set of vectors $\{v_1, \dots, v_k\}$ is called **orthonormal** if it is orthogonal and each vector has length one. That is,

$$\{v_1, \dots, v_k\} \quad \text{is orthonormal if} \quad \langle v_i, v_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Remark. Any orthogonal set of non-zero vectors can be made orthonormal by multiplying each vector v by $\frac{1}{\|v\|}$.

Examples 30.9.

1. In $\mathbb{R}^3$ with the dot product:
   (a) $\{(1,0,0), (0,1,0), (0,0,1)\}$ is orthonormal
   (b) $\{(1,1,1), (1,-1,0), (1,1,-2)\}$ is orthogonal but is not orthonormal
   (c) $\{\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2)\}$ is orthonormal

2. In $\mathcal{P}_2(\mathbb{R})$ with the inner product $\langle p, q \rangle = \int_0^1 p(x)q(x)\,dx$:
   (a) The set $\{1, x, x^2\}$ is not orthogonal.
   (b) The set $\{1, 2x-1, 6x^2-6x+1\}$ is orthogonal but not orthonormal.
   (c) The set $\{1, \sqrt{3}(2x-1)\}, \sqrt{5}(6x^2-6x+1)\}$ is orthonormal.

3. In $C[0, 2\pi]$ with the inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx$$

   (a) The set $\{1, \sin(x), \cos(x)\}$ is orthogonal but not orthonormal.
   (b) The set $\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x)\}$ is orthonormal
   (c) The (infinite) set

$$\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{\pi}}\sin(2x), \frac{1}{\sqrt{\pi}}\cos(2x), \dots\}$$

   is orthonormal.

Definition 30.10

Let V be an inner product space. An **orthonormal basis** for V is a basis that is an orthonormal set.

Examples 30.11.

1. In $\mathbb{R}^3$ with the dot product:
   (a) $\{(1,0,0),(0,1,0),(0,0,1)\}$ is an orthonormal basis
   (b) $\{\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2)\}$ is an orthonormal basis
2. For $V = \mathcal{P}_2(\mathbb{R})$ with the inner product $\langle p, q\rangle = \int_0^1 p(x)q(x)\,dx$, The set $\{1, \sqrt{3}(2x-1)\}, \sqrt{5}(6x^2 - 6x+1)\}$ is an orthonormal basis.

Bases that are orthonormal are particularly convenient to work with. For example, we have the following.

Lemma 30.12

Let V be an inner product space and let $\mathcal{B} = \{b_1, \dots, b_n\}$ be an orthonormal basis for V. Then for all $v \in V$ we have that

$$v = \langle v, b_1\rangle b_1 + \cdots + \langle v, b_n\rangle b_n$$

Proof. Let $v = \alpha_1 b_1 + \cdots + \alpha_n b_n$. We need to show that $\alpha_i = \langle v, b_i\rangle$. We have

$$\langle v, b_i\rangle = \langle\sum_{j=1}^{n} \alpha_j b_j, b_i\rangle = \sum_{j=1}^{n} \alpha_j \langle b_j, b_i\rangle = \alpha_i$$

□

Example 30.13. Let $V = \mathbb{R}^3$ equipped with the dot product and let

$$\mathcal{B} = \{b_1 = \frac{1}{\sqrt{3}}(1,1,1), b_2 = \frac{1}{\sqrt{2}}(1,-1,0), b_3 = \frac{1}{\sqrt{6}}(1,1,-2)\}$$

We saw above that this is an orthonormal basis. To find coordinates with respect to $\mathcal{B}$, we can just use the inner product. For example

$$\begin{aligned}(1,2,3) &= ((1,2,3)\cdot b_1)b_1 + ((1,2,3)\cdot b_2)b_2 + ((1,2,3)\cdot b_3)b_3 && \text{(Lemma 30.12)}\\ &= \frac{1}{\sqrt{3}}(1,2,3)\cdot(1,1,1)\,b_1 + \frac{1}{\sqrt{2}}(1,2,3)\cdot(1,-1,0)\,b_2 + \frac{1}{\sqrt{6}}(1,2,3)\cdot(1,1,-2)\,b_3\\ &= 2\sqrt{3}\,b_1 - \frac{1}{\sqrt{2}}b_2 - \frac{\sqrt{3}}{\sqrt{2}}b_3\end{aligned}$$

That is, $[(1,2,3)]_{\mathcal{B}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2\sqrt{6} \\ -1 \\ -\sqrt{3}\end{bmatrix}$

Example 30.14. Let $W = \{(x,y,z) \in \mathbb{R}^3 \mid x+y+z = 0\} \leqslant \mathbb{R}^3$ equipped with the dot product. The set

$$\mathcal{B} = \{\frac{1}{\sqrt{2}}(-1,1,0), \frac{1}{\sqrt{6}}(-1,-1,2)\}$$

is an orthonormal basis for W. For $(-1,0,1) \in W$ we have

$$(-1,0,1)\cdot b_1 = \frac{1}{\sqrt{2}} \qquad (-1,0,1)\cdot b_2 = \frac{\sqrt{3}}{\sqrt{2}}$$

$$(-1,0,1) = \frac{1}{\sqrt{2}}\, b_1 + \frac{\sqrt{3}}{\sqrt{2}}\, b_2 = \frac{1}{2}(-1,1,0) + \frac{1}{2}(-1,-1,2)$$

30.4 Exercises

283. Use Lemma 30.12 to express the given vector as a linear combination of the vectors in the following orthonormal basis (with respect to the dot product) for $\mathbb{R}^3$.

$$\mathcal{B} = \left\{ \left(\tfrac{1}{3}, -\tfrac{2}{3}, \tfrac{2}{3}\right), \left(-\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{2}{3}\right), \left(\tfrac{2}{3}, \tfrac{2}{3}, \tfrac{1}{3}\right) \right\}$$

(a) $(1,2,3)$ (b) $(-1,0,1)$

284. Let V be the be the vector space $C[0,2\pi]$ of real-valued continuous functions on the interval $[0,2\pi]$ equipped with the inner product $\langle f,g\rangle = \int_0^{2\pi} f(x)g(x)\,dx$. Show that the set

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x) \right\} \subseteq V$$

is an orthonormal set

285. Let $\langle x,y\rangle$ be an inner product on a vector space V, and let $\mathcal{B} = \{e_1, e_2, \dots, e_n\}$ be an orthonormal basis for V. Prove that:

(a) $\langle \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n, \beta_1 e_1 + \beta_2 e_2 + \cdots + \beta_n e_n\rangle = \alpha_1\overline{\beta}_1 + \alpha_2\overline{\beta}_2 + \cdots + \alpha_n\overline{\beta}_n$

(b) $\langle x,y\rangle = \langle x,e_1\rangle\overline{\langle y,e_1\rangle} + \cdots + \langle x,e_n\rangle\overline{\langle y,e_n\rangle}$

(c) The matrix representation, with respect to $\mathcal{B}$, of the inner product is I_n.

286. Let V be a finite dimensional inner product space and let $\mathcal{B} = \{b_1, \dots, b_n\}$ be an orthonormal basis for V. Let $T : V \to V$ be a linear transformation and let $A = [T]_{\mathcal{B}}$ be the matrix of T with respect to $\mathcal{B}$. Prove that $A_{ij} = \langle T(b_j), b_i\rangle$

Extra material for lecture 30

- ▷ Let V be the inner product space of Exercise 284. Show that the infinite set

  $$\left\{ \tfrac{1}{\sqrt{2\pi}}, \tfrac{1}{\sqrt{\pi}}\sin(x), \tfrac{1}{\sqrt{\pi}}\cos(x), \tfrac{1}{\sqrt{\pi}}\sin(2x), \tfrac{1}{\sqrt{\pi}}\cos(2x), \tfrac{1}{\sqrt{\pi}}\sin(3x), \tfrac{1}{\sqrt{\pi}}\cos(3x), \dots \right\} \subseteq V$$

  is an orthonormal set. The set is therefore linearly independent.

- ▷ A **Hilbert space** is an inner product space with the property that the associated metric is *complete*. That the metric is complete is to say that all Cauchy sequences converge.

  An example of an infinite-dimensional Hilbert space is the space of "square summable" sequences of complex numbers

  $$\ell^2 = \{ \ (z_1, z_2, z_3, \dots) \mid z_i \in \mathbb{C},\ \sum_{i=1}^{\infty} |z_i|^2 < \infty \}$$

  with inner product

  $$\langle (x_1, \dots), (y_1, \dots) \rangle = \sum_{i=1}^{\infty} x_i \, \overline{y_i}$$

The Gram-Schmidt orthogonalisation procedure and orthogonal projection

We have seen how to find a basis for a vector space, but what about finding an *orthonormal* basis? In this lecture we discuss a technique, based on the idea contained in Exercise 287 below, for converting a basis of a finite-dimensional inner product space into an orthonormal basis.

Exercise 287. Let V be an inner product space and let $v \in V$.

a) Let $u \in V$ be a unit vector. Show that $v - \langle v, u\rangle u$ is orthogonal to u.

b) Let $\{u_1, \ldots, u_k\} \subseteq V$ be an orthonormal set. Show that

$$v - \langle v, u_1\rangle u_1 - \langle v, u_2\rangle u_2 - \cdots - \langle v, u_k\rangle u_k$$

is orthogonal to every element of $\operatorname{span}\{u_1, \ldots, u_k\}$.

31.1 Gram-Schmidt orthogonalisation procedure

Let V be an inner product space and let $\{u_1, \ldots, u_k\} \subseteq V$ be an orthonormal set. Suppose that $v \in V$ is such that $v \notin W = \operatorname{span}\{u_1, \ldots, u_k\}$. Defining $w = v - \langle v, u_1\rangle u_1 - \langle v, u_2\rangle u_2 - \cdots - \langle v, u_k\rangle u_k$ we have that $w \neq \vec{0}$ and w is orthogonal to W. Therefore, if we define $u_{k+1} = w/\|w\|$ and add it to the above orthonormal set, the now larger set $\{u_1, \ldots, u_{k+1}\} \subseteq V$ is still orthonormal. Applying this observation repeatedly allows us to construct an orthonormal basis for V.

Algorithm 31.1: Gram-Schmidt procedure

Let V be a finite-dimensional inner product space and let $\{b_1, \ldots, b_n\}$ be a basis for V.

We define $u_1, w_2, u_2, \ldots, w_n, u_n \in V$ as follows:

1) $u_1 = \frac{1}{\|b_1\|} b_1$

2) $w_2 = b_2 - \langle b_2, u_1\rangle u_1$ and $u_2 = \frac{1}{\|w_2\|} w_2$

3) $w_3 = b_3 - \langle b_3, u_1\rangle u_1 - \langle b_3, u_2\rangle u_2$ and $u_3 = \frac{1}{\|w_3\|} w_3$

$\vdots$ $\vdots$

n) $w_n = b_n - \langle b_n, u_1\rangle u_1 - \cdots - \langle b_n, u_{k-1}\rangle u_{k-1}$ and $u_n = \frac{1}{\|w_n\|} w_n$

Then $\{u_1, \ldots, u_n\}$ is an orthonormal basis for V.

Let's note explicitly the following consequence.

Theorem 31.2

Every finite-dimensional inner product space has an orthonormal basis. □

Example 31.3. We find an orthonormal basis for the subspace W of $\mathbb{R}^4$ (with the dot product) spanned by $\{(1,1,1,1),(2,4,2,4),(1,5,-1,3)\}$.

Following the Gram-Schmidt procedure gives:

$$\begin{aligned}
u_1 &= (1,1,1,1)/(\|(1,1,1,1)\|) = \frac{1}{2}(1,1,1,1)\\
w_2 &= (2,4,2,4) - \langle (2,4,2,4), u_1\rangle u_1 = (2,4,2,4) - \frac{1}{4}\langle (2,4,2,4),(1,1,1,1)\rangle (1,1,1,1)\\
&= (2,4,2,4) - 3(1,1,1,1) = (-1,1,-1,1)\\
u_2 &= (-1,1,-1,1)/\|(-1,1,-1,1)\| = \frac{1}{2}(-1,1,-1,1)\\
w_3 &= (1,5,-1,3) - \langle (1,5,-1,3), u_1\rangle u_1 - \langle (1,5,-1,3), u_2\rangle u_2\\
&= (1,5,-1,3) - 2(1,1,1,1) - 2(-1,1,-1,1) = (1,1,-1,-1)\\
u_3 &= (1,1,-1,-1)/\|(1,1,-1,-1)\| = \frac{1}{2}(1,1,-1,-1)
\end{aligned}$$

So $\{\frac{1}{2}(1,1,1,1), \frac{1}{2}(-1,1,-1,1), \frac{1}{2}(1,1,-1,-1)\}$ is an orthonormal basis for W.

31.2 Orthogonal projection

Definition 31.4: Orthogonal complement

Let V be an inner product space and let $W \leqslant V$ be a subspace of V. The **orthogonal complement** of W is defined to be

$$W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in W\}$$

Exercise 288. Let V be an inner product space and let $W \leqslant V$ be a subspace of V. Prove the following properties of the orthogonal complement. Note that we are not assuming that V (or W) is finite dimensional.

a) $W^\perp$ is a subspace of V

b) $W \cap W^\perp = \{\vec{0}\}$

c) $W \subseteq (W^\perp)^\perp$

Lemma 31.5

Let V be a finite-dimensional inner product space and let $W \leqslant V$ be a subspace. Every vector $v \in V$ can be written in a unique way as $v = w + w'$ where $w \in W$ and $w' \in W^\perp$.

Proof. Let $v \in V$. We first need to show that there exist $w \in W$ and $w' \in W^\perp$ such that $v = w + w'$. Let $\{w_1, \dots, w_k\}$ be an orthonormal basis for W. Define

$$w = \langle v, w_1\rangle w_1 + \cdots + \langle v, w_k\rangle w_k$$

Then clearly $w \in W$. Now define $w' = v - w$. We have that $v = w + w'$. We have to show that $w' \in W^\perp$. For $i \in \{1, \dots, k\}$ we have

$$\begin{aligned}
\langle w, w_i\rangle &= \langle \sum_{j=1}^{k} \langle v, w_j\rangle w_j, w_i\rangle && \text{(definition of } w\text{)}\\
&= \sum_{j=1}^{k} \langle v, w_j\rangle \langle w_j, w_i\rangle && \text{(linearity)}\\
&= \langle v, w_i\rangle && (\{w_1, \dots, w_k\} \text{ is an orthonormal set})
\end{aligned}$$

Therefore $\langle w - v, w_i \rangle = 0$ (for all i) and it follows that $w' \in W^\perp$.

It remains to show that w and w' are unique. Suppose that $u \in W$ and $u' \in W^\perp$ are such that $v = u + u'$. Then we have

$$\begin{aligned} & w + w' = u + u' \\ \implies & w - u = u' - w' \in W \cap W^\perp \\ \implies & w - u = u' - w' = \vec{0} \qquad \text{(Exercise 288)} \end{aligned}$$

□

Definition 31.6: Orthogonal projection

Let V be a finite-dimensional inner product space and $W \leqslant V$ a subspace. The **orthogonal projection** of V onto W is the map

$$\operatorname{proj}_W : V \to V$$

defined as follows.

Given $v \in V$ we have $v = w + w'$ for some (unique) $w \in W$ and $w' \in W^\perp$. Define $\operatorname{proj}_W(v) = w$.

The element $\operatorname{proj}_W(v)$ is sometimes called the **projection of v to W**.

Exercise 289. Show that proj_W is a linear transformation and that $(\operatorname{proj}_W)^2 = \operatorname{proj}_W$.

Exercise 290. Let V be an finite-dimensional inner product space and let $W \leqslant V$ be a subspace of V.

a) Use the rank-nullity theorem to show that $\dim(V) = \dim(W) + \dim(W^\perp)$.

b) Show that $(W^\perp)^\perp = W$.

From the proof of Lemma 31.5 we have the following.

Lemma 31.7

Let V be a finite-dimensional inner product space and $W \leqslant V$ a subspace. Let $\{w_1, \dots, w_k\}$ be an orthonormal basis for W. Then for any $v \in V$

$$\operatorname{proj}_W(v) = \langle v, w_1 \rangle w_1 + \cdots + \langle v, w_k \rangle w_k$$

Example 31.8. Consider $\mathcal{P}_2(\mathbb{R})$ with the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

The orthogonal projection of $v = 1 + 2x + 3x^2$ onto the unit vector $u = \sqrt{3}\,x$ (i.e., the projection onto the subspace $W = \operatorname{span}\{u\}$) is

$$\operatorname{proj}_W(v) = \langle v, u \rangle u = (3\int_0^1 x + 2x^2 + 3x^3 dx)x = \frac{23}{4}x$$

Example 31.9. Let $W = \{(x, y, z) \mid x + y + z = 0\}$ in $V = \mathbb{R}^3$ equipped with the dot product.
The set $\{b_1 = \frac{1}{\sqrt{2}}(1, -1, 0), b_2 = \frac{1}{\sqrt{6}}(1, 1, -2)\}$ is an orthonormal basis for W. For $v = (1, 2, 3)$ we have

$$\operatorname{proj}_W(v) = \langle v, b_1 \rangle b_1 + \langle v, b_2 \rangle b_2 = \left(\frac{-1}{2}\right)(1, -1, 0) + \left(\frac{-1}{2}\right)(1, 1, -2) = (-1, 0, 1)$$

Note that $v - \operatorname{proj}_W(v) = (2, 2, 2)$ is orthogonal to W.

Lemma 31.10

Let V be an inner product space and $W \leqslant V$ a subspace. Then, for all $v \in V$ $\mathrm{proj}_W(v)$ is the vector in W that is closest to v. (That is, $\forall v \in V\ \forall\, w \in W$, $\|v - \mathrm{proj}_W(v)\| \leqslant \|v - w\|$)

Proof. Let $p = \mathrm{proj}_W(v)$ and $w \in W$

$$\begin{aligned}\|v-w\|^2 &= \langle v-w, v-w\rangle = \langle (v-p)+(p-w), (v-p)+(p-w)\rangle \\ &= \langle v-p, v-p\rangle + \langle v-p, p-w\rangle + \langle p-w, v-p\rangle + \langle p-w, p-w\rangle \\ &= \|v-p\|^2 + \|p-w\|^2 \\ &\geqslant \|v-p\|^2\end{aligned}$$

□

Example 31.11. Let $W \leqslant \mathbb{R}^4$ be as in Example 31.3, that is, $W = \mathrm{span}\{(1,1,1,1), (2,4,2,4), (1,5,-1,3)\}$. We saw that $\{\frac{1}{2}(1,1,1,1), \frac{1}{2}(-1,1,-1,1), \frac{1}{2}(1,1,-1,-1)\}$ is an orthonormal basis for W.

Using the orthonormal basis we find the point $p \in W$ that is closest to $v = (2,2,1,3)$. From Lemma 31.10 we know that $p = \mathrm{proj}_W(v)$.

$$\begin{aligned}p = \mathrm{proj}_W(v) &= \langle v, u_1\rangle u_1 + \langle v, u_2\rangle u_2 + \langle v, u_3\rangle u_3 = 2(1,1,1,1) + \frac{1}{2}(-1,1,-1,1) + 0(1,1,-1,-1) \\ &= \frac{1}{2}(3,5,3,5)\end{aligned}$$

31.3 Exercises

291. Use the Gram-Schmidt procedure to construct orthonormal bases for the subspaces of $\mathbb{R}^n$ spanned by the following sets of vectors (using the dot product):

    (a) $(1,0,1,0)$, $(2,1,1,1)$, $(1,-1,1,-1)$

    (b) $(2,2,-1,0)$, $(2,3,1,-2)$, $(3,4,5,-2)$

    (c) $(1,-2,1,3,-1)$, $(0,6,-2,-6,0)$, $(4,-2,2,6,-4)$

292. Let $\mathcal{P}_2(\mathbb{R})$ be the vector space of polynomials of degree at most two with the inner product

    $$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

    Apply Gram-Schmidt to the basis $\{1, x, x^2\}$ to obtain an orthonormal basis.

293. Find the orthogonal projection of (x,y,z) onto the subspace of $\mathbb{R}^3$ spanned by the vectors

    (a) $\{(1,\ 2,\ 2), (-2,\ 2,\ -1)\}$

    (b) $\{(1,\ 2,\ -1), (0,\ -1,\ 2)\}$

294. Consider $\mathbb{R}^3$ equipped with the dot product. Let $w = (2,-1,-2)$ and $v = (2,1,3)$. Find vectors v_1 and v_2 such that $v = v_1 + v_2$, v_1 is parallel to w, and v_2 is perpendicular to w.

295. Find the standard matrices of the transformations $T : \mathbb{R}^3 \to \mathbb{R}^3$ which orthogonally project a point (x,y,z) onto the following subspaces of $\mathbb{R}^3$. Use the matrix to show the transformation is idempotent (i.e., $T \circ T = T$).

    (a) The z-axis.

    (b) The straight line $x = y = 2z$.

    (c) The plane $x + y + z = 0$.

296. Let Π be the plane in $\mathbb{R}^3$ given by $x+y+z=0$. Use orthogonal projection to find the point on Π that is as close as possible to $(4,5,0)$.

297. Let V be a finite-dimensional inner product space and let $W \leqslant V$ be a subspace. Show that $\dim W + \dim W^{\perp} = \dim V$.

298. Let V be a finite-dimensional inner product space and let $W \leqslant V$ be a subspace. Let $P : V \to V$ be projection onto W. Show that

$$\forall\, u, v \in V, \quad \langle P(u), v\rangle = \langle u, P(v)\rangle$$

Extra material for lecture 31

- ▷ What happens if we apply Gram-Schmidt to a linearly dependent set? It can be used to produce an orthonormal basis for the span of the original set of vectors.
- ▷ Not all infinite-dimensional inner product spaces have an orthonormal basis.
- ▷ Reflection across a subspace

  Let V be a finite-dimensional inner product space and $W \leqslant V$ a subspace. **Reflection** across W is the linear transformation $R : V \to V$ defined as follows. Given $v \in V$ we have $v = w + w'$ for some (unique) $w \in W$ and $w' \in W^{\perp}$. Define

  $$R(w + w') = w - w'$$

  Notice that $R(v) = v - 2\,\mathrm{proj}_{W^{\perp}}(v)$

Orthogonal diagonalisation

In this lecture we look at matrix representations with respect to orthonormal bases. Suppose that $\mathcal{B} = \{u_1, \ldots, u_n\}$ is an orthonormal basis for $\mathbb{R}^n$ equipped with the standard inner product (dot product). Then the transition matrix $P = P_{\mathcal{S},\mathcal{B}} = [\, [u_1]_{\mathcal{S}} \cdots [u_n]_{\mathcal{S}} \,]$ has the property that $P^T P = I_n$.

32.1 Orthogonal matrices

Definition 32.1

A matrix $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ is called **orthogonal** if $Q^T Q = I_n$.

Examples 32.2.

1) The following are all orthogonal:

$$\frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ -2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \in \mathcal{M}_{3.3}(\mathbb{R}) \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$$

2) The following are not orthogonal:

$$\begin{bmatrix} 1 & 2 & -2 \\ -2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \in \mathcal{M}_{3.3}(\mathbb{R}) \qquad \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \frac{1}{5}\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{R})$$

Exercise 299. Suppose that $Q, P \in \mathcal{M}_{n,n}(\mathbb{R})$ are orthogonal matrices. Show that

(a) Q^T is orthogonal

(b) Q^{-1} is orthogonal

(c) PQ is orthogonal

(d) $\det(Q) = \pm 1$

Lemma 32.3: Conditions equivalent to orthogonality

Let $Q \in \mathcal{M}_{n,n}(\mathbb{R})$. The following are equivalent

1. Q is orthogonal
2. the columns of Q form an orthonormal basis of $\mathcal{M}_{n,1}(\mathbb{R})$ (with respect to the dot product)
3. the rows of Q form an orthonormal basis of $\mathcal{M}_{n,1}(\mathbb{R})$ (with respect to the dot product)
4. $\|Qu\| = \|u\|$ for all $u \in \mathcal{M}_{n,1}(\mathbb{R})$
5. $\langle Qu, Qv\rangle = \langle u, v\rangle$ for all $u, v \in \mathcal{M}_{n,1}(\mathbb{R})$

Proof. That $1 \Leftrightarrow 2$ follows from the way in which matrix multiplication is defined. That $1 \Leftrightarrow 3$ then follows from the fact that the transpose of an orthogonal matrix is orthogonal.

For $1 \Rightarrow 4$ we have:

$$\begin{aligned}
\|Qu\|^2 = \langle Qu, Qu\rangle &= (Qu)^T(Qu) && \text{(coordinate matrices with respect to the standard basis)}\\
&= u^T Q^T Q u \\
&= u^T I_n u && \text{(since } Q^TQ = I_n\text{)}\\
&= u^T u = \|u\|^2
\end{aligned}$$

For $4 \Rightarrow 5$ we have:

$$\begin{aligned}
\langle Qu, Qv\rangle &= \frac{1}{4}\left(\langle Qu+Qv, Qu+Qv\rangle - \langle Qu-Qv, Qu-Qv\rangle\right)\\
&= \frac{1}{4}\left(\langle Q(u+v), Q(u+v)\rangle - \langle Q(u-v), Q(u-v)\rangle\right)\\
&= \frac{1}{4}\left(\langle u+v, u+v\rangle - \langle u-v, u-v\rangle\right) && \text{(since 4 holds)}\\
&= \langle u, v\rangle
\end{aligned}$$

It only remains to show that $5 \Rightarrow 1$.

$$\langle Qu, Qv\rangle = \langle u, v\rangle \implies u^TQ^TQv = u^Tv \implies u^T(Q^TQ - I_n)v = 0$$

Since this holds for all $u, v \in \mathcal{M}_{n,1}(\mathbb{R})$ we must have that $Q^TQ - I_n = \vec{0}$. □

32.2 Orthogonal diagonalisation

Definition 32.4

A matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$ is said to be **orthogonally diagonalisable** if there is an orthogonal matrix $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ and a diagonal matrix $D \in \mathcal{M}_{n,n}(\mathbb{R})$ such that $A = Q\,D\,Q^T$.

Example 32.5. The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ is orthogonally diagonalisable since

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^T = QDQ^T$$

(We'll see soon how Q can be found.)

We have the following version of Theorem 25.4.

Theorem 32.6

A matrix $A \in \mathcal{M}_{n,n}(\mathbb{R})$ is orthogonally diagonalisable if and only if there is an orthonormal[*] basis $\mathcal{B}$ for $\mathcal{M}_{n,1}(\mathbb{R})$ with the property that all elements of $\mathcal{B}$ are eigenvectors of A.

Sketch of proof. The proof is the same as for Theorem 25.4, but with the extra observation that the basis $\mathcal{B}$ is orthonormal if and only if the transition matrix is orthogonal. □

[*] with respect to the dot product

32.3 Real symmetric matrices

Theorem 32.7

Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. If A is symmetric (i.e., $A^T = A$), then A is orthogonally diagonalisable.

Although we postpone the proof of this theorem, we note the following

Proposition 32.8

Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$. If A is symmetric, then:

1. All roots of $c_A(x)$ are real.
2. Eigenvectors having different eigenvalues are orthogonal.

Proof. (Recall that for a matrix $M \in \mathcal{M}_{m,n}(\mathbb{C})$, A^* denotes the conjugate transpose $A^* = \overline{A}^T$.)

Suppose that $\lambda \in \mathbb{C}$ and $v \in \mathcal{M}_{n,1}(\mathbb{C}) \setminus \{\vec{0}\}$ are such that $Av = \lambda v$. Then $v^* A\, v$ is a 1×1 matrix and $(v^* A\, v)^T = v^* A\, v$ (since any 1×1 matrix is symmetric). It follows that

$$\overline{v^* A\, v} = \overline{(v^* A\, v)^T} = \overline{v^T A^T \overline{v}} = v^* A^*\, v = v^* A\, v$$

Therefore $v^*\, A\, v \in \mathbb{R}$. Since $v^*\, A\, v = v^* \lambda v = \lambda v^* v$, and both $v^*\, A\, v$ and $v^* v$ are real numbers (and $v^* v > 0$), it follows that λ is real.

Suppose $v_1, v_2 \in \mathcal{M}_{n,1}(\mathbb{R})$ are two eigenvectors of A with $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$.
Then

$$v_1^T A v_2 = v_1^T(\lambda_2 v_2) = \lambda_2 v_1^T v_2$$

and also

$$v_1^T A v_2 = v_1^T A^T v_2 = (Av_1)^T v_2 = (\lambda_1 v_1)^T v_2 = \lambda_1 v_1^T v_2$$

Therefore $\lambda_2 v_1^T v_2 = \lambda_1 v_1^T v_2$ which, if $\lambda_1 \neq \lambda_2$, implies $v_1^T v_2 = 0$. □

Algorithm 32.9: To orthogonally diagonalise a real symmetric matrix

Let $A \in \mathcal{M}_{n,n}(\mathbb{R})$ and suppose that $A^T = A$.

1. Find the eigenvalues of A.
2. For each eigenvalue
   (a) Find a basis for the eigenspace
   (b) Use Gram-Schmidt to convert to an orthonormal basis
3. The union of the eigenspace bases will be an orthonormal basis $\{u_1, \dots, u_n\}$ for $\mathbb{R}^n$. Letting Q be the matrix whose columns are given by the u_i and letting D be the diagonal matrix whose diagonal entries are the corresponding eigenvalues[†]we then have

$$A = QDQ^T$$

[†] The order of the eigenvalues must correspond to the order of the eigenvectors u_i

Example 32.10. We apply the above to the matrix $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$,

To find the eigenvalues:

$$\begin{aligned}
\det(A - xI_3) &= \det \begin{bmatrix} 4-x & 2 & 2 \\ 2 & 4-x & 2 \\ 2 & 2 & 4-x \end{bmatrix} = \det \begin{bmatrix} 4-x & 2 & 2 \\ x-2 & 2-x & 0 \\ x-2 & 0 & 2-x \end{bmatrix} \qquad (R_2 - R_1, R_3 - R_1) \\
&= (x-2)^2 \det \begin{bmatrix} 4-x & 2 & 2 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \\
&= (x-2)^2 \left(-\det \begin{bmatrix} 2 & 2 \\ 0 & -1 \end{bmatrix} - \det \begin{bmatrix} 4-x & 2 \\ 1 & -1 \end{bmatrix} \right) \qquad \text{(expanding along the second row)} \\
&= (x-2)^2 \left(2 - (x-6)\right) = (x-2)^2(8-x)
\end{aligned}$$

The eigenvalues are 2 and 8.

To find a basis for the eigenspace with $\lambda = 2$:

$$A - 2I_3 = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A basis for the eigenspace is therefore $\{(-1,1,0), (-1,0,1)\}$.

We apply Gram-Schmidt to find an orthonormal basis for the ($\lambda = 2$) eigenspace:

$$\begin{aligned}
u_1 &= \frac{1}{\sqrt{2}}(-1,1,0) \\
w_2 &= (-1,0,1) - \langle(-1,0,1), u_1\rangle u_1 \\
&= (-1,0,1) - \frac{1}{2}\langle(-1,0,1),(-1,1,0)\rangle(-1,1,0) \\
&= (-1,0,1) - \frac{1}{2}(-1,1,0) \\
&= (-\frac{1}{2}, -\frac{1}{2}, 1) = \frac{1}{2}(-1,-1,2) \\
u_2 &= \frac{w_2}{\|w_2\|} = \frac{(-1,-1,2)}{\|(-1,-1,2)\|} = \frac{1}{\sqrt{6}}(-1,-1,2)
\end{aligned}$$

Therefore $\{\frac{1}{\sqrt{2}}(-1,1,1), \frac{1}{\sqrt{6}}(-1,-1,2)\}$ is an orthonormal basis for the eigenspace.

Now for the $\lambda = 8$ eigenspace:

$$A - 8I_3 = \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

A basis for the eigenspace is therefore $\{(1,1,1)\}$.

We apply Gram-Schmidt to find an orthonormal basis for the ($\lambda = 8$) eigenspace:

$$u_3 = \frac{1}{\sqrt{3}}(1,1,1)$$

Finally, if we take

$$Q = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

we have that $A = QDQ^T$.

32.4 Exercises

300. Determine whether or not the given matrix A is orthogonal.

(a) $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

301. Show that the rotation matrix $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal.

302. For each symmetric matrix A below find a decomposition $A = QDQ^T$, where Q is orthogonal and D diagonal.

(a) $\begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix}$

(b) $\begin{bmatrix} 7 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} -2 & 0 & -36 \\ 0 & -3 & 0 \\ -36 & 0 & -23 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Extra material for lecture 32

▷ Reference

Elementary Linear Algebra by Anton and Rorres, §7, p401

Proof of the spectral theorem

Our goal is to show that every real symmetric matrix is orthogonally diagonalisable (Theorem 32.7). The proof is more easily understood when presented in terms of linear transformations rather than matrices.

Definition 33.1

Let V be a finite dimensional real inner product space. We will call a linear transformation $T : V \to V$ **symmetric** if it has the property that $\langle T(u), v\rangle = \langle u, T(v)\rangle$ for all $u, v \in V$.

Exercise 303. Let V be a finite dimensional real inner product space. Let $\mathcal{B}$ be an orthonormal basis for V and let $T : V \to V$ be a linear transformation. Show that T is symmetric if and only if the matrix $[T]_{\mathcal{B}}$ is symmetric.

Theorem 33.2

Let V be a finite-dimensional real inner product space and $T : V \to V$ a linear transformation. If T is symmetric, then there exists an orthonormal basis for V all of whose elements are eigenvectors of T.

Proof. We use induction on $n = \dim(V)$.

Base case: If n=1, then we choose any non-zero vector $u \in V$ and take $\mathcal{B} = \{u/\|u\|\}$.

Induction step: Assume that for any $(n-1)$-dimensional inner product space W and every symmetric linear transformation $S : W \to W$ there exsits an orthonormal basis for W made up of eigenvectors of S.

Let λ be an eigenvalue of T. From Proposition 32.8 we know that $\lambda \in \mathbb{R}$. Let $u \in V$ be an eigenvector of T with eigenvalue λ and with $\|u\| = 1$. Let $W = \{v \in V \mid \langle v, u\rangle = 0\}$. Then we have

(a) W is a subspace of V (b) $T(W) \subseteq W$ (c) $\dim(W) = n - 1$

Let $S : W \to W$ be given by $S(w) = T(w)$. Then S is a symmetric linear transformation. By the induction hypothesis, there exists an orthonormal basis $\mathcal{C}$ for W such that all elements in $\mathcal{C}$ are eigenvectors of S. We therefore have that $\mathcal{C} \subseteq V$ is an orthonormal set and all its elements are eigenvectors of T. Hence $\{u\} \cup \mathcal{C}$ is an orthonormal basis for V made up of eigenvectors of T.

Therefore, by mathematical induction a symmetric linear transformation on a finite dimensional real inner product space is orthogonally diagonalisable. □

33.1 An application: conic sections

Suppose that we would like to plot the set of points $(x, y) \in \mathbb{R}^2$ that satisfy the equation

$$6x^2 - 4xy + 3y^2 = 1$$

We can use orthogonal diagonalisation to eliminate the cross terms in the above equation.

The equation can be written as $X^TAX = 1$ where $X = \begin{bmatrix} x \\ y \end{bmatrix}$ and $A = \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$.

Since A is real symmetric we know that it can be orthogonally diagonalised. Calculation gives $A = QDQ^T$ with

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \quad \text{and} \quad Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

Note that

$$X^TAX = X^TQDQ^TX = (Q^TX)^TD(Q^TX) \tag{$*$}$$

Let $\mathcal{B} = \{b_1 = \frac{1}{\sqrt{5}}(1,2), b_2 = \frac{1}{\sqrt{5}}(-2,1)\}$ be the orthonormal basis of $\mathbb{R}^2$ corresponding to the columns of Q. We rewrite the above equation using coordinates with respect to $\mathcal{B}$.

Let $X' = \begin{bmatrix} x' \\ y' \end{bmatrix}$ be the coordinates of the point (x, y) with respect to $\mathcal{B}$. Note that $P_{\mathcal{S},\mathcal{B}} = Q$ and therefore $P_{\mathcal{B},\mathcal{S}} = Q^{-1} = Q^T$.

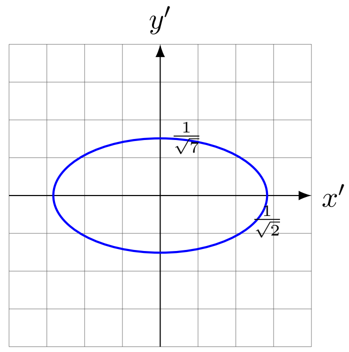
We have

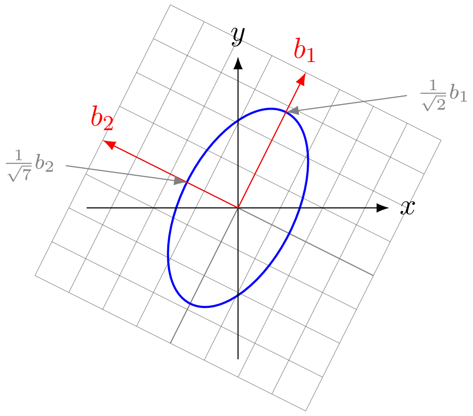
$$X' = [(x,y)]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{S}}[(x,y)]_{\mathcal{S}} = Q^TX$$

Then we have

$$\begin{aligned}
6x^2 - 4xy + 3y^3 = 1 &\iff X^TAX = 1 \\
&\iff (Q^TX)^TD(Q^TX) = 1 && \text{(from $*$)} \\
&\iff (X')^TDX' = 1 \\
&\iff \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = 1 \\
&\iff 2(x')^2 + 7(y')^2 = 1
\end{aligned}$$

The curve is now much easier to recognise as an ellipse.





33.2 Exercises

304. Use orthogonal diagonalisation to sketch the curve given by the following equation:

$$5x^2 - 4xy + 8y^2 = 36$$

305. Prove the statements (a), (b), and (c) about W in the above proof of Theorem 33.2.

306. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(x, y) = (-6x, -5x+4y)$. The following defines an inner product on $\mathbb{R}^2$

$$\langle (a,b), (x,y) \rangle = ax - ay - bx + 2by$$

   (a) Show that T is symmetric (with respect to the inner product above).
   (b) Find a basis for $\mathbb{R}^2$ that is orthonormal (with respect to the inner product above) and composed of eigenvectors of T.

Extra material for lecture 33

- ▷ For a longer discussion of the application of diagonalisation to conic sections see

  Elementary Linear Algebra by Anton and Rorres, §7.2, p417

  The same approach can be applied to *quadric surfaces*.

- ▷ Let V be a finite-dimensional inner product space and $W \leqslant V$ a subspace.

  (a) Show that projection onto W $\text{proj}_W : V \to V$ is symmetric.

  (b) Show that reflection across W is symmetric.

- ▷ Let $V \leqslant \mathcal{F}([0,1],\mathbb{R})$ be the real inner product space of all smooth functions $f : [0,1] \to \mathbb{R}$ that satisfy $f(0) = f(1) = 0$. That is

$$V = \{f \in \ \mathcal{F}([0,1],\mathbb{R}) \mid f(0) = f(1) = 0, \text{and } f \text{ is smooth}\}$$

  with inner product

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$$

  Let $D : V \to V$ be the linear transformation given by $D(f(x)) = \frac{df}{dx}$

  (a) Use integration by parts to show that $\forall f, g \in V$, $\langle D(f), g\rangle = -\langle f, D(g)\rangle$

  (b) Show that D^2 is symmetric. (Explicitly, $D^2 : V \to V$,$D^2(f(x)) = \frac{d^2f}{dx^2}$)

Unitary diagonalisation

We note that the results on orthogonal diagonalisation carry over to complex matrices. The proofs given in the real case apply with only minor changes, and will not be repeated.

34.1 Unitary matrices

Recall that for a matrix $A \in \mathcal{M}_{m,n}(\mathbb{C})$ we denote by A^* the conjugate transpose $A^* = \overline{A}^T$.

Definition 34.1

A matrix $U \in \mathcal{M}_{n,n}(\mathbb{C})$ is called **unitary** if $U^*U = I_n$.

Examples 34.2.

$$\frac{1}{\sqrt{2}}\begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}, \qquad \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} \in \mathcal{M}_{2,2}(\mathbb{C})$$

Exercise 307. Suppose the $U, P \in \mathcal{M}_{n,n}(\mathbb{C})$ are unitary matrices. Show that

(a) U^* is unitary

(b) U^{-1} is unitary

(c) PU is unitary

(d) $|\det(U)| = 1$

Lemma 34.3: Conditions equivalent to being unitary

Let $U \in \mathcal{M}_{n,n}(\mathbb{C})$. The following are equivalent

1. U is unitary
2. the columns of U form an orthonormal basis of $\mathbb{C}^n$ (with respect to the complex dot product)
3. the rows of U form an orthonormal basis of $\mathbb{C}^n$ (with respect to the complex dot product)
4. $\|Uv\| = \|v\|$ for all $v \in \mathbb{C}^n$ (complex dot product)
5. $\langle Uu, Uv\rangle = \langle u, v\rangle$ for all $u, v \in \mathbb{C}^n$ (complex dot product)

Proof. We prove that 4 implies 5.

$$\langle u+v, u+v\rangle - \langle u-v, u-v\rangle = 2(\langle u,v\rangle + \langle v,u\rangle) = 4\Re(\langle u,v\rangle)$$
$$\langle U(u+v), U(u+v)\rangle - \langle U(u-v), U(u-v)\rangle = 2(\langle Uu,Uv\rangle + \langle Uv,Uu\rangle) = 4\Re(\langle Uu,Uv\rangle)$$

Since 4 holds, the left hand sides above are equal and therefore $\Re(\langle u,v\rangle) = \Re(\langle Uu,Uv\rangle$. Putting iv in place of v in the above calculation gives

$$\langle u,v\rangle - \langle v,u\rangle = \langle Uu,Uv\rangle - \langle Uv,Uu\rangle$$

and therefore $\Im(\langle u,v\rangle) = \Im(\langle Uu,Uv\rangle$ (the imaginary parts are equal). □

34.2 Unitary diagonalisation

Definition 34.4

A matrix $A \in \mathcal{M}_{n,n}(\mathbb{C})$ is said to be **unitarily diagonalisable** if there is a unitary matrix $U \in \mathcal{M}_{n,n}(\mathbb{C})$ and a diagonal matrix $D \in \mathcal{M}_{n,n}(\mathbb{C})$ such that $A = U\,D\,U^*$.

Example 34.5. The matrix $A = \begin{bmatrix} 1 & 2i \\ -2i & -2 \end{bmatrix}$ is orthogonally diagonalisable since

$$A = \begin{bmatrix} 1 & 2i \\ -2i & -2 \end{bmatrix} = \begin{bmatrix} -\frac{i}{\sqrt{5}} & \frac{2i}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{i}{\sqrt{5}} & \frac{2i}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^* = UDU^*$$

Theorem 34.6

A matrix $A \in \mathcal{M}_{n,n}(\mathbb{C})$ is unitarily diagonalisable if and only if there is an orthonormal basis $\mathcal{B}$ for $\mathbb{C}^n$ with the property that all elements of $\mathcal{B}$ are eigenvectors of A.

34.3 Hermitian matrices

Theorem 34.7

Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$. If A is Hermitian (i.e., $A^* = A$), then A is unitarily diagonalisable.

Proposition 34.8

Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$. If A is Hermitian (i.e., $A^* = A$), then:

1. All eigenvalues of A are real;
2. Eigenvectors having different eigenvalues are orthogonal.

Algorithm 34.9: To unitarily diagonalise a Hermitian matrix

Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$ and suppose that $A^* = A$.

1. Find the eigenvalues of A.
2. For each eigenvalue
   (a) Find a basis for the eigenspace
   (b) Use Gram-Schmidt to convert to an orthonormal basis
3. The union of the eigenspace bases will be an orthonormal basis $\{u_1, \dots, u_n\}$ for $\mathbb{C}^n$. Letting U be the matrix whose columns are given by the u_i and letting D be the diagonal matrix whose diagonal entries are the corresponding eigenvalues, we then have

$$A = UDU^*$$

Example 34.10. Let $A = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$.

To find the eigenvalues of A:

$$\det(xI_2 - A) = \det \begin{bmatrix} x-1 & -1-i \\ -1+i & x-2 \end{bmatrix} = (x-1)(x-2) - (1+i)(1-i) = x^2 - 3x = x(x-3)$$

The eigenvalues of A are: $0, 3$.

To find the eigenvectors of A:

$$A - 0I_2 = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1+i \\ 0 & 0 \end{bmatrix}$$

A basis for the $\lambda = 0$ eigenspace is $\{(1+i, -1)\}$. An orthonormal basis is $\{\frac{1}{\sqrt{3}}(1+i, -1)\}$.

$$A - 3I_2 = \begin{bmatrix} -2 & 1+i \\ 1-i & -1 \end{bmatrix} \sim \begin{bmatrix} -2 & 1+i \\ 0 & 0 \end{bmatrix}$$

A basis for the $\lambda = 3$ eigenspace is $\{(1+i, 2)\}$. An orthonormal basis is $\{\frac{1}{\sqrt{6}}(1+i, 2)\}$.

(Notice that $(1+i, -1) \cdot (1+i, 2) = (1+i)(1-i) - 2 = 0$)

Letting $U = \begin{bmatrix} \frac{1+i}{\sqrt{3}} & \frac{1+i}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix}$ and $D = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$ we have that $A = UDU^*$.

34.4 Exercises

308. Unitarily diagonalse the Hermitian matrix $A = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix}$.

309. Let $M \in \mathcal{M}_{n,n}(\mathbb{C})$ be an Hermitian matrix. Show that M is positive definite (see definition 30.1) if and only if all eigenvalues of M are strictly positive. (Hint: unitarily diagonalise M.)

310. A linear transformation $T : V \to V$ on a complex inner product space V is called **self-adjoint** if
$$\forall u, v \in V, \quad \langle T(u), v \rangle = \langle u, T(v) \rangle$$
    (a) Show that all eigenvalues of T are real.
    (b) Show that eigenvectors corresponding to distinct eigenvalues are orthogonal

311. Let $A \in \mathcal{M}_{n,n}(\mathbb{C})$ and suppose that $A^* = -A$. Suppose that $\lambda \in \mathbb{C}$ is an eigenvalue of A. Show that $\lambda = iy$ for some $y \in \mathbb{R}$.

Further material for lecture 34

▷ A matrix $A \in \mathcal{M}_{n,n}(\mathbb{C})$ is called **normal** if $A^*A = AA^*$. Normal matrices are unitarily diagonalisable. (This will be covered in the subject *MAST20022 Group Theory and Linear Algebra*.)

All Hermitian matrices are normal. The matrix $\left[\begin{smallmatrix} 2+i & 2-i \\ 2-i & 2+i \end{smallmatrix}\right]$ is normal but not Hermitian.

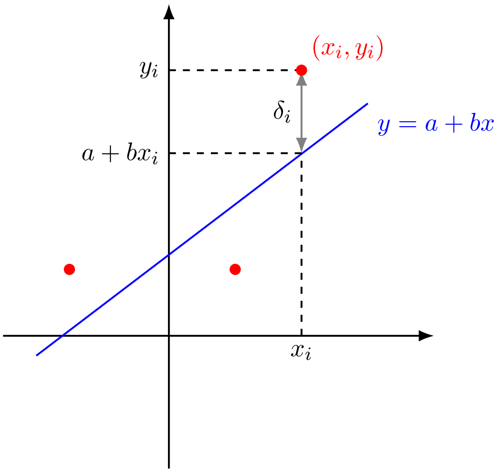
Least squares approximation

We give two applications related to the dot product in $\mathbb{R}^n$.

35.1 Least squares line of best fit

Given a set of data points (x_1, y_1), (x_2, y_2),..., (x_n, y_n) we want to find the straight line $y = a + bx$ which best approximates the data. A common approach is to minimise the **least squares error**:

$$\begin{aligned} E &= \text{sum of the squares of the vertical errors } \delta_i \\ &= \sum_{i=1}^{n} \delta_i^2 \\ &= \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \end{aligned}$$



Given (x_1, y_1),..., (x_n, y_n) we want to find $a, b \in \mathbb{R}$ that minimise the quantity $\sum_{i=1}^{n}(y_i - (a + bx_i))^2$ This can be written as

$$E = \sum_{i=1}^{n}(y_i - (a + bx_i))^2 = \|\mathbf{y} - A\mathbf{u}\|^2$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$$

The length is that coming from the inner product on $\mathcal{M}_{n,1}(\mathbb{R})$ that corresponds to the dot product on $\mathbb{R}^n$, that is $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \mathbf{v}_1^T \mathbf{v}_2$.

To minimise $\|\mathbf{y} - A\mathbf{u}\|$ we want $\mathbf{u}$ to be such that $A\mathbf{u}$ is as close as possible to $\mathbf{y}$ (which is fixed).

That is, we want the vector in

$$W = \{A\mathbf{v} \mid \mathbf{v} \in \mathcal{M}_{2,1}(\mathbb{R})\} \leqslant \mathcal{M}_{n,1}(\mathbb{R}) \qquad (W \text{ is the column space of } A)$$

that is closest to $\mathbf{y}$.

The closest vector is precisely $\mathbf{p} = \text{proj}_W(\mathbf{y})$. To find $\mathbf{u}$ we *could* project $\mathbf{y}$ to W to get $\mathbf{p}$ and then solve $A\mathbf{u} = \mathbf{p}$ to get $\mathbf{u}$ (by solving a linear system).

However, we can calculate $\mathbf{u}$ more directly (without finding an orthonormal basis for W) by using properties of the projection:

$$\begin{aligned}
& & \mathbf{w}^T(\mathbf{y} - \mathrm{proj}_W\,\mathbf{y}) = 0 \quad & \forall\, \mathbf{w} \in W \\
&\implies & (A\mathbf{v})^T(\mathbf{y} - A\mathbf{u}) = 0 \quad & \forall\, \mathbf{v} \in \mathcal{M}_{2,1}(\mathbb{R}) \qquad \text{(since } \mathbf{w} = A\mathbf{v}\text{)} \\
&\implies & \mathbf{v}^T A^T(\mathbf{y} - A\mathbf{u}) = 0 \quad & \forall\, \mathbf{v} \in \mathcal{M}_{2,1}(\mathbb{R}) \\
&\implies & A^T(\mathbf{y} - A\mathbf{u}) = \vec{0} & \\
&\implies & A^T\mathbf{y} - A^TA\mathbf{u} = \vec{0} & \\
&\implies & (A^TA)\mathbf{u} = A^T\mathbf{y} & \qquad (*)
\end{aligned}$$

From this we can calculate $\mathbf{u}$, given that we know A and $\mathbf{y}$. It's just a matter of solving the linear system.

Note that $A^TA \in \mathcal{M}_{2,2}(\mathbb{R})$. If A^TA is invertible (and it usually is), the solution to $(*)$ is given by

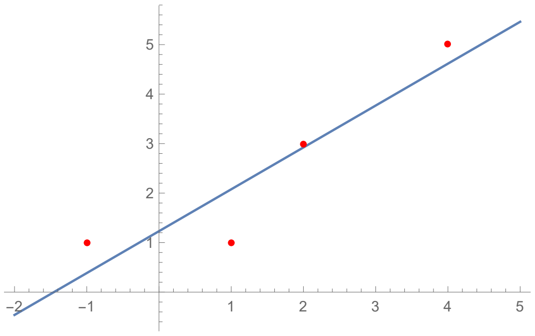
$$\mathbf{u} = (A^TA)^{-1}A^T\mathbf{y}$$

Example 35.1. We find the straight line which best fits the data points: $(-1, 1)$, $(1, 1)$, $(2, 3)$, $(4, 5)$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 22 \end{bmatrix}$$

$$(A^TA)^{-1}A^T\mathbf{y} = \frac{1}{52}\begin{bmatrix} 22 & -6 \\ -6 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \\ 5 \end{bmatrix} = \frac{1}{13}\begin{bmatrix} 16 \\ 11 \end{bmatrix}$$

The line of best fit is $y = \frac{16}{13} + \frac{11}{13}x$



35.2 Polynomial of best fit

The same method works for finding quadratic (or higher degree) fitting curves.

To find the quadratic $y = a + bx + cx^2$ which best fits data (x_1, y_1), (x_2, y_2),..., (x_n, y_n) we take

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

and solve

$$A^TA\mathbf{u} = A^T\mathbf{y}$$

for

$$\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

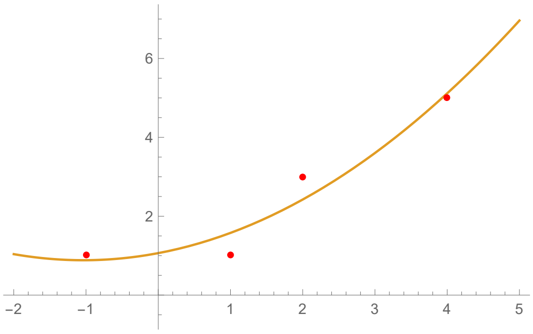
Example 35.2. We find the parabola which best fits the data points: $(-1, 1)$, $(1, 1)$, $(2, 3)$, $(4, 5)$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 4 \\ 1 & 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 22 \\ 6 & 22 & 72 \\ 22 & 72 & 274 \end{bmatrix}$$

$$(A^TA)^{-1}A^T\mathbf{y} = \begin{bmatrix} \frac{83}{78} \\ \frac{9}{26} \\ \frac{1}{6} \end{bmatrix}$$

The parabola of best fit is

$y = \frac{83}{78} + \frac{9}{26}x + \frac{1}{6}x^2$



35.3 Exercises

312. Find the (least squares) line of best fit for the given data sets.

    (a) $\{(0,0),(1,0),(2,1),(3,3),(4,5)\}$

    (b) $\{(-2,2),(-1,1),(0,-1),(1,0),(2,3)\}$

313. A maths lecturer was placed on a rack by his students and stretched to lengths $L = 1.7, 2.0$ and 2.3 metres when forces of $F = 1, 2$ and 4 tonnes were applied. Assuming Hooke's law $L = a + bF$, estimate the lecturer's normal length a.

314. A firm that manufactures widgets finds the daily consumer demand $d(x)$ for widgets as a function of their price x is as in the following table:

x	1	1.5	2	2.5	3
$d(x)$	200	180	150	100	25

Using least squares, approximate the daily consumer demand by a linear function.

315. Find the parabola of best fit for the data in Exercise 312(a).
(Use MATLAB for the matrix algebra in this question!)

Extra material for lecture 35

- ▷ Reference

  Elementary Linear Algebra by Anton and Rorres, §6.5, p387

APPENDIX A

Cardinality

What should it mean for two sets to have the same 'size'? Does it make sense to say that there are more rational numbers than there are natural numbers?[*] Are there more real numbers than there are rationals?[†]

Rather than trying to define the size of a set directly, it is convenient to introduce the notion of two sets 'having the same number of elements'. The notion of a bijective function is clearly just what is required.

Definition A.1

Two sets A and B are said to have the **same cardinality** if there exists a bijection $A \to B$. A set is **finite** if it is either empty or has the same cardinality as the set $\{1, 2, \ldots, n\}$ for some n. A set that is not finite is called **infinite**. A set is called **countably infinite** is it has the same cardinality as $\mathbb{N}$. A set is called **countable** if it is either finite or countably infinite. A set that is not countable is called **uncountable**.

Lemma A.2

Let A be a set.

1. If there exists an injective function $A \to \mathbb{N}$, then A is countable.
2. If there exists a surjective function $\mathbb{N} \to A$, then A is countable.

Proof. For the first statement it suffices to show that if B is an infinite subset of $\mathbb{N}$, then there is a bijection $\varphi : \mathbb{N} \to B$. We inductively define a sequence of subsets $B_i \subset \mathbb{N}$ together with the required function $\varphi : \mathbb{N} \to B$. Let $B_1 = B$. Suppose that B_i has been defined and is infinite. Let $b_i = \min(B_i)$ and define $\varphi(i) = b_i$ and $B_{i+1} = B_i \setminus \{b_n\}$. Since B_i is infinite, B_{i+1} is infinite. The finction $\varphi : \mathbb{N} \to B$ defined inductively in this way is clearly a bijection.

For the second statement, suppose now that there exists a surjective function $f : \mathbb{N} \to A$. If A is finite, then A is countable by definition and there is nothing to prove. We can assume, therefore, that A is infinite. We define a map $g : \mathbb{N} \to A$ as follows. Define $M \subseteq \mathbb{N}$ by

$$M = \{m \in \mathbb{N} \mid f(m) \notin \{f(1), f(2), \ldots, f(m-1)\}\}$$

Note that

1. M is infinite since A is infinite
2. $f(M) = A$
3. If $m, n \in M$ and $m \neq n$, then $f(m) \neq f(n)$

Denote by m_i the i-th element of M (using the usual ordering on $\mathbb{N}$) and define

$$g : \mathbb{N} \to A, \quad g(i) = f(m_i)$$

[*] no

[†] yes

That g is bijective follows from properties 2 and 3 above. □

Proposition A.3

1. $\mathbb{Z}$ and $\mathbb{N}$ have the same cardinality (i.e., $\mathbb{Z}$ is countably infinite)
2. $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$ have the same cardinality (i.e., $\mathbb{N} \times \mathbb{N}$ is countably infinite)
3. $\mathbb{Q}$ and $\mathbb{N}$ have the same cardinality (i.e., $\mathbb{Q}$ is countably infinite)

Proof.

1. The map $f : \mathbb{N} \to \mathbb{Z}$ given by $f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ \frac{1-n}{2} & \text{if } n \text{ is odd} \end{cases}$ is a surjection.

   (In fact it's a bijection.)

2. The map $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $g(m, n) = 2^m 3^n$ is injective.

3. From the first two parts, we know that $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ has the same cardinality as $\mathbb{N}$. Then note that the map $h : \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) \to \mathbb{Q}$ given by $h(m, n) = \frac{m}{n}$ is surjective.

□

At this point we might start to think that all infinite sets are countably infinite. But we'd be wrong.

Theorem A.4

$\mathbb{R}$ is uncountable.

Proof. Suppose, for a contradiction, that the interval $(0, 1) \subset \mathbb{R}$ is countable and let $f : \mathbb{N} \to (0, 1)$ be a bijection. We consider the decimal expansion of each element:

$$\begin{aligned} f(1) &= 0.a_{11}a_{12}a_{13}\ldots \\ f(2) &= 0.a_{21}a_{22}a_{23}\ldots \\ &\vdots \\ f(i) &= 0.a_{i1}a_{i2}a_{i3}\ldots \\ &\vdots \end{aligned}$$

Each $a_{ij} \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and a_{ij} is the j-th digit in the decimal expansion of $f(i)$. Define $b \in (0, 1)$ as follows:

$$b = 0.b_1b_2b_3\ldots \qquad \text{where} \qquad b_i = \begin{cases} 7 & \text{if } a_{ii} = 8 \\ 8 & \text{if } a_{ii} \neq 8 \end{cases}$$

Notice that for all $i \in \mathbb{N}$, $b_i \neq a_{ii}$. It follows that for all $i \in \mathbb{N}$, $b \neq f(i)$. This contradicts the assumption that f is surjective. □

A similar 'diagonal argument' establishes the following.

Theorem A.5

Let A be a set. The power set of A, $\mathcal{P}(A)$, does not have the same cardinality as A.

Proof. Suppose, for a contradiction, that there exists a bijection $f : A \to \mathcal{P}(A)$. Define $B = \{a \in A \mid a \notin f(A)\}$. Since $B \subseteq A$ and f is surjective, there exists $b \in A$ such that $f(b) = B$. Then we have

$$\begin{aligned} b \in f(b) &\iff b \in B && \text{(since } f(b) = B\text{)} \\ &\iff b \notin f(b) && \text{(definition of } B\text{)} \end{aligned}$$

□

Existence of bases

THIS IS FOR INTEREST ONLY!

The goal is to present a proof of the following theorem.

Theorem B.1

Let V be a vector space.

1. Every spanning set of V contains a basis of V.
2. Every linearly independent set in V can be extended to a basis of V.
3. Any two bases of V have the same cardinality.

If we assume that V is finite dimensional, the theorem is much easier to prove (see Lecture 17). In the general case, which we shall consider here, the proof uses some fundamental results from the theory of infinite sets which are stated without proof in the second section.

Notice that, since V itself is a spanning set for V, we have the following consequence of the theorem.

Corollary B.2

Every vector space has a basis.

B.1 Proof of the theorem

We start by proving the following.

Lemma B.3

Let V be a vector space over $\mathbb{F}$ and let $X, Y \subseteq V$ be two subsets. If X is linearly independent and Y spans V, then there is a subset $Y' \subseteq Y$ such that $X \cup Y'$ is a basis of X.

Proof of Lemma B.3. Define $\mathcal{S}$ to be the collection of all subsets Z of Y such that $X \cup Z$ is linearly independent, that is:

$$\mathcal{S} = \{Z \mid Z \subseteq Y \text{ and } X \cup Z \quad \text{is linearly independent}\}$$

Let $Y' \in \mathcal{S}$ be **maximal** in $\mathcal{S}$, that is, for all $Z \in \mathcal{S}$ we have:

$$Z \supseteq Y' \implies Z = Y'$$

How do we know that such a maximal element exists? If Y is a finite set, then $\mathcal{S}$ is also finite and the existence is clear. If Y is infinite, the existence of a maximal element Y' is less obvious, and is in fact a fundamental property of set theory. It is called "Zorn's Lemma"[*] (see the next section).

[*] In fact, it is not really a lemma. It is equivalent to something called the "Axiom of Choice," which is independent of the other axioms of set theory.

By construction $X \cup Y'$ is linearly independent. We claim that it is also a spanning set for V. We know that for all $y \in Y$, $y \in \operatorname{span}(X \cup Y')$ since otherwise $X \cup Y' \cup \{y\}$ would be linearly independent, which contradicts the maximality of Y'. Thus $Y \subseteq \operatorname{span}(X \cup Y')$ and hence $\operatorname{span}(Y) \subseteq \operatorname{span}(X \cup Y')$. Since $\operatorname{span}(Y) = V$ and $\operatorname{span}(X \cup Y') \subseteq V$, it follows that $\operatorname{span}(X \cup Y') = V$. Therefore $X \cup Y'$ is a spanning set for V, and hence a basis. □

We can now prove the first two parts of the theorem.

Proof (of parts 1 and 2 of the theorem).

1. Suppose that Y is a spanning set. Applying Lemma B.3 with this Y and $X = \emptyset$ yields a subset $Y' \subseteq Y$ such that Y' is a basis.

2. Suppose now that X is any linearly independent set. Taking $Y = V$ and applying Lemma B.3 yields a basis that contains X. □

To prove the third part of the theorem, we will use the following lemma.

Lemma B.4

Let V be a vector space over $\mathbb{F}$, let $Y \subseteq V$ be a spanning set for V, and let $\{x_1, \dots, x_m\} \subseteq V$ be a linearly independent set of vectors. If $|Y| \geqslant m$, then there are elements $y_1, \dots, y_m \in Y$ such that $\{Y \setminus \{y_1, \dots, y_m\}\} \cup \{x_1, \dots, x_m\}$ is a spanning set for V. (That is, replacing the y_i by the x_i still gives a spanning set.)

Proof. We first note that since $\{x_1, \dots, x_m\}$ is linearly independent, all of the x_i are non-zero. Since Y is a spanning set, there exist $a_1, \dots, a_k \in Y$ and $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that $x_1 = \alpha_1 a_1 + \cdots + \alpha_k a_k$. As x_1 is non-zero, at least one of the α_i is non-zero. By re-ordering the a_i if necessary we may assume that $\alpha_1 \neq 0$. We then have that

$$a_1 = \left(\frac{1}{\alpha_1}\right) x_1 - \left(\frac{\alpha_2}{\alpha_1}\right) a_2 - \cdots - \left(\frac{\alpha_k}{\alpha_1}\right) a_k$$

Let $y_1 = a_1$. Using the above expression any linear combination of elements from Y can be rewritten as a linear combination of vectors from $Y_1 = \{Y \setminus \{y_1\}\} \cup \{x_1\}$. We simply replace any occurance of y_1 by the right hand side of the above expression. This then gives a linear combination which does not involve y_1, but does involve x_1. It follows that $\operatorname{span}(Y) \subseteq \operatorname{span}(Y_1)$, and therefore $\operatorname{span}(Y_1) = V$.

Suppose now that we have found $y_1, \dots, y_l \in Y$ (where $1 \leqslant l < m$) such that $Y_l = \{Y \setminus \{y_1, \dots, y_l\}\} \cup \{x_1, \dots, x_l\}$ is a spanning set for V. Since Y_l is a spanning set, there exist $b_1, \dots, b_k \in Y \setminus \{y_1, \dots, y_l\}$ and $\beta_1, \dots, \beta_k, \gamma_1 \dots, \gamma_l \in \mathbb{F}$ such that $x_{l+1} = \beta_1 b_1 + \cdots + \beta_k b_k + \gamma_1 x_1 + \cdots + \gamma_l x_l$. Since x_{l+1} is non-zero, at least one of the β_i or γ_j is non-zero. Indeed, not all the β_i can be zero, as that would contradict the linear independence of the set $\{x_1, \dots, x_{l+1}\}$. Re-ordering if necessary, we may assume that $\beta_1 \neq 0$. Then

$$b_1 = \left(\frac{1}{\beta_1}\right) x_{l+1} - \left(\frac{\beta_2}{\beta_1}\right) b_2 - \cdots - \left(\frac{\beta_k}{\beta_1}\right) b_k - \left(\frac{\gamma_1}{\beta_1}\right) x_1 - \cdots - \left(\frac{\gamma_l}{\beta_1}\right) x_l$$

Letting $y_{l+1} = b_1$ and $Y_{l+1} = \{Y \setminus \{y_1, \dots, y_{l+1}\}\} \cup \{x_1, \dots, x_{l+1}\}$, we have, as above, that $\operatorname{span}(Y_{l+1}) = V$. The lemma then follows by induction. □

The third part of the theorem follows from:

Lemma B.5

Let V be a vector space over $\mathbb{F}$ and let $X, Y \subseteq V$ be two subsets. If X is linearly independent and Y spans V, then $|X| \leqslant |Y|$.

Proof. We first prove the lemma under the assumption that Y is finite. Let $Y = \{y_1, \dots, y_k\}$. Suppose, in order to get a contradiction, that $|X| > k$. Choose distinct elements $x_1, \dots, x_k \in X$. Then $\{x_1, \dots, x_k\}$ is linearly independent (since X is), and applying Lemma B.4 we know that

$$\{x_1, \dots, x_k\} = (Y \setminus \{y_1, \dots, y_k\}) \cup \{x_1, \dots, x_k\}$$

is a spanning set for V. Since $|X| > k$, there is an element $x \in X \setminus \{x_1, \dots, x_k\}$. As $\{x_1, \dots, x_k\}$ is a spanning set for V, x can be expressed as a linear combination of the x_i. This contradicts the linear independence of X, so we must in fact have $|X| \leqslant k$.

Consider now the case where Y is not finite. Denote by $F(Y)$ the set of all finite subsets of Y. Then $|F(Y)| = |Y|$ (See Lemma B.6). Define a map $\Phi : X \to F(Y)$ as follows: For each $x \in X$ we choose a a finite subset $S_x \subset Y$ such that x is a linear combination of S_x, and define $\Phi(x) = S_x$. If $|X| > |F(Y)|$ then there is some element $S \in F(X)$ with infinite preimage (see Lemma B.6). We would then have a finite set $S \subset Y$ such that $\Phi^{-1}(S) \subset X$ is an infinite, linearly independent subset of $\mathrm{span}(S)$. This would contradict the first case of this proof. □

Proof of part 3 of the theorem. Let B_1 and B_2 be two bases for V. Since B_1 is linearly independent and B_2 is a spanning set, Lemma B.5 implies that $|B_1| \leqslant |B_2|$. On the other hand, B_2 is linearly independent and B_1 is a spanning set, so we also have $|B_2| \leqslant |B_1|$. It follows that $|B_1| = |B_2|$. □

B.2 Results from set theory

In the above proof we used some results about infinite sets. We state them here without proof. The interested reader should consult an introductory textbook on set theory.

Lemma B.6

Suppose that X and Y are infinite sets and $f : X \to Y$ is a function. Then

1. $|F(Y)| = |Y|$ (where $F(Y)$ is the set of all *finite* subsets of Y)
2. If $|X| > |Y|$, then there exists an element $y \in Y$ whose preimage, $\{x \in X \mid f(x) = y\}$, is infinite. □

Zorn's Lemma is a fundamental statement in the theory of infinite sets, and is equivalent to the Axiom of Choice. In order to state Zorn's Lemma (in the form we use) we make the following definition. If S is a collection of sets, a non-empty subset $\mathcal{C} \subseteq S$ is called a **chain** if

$$\forall\ A, B \in \mathcal{C} \quad \text{either} \quad A \subseteq B \quad \text{or} \quad B \subseteq A$$

Zorn's Lemma

Let S be a non-empty collection of sets. Suppose that whenever $\mathcal{C}$ is a chain in S the union $\cup_{C \in \mathcal{C}} C$ is also an element of S. Then S contains a maximal element. □