MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 1 & 2

Tuesday, September 1 – Wednesday September 2

One of the main objectives of the course

To introduce and study different Algebraic Structures!

Most important algebraic structures for us this term:

- Vector Spaces
- Fields
- Rings

Some Set Theory Notation

- Let S be a set, let a be an element.

$a \in S$	means a is contained in S,
	or equivalently a is an element of S.
$a \notin S$	means a is not contained in S
	(a is not an element of S).

- If S_1, S_2 are two sets, then

$S_1 \subseteq S_2$	means S_1 is a subset of S_2,
	that is, every element of S_1 is also an element of S_2

(note that S_2 may contain other elements as well, but it certainly contains all the elements of S_1)

- $S_1 \times S_2$ is the Cartesian product of S_1 and S_2, that is, the set of all ordered pairs (a, b) with $a \in S_1$, $b \in S_2$.

Some very important and well-known examples of sets

- $\mathbb{N}_0 = \{0, 1, 2, 3, 4, 5, 6, 7, \ldots\}$ is the set of natural (or whole) numbers
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$ the set of integers
- $\mathbb{Q}$ the set of rational numbers:

$$\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0 \right\}$$

- $\mathbb{R}$ the set of real numbers

Recall that we have

$$\mathbb{N}_0 \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

Definition of the notion of 'field'

Definition. A *field* $\mathbb{F}$ is a set (i.e. collection) of elements (i.e. "mathematical" objects) together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x,y) \in \mathbb{F}\times\mathbb{F} \mapsto x+y \in \mathbb{F} \\ \textit{multiplication} \qquad & (x,y) \in \mathbb{F}\times\mathbb{F} \mapsto x\cdot y \in \mathbb{F} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathbb{F}$, $\quad x+y = y+x$ *(commutativity)*

(ii) for all $x, y, z \in \mathbb{F}$, $\quad (x+y)+z = x+(y+z)$ *(associativity)*

(iii) there exists an element 0 in $\mathbb{F}$ such that

$$\text{for all } x \in \mathbb{F}, \quad 0+x = x+0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathbb{F}$, there exists an element $w = w_x$ such that

$$x+w = w+x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

Definition of the notion of 'field' (cont.)

(i') for all $x, y \in \mathbb{F}$, $\quad x \cdot y = y \cdot x$ *(commutativity)*

(ii') for all $x, y, z \in \mathbb{F}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathbb{F}$, **which is different from the zero element** 0 **guaranteed by property (iii) above**, such that

$$\text{for all } x \in \mathbb{F}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

(iv') for every $x \in \mathbb{F}$, **if** $x \neq 0$, then there exists an element $u = u_x$ such that

$$x \cdot u = u \cdot x = 1$$

(reciprocal element, or multiplicative inverse) usually denoted by $1/x$ or by x^{-1}

(ix) for all $x, y, z \in \mathbb{F}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$

$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

The last property, called the Distributive Law, relates the operations of addition and multiplication in $\mathbb{F}$.

Does the definition seem "scary" or bizarre? Actually we encounter/work with such structures all the time, and you have seen them in school or unsuspectingly used their properties in everyday applications!

Very Important Examples:

- The set $\mathbb{R}$ of real numbers is a field.
- The set $\mathbb{Q}$ of rational numbers is a field.
- The set $\mathbb{C}$ of complex numbers is a field.
  (will define this properly very soon)

Furthermore, we will see several more examples in this course.

Note finally that neither $\mathbb{N}_0$ or $\mathbb{Z}$ are fields.
Question: How could we justify this statement?

Some comments on the definition

Note that all the claims below can be derived from the properties/axioms we already stated, we don't need to add anything else to how these axioms are stated.

- The element whose existence is guaranteed by property (iii) (neutral element of addition) is **unique**, that is, there is only one element in $\mathbb{F}$ that satisfies the conclusion of the property.

  This allows us to call this element **the** neutral element of addition in $\mathbb{F}$, and to set aside a special symbol for it, the symbol 0.

- For every $x \in \mathbb{F}$, the element whose existence is guaranteed by property (iv) (additive inverse of x) is **unique**. In other words, if we have $w_1, w_2 \in \mathbb{F}$ such that

$$\begin{aligned} & x + w_1 = w_1 + x = 0 \\ \text{and similarly} \quad & x + w_2 = w_2 + x = 0, \end{aligned}$$

  then $w_1 = w_2$.

  This allows us to call this element **the** negative element of x (or **the** additive inverse of x), and to denote it in a special way, as $-x$.

Some comments on the definition (cont.)

- The element whose existence is guaranteed by property (iii') (neutral element of multiplication) is **unique**, that is, there is only one element in $\mathbb{F}$ that satisfies the conclusion of the property.

  This allows us to call this element **the** neutral element of multiplication in $\mathbb{F}$, and to set aside a special symbol for it, the symbol 1.

- For every $x \in \mathbb{F}$, $x \neq 0$, the element whose existence is guaranteed by property (iv') (multiplicative inverse of x) is **unique**.

  This allows us to call this element **the** reciprocal element of x (or **the** multiplicative inverse of x), and to be able to denote it in a special way, as $\frac{1}{x}$ or as x^{-1}.

- Given that we stipulate that $0 \neq 1$, every field has **at least** two elements! We will soon see that there is a field which has exactly two elements (and then necessarily one of them will be the neutral element of addition, while the other element will be the neutral element of multiplication).

Question: Is it enough to single out these properties/axioms? We know e.g. that in $\mathbb{R}$ more properties and identities hold true.

Answer: YES, many other properties (especially all the properties we will be interested in in this course) can be derived from the above.

Let's see one such derivation:

Proposition. Suppose $\mathbb{F}$ is a field, x, y elements of $\mathbb{F}$, and let 0 be the zero element of $\mathbb{F}$. Then

(i) $0 \cdot x = 0$.

(ii) If $x \cdot y = 0$, then $x = 0$ or $y = 0$.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 3

Friday September 4

Reminder: Definition of field

Definition. A *field* $\mathbb{F}$ is a set (i.e. collection) of elements (i.e. "mathematical" objects) together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x, y) \in \mathbb{F} \times \mathbb{F} \mapsto x + y \in \mathbb{F} \\ \textit{multiplication} \qquad & (x, y) \in \mathbb{F} \times \mathbb{F} \mapsto x \cdot y \in \mathbb{F} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathbb{F}$, $\quad x + y = y + x$ *(commutativity)*

(ii) for all $x, y, z \in \mathbb{F}$, $\quad (x + y) + z = x + (y + z)$ *(associativity)*

(iii) there exists an element 0 in $\mathbb{F}$ such that

$$\text{for all } x \in \mathbb{F}, \quad 0 + x = x + 0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathbb{F}$, there exists an element $w = w_x$ such that

$$x + w = w + x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

(i') for all $x, y \in \mathbb{F}$, $\quad x \cdot y = y \cdot x$ *(commutativity)*

(ii') for all $x, y, z \in \mathbb{F}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathbb{F}$, **which is different from the zero element** 0 **guaranteed by property (iii) above**, such that

$$\text{for all } x \in \mathbb{F}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

(iv') for every $x \in \mathbb{F}$, **if** $x \neq 0$, then there exists an element $u = u_x$ such that

$$x \cdot u = u \cdot x = 1$$

(reciprocal element, or multiplicative inverse) usually denoted by $1/x$ or by x^{-1}

(ix) for all $x, y, z \in \mathbb{F}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

The last property, called the Distributive Law, relates the operations of addition and multiplication in $\mathbb{F}$.

Recall: Important Examples

- The set $\mathbb{R}$ of real numbers with the standard operations of addition and multiplication is a field.
- The set $\mathbb{Q}$ of rational numbers with the standard operations of addition and multiplication is a field.
- The set $\mathbb{C}$ of complex numbers with the standard operations of addition and multiplication, which we will recall/introduce very soon, is a field.

Thus, for example, for all $x, y, z \in \mathbb{R}$, we have that

$$x + y = y + x$$

$$\text{and} \quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$$

$$\text{and} \quad y \cdot (x + z) = y \cdot x + y \cdot z.$$

We will be citing and using these properties throughout the term.

Question: Is it enough to single out these properties/axioms? We know e.g. that in $\mathbb{R}$ more properties and identities hold true.

Answer: YES, many other properties (especially all the properties we will be interested in in this course) can be derived from the above.

Let's see one such derivation:

Proposition. Suppose $\mathbb{F}$ is a field, x, y elements of $\mathbb{F}$, and let 0 be the zero element of $\mathbb{F}$. Then

(i) $0 \cdot x = 0$.

(ii) If $x \cdot y = 0$, then $x = 0$ or $y = 0$.

Proof of the Proposition

Part (i). We can write

$$\begin{aligned} 0 \cdot x + x &= 0 \cdot x + 1 \cdot x && \text{(multiplicative identity)} \\ &= (0 + 1) \cdot x && \text{(right distributive property)} \\ &= 1 \cdot x && \text{(additive identity)} \\ &= x\,. && \text{(multiplicative identity)} \end{aligned}$$

Therefore we have $0 \cdot x + x = x$.
But then

$$\begin{aligned} 0 \cdot x &= 0 \cdot x + 0 && \text{(additive identity)} \\ &= 0 \cdot x + (x + (-x)) && \text{(additive inverse)} \\ &= (0 \cdot x + x) + (-x) && \text{(associativity of addition)} \\ &= x + (-x) && \text{(by what we showed above)} \\ &= 0\,. && \text{(additive inverse)} \end{aligned}$$

Note that this is one of several ways we can prove Part (i), so if you have an idea about an alternative justification don't dismiss it just because it doesn't begin exactly like in this proof (of course be careful that you are using the axioms correctly).

Proof of the Proposition (cont.)

Part (ii). Assume that we have $x \cdot y = 0$.
There are two cases/possibilities for x.

Case 1: $x = 0$. Then we already have the conclusion.

Case 2: $x \neq 0$. Then, by Property (iv'), x has a multiplicative inverse (and recall that we denote this by x^{-1}). We can write

$$\begin{aligned} y &= 1 \cdot y \\ &= (x^{-1} \cdot x) \cdot y \\ &= x^{-1} \cdot (x \cdot y) \\ &= x^{-1} \cdot 0 && \text{(by assumption above)} \\ &= 0 \cdot x^{-1} \\ &= 0\,. && \text{(by Part (i))} \end{aligned}$$

So here we showed that, if $x \neq 0$, then necessarily $y = 0$.

In the end, we do have either $x = 0$ or $y = 0$ (or both), as we wanted.

Practice. Write above which properties/axioms are used to go from one line to the next.

More examples of fields

The following problem is similar to homework and exam problems from last year.

Problem 1. Consider $\mathbb{R}^2$ with operations of addition and multiplication defined as follows: for every two ordered pairs $(a, b), (c, d) \in \mathbb{R}^2$,

$$(a, b) + (c, d) \stackrel{\text{def}}{=} (a + c, b + d)$$

$$\text{and} \quad (a, b) \cdot (c, d) \stackrel{\text{def}}{=} (ac - bd,\ ad + bc).$$

Show that the triple $\left(\mathbb{R}^2, +, \cdot\right)$ is a field.

Clarification. Here the operations on the right-hand sides are the standard operations in $\mathbb{R}$. Also $ac - bd$ is the same as $ac + (-bd)$.

More examples of fields

Problem 1. Consider $\mathbb{R}^2$ with operations of addition and multiplication defined as follows: for every two ordered pairs $(a, b), (c, d) \in \mathbb{R}^2$,

$$(a, b) + (c, d) \stackrel{\text{def}}{=} (a + c, b + d)$$

$$\text{and} \quad (a, b) \cdot (c, d) \stackrel{\text{def}}{=} (ac - bd,\ ad + bc).$$

Show that the triple $\left(\mathbb{R}^2, +, \cdot\right)$ is a field.

Remark. Observe that the given operations are of the type we want: that is, for every two ordered pairs $(a, b), (c, d) \in \mathbb{R}^2$, we have by definition that $a, b, c, d \in \mathbb{R}$

$$\text{and therefore} \quad a + c \in \mathbb{R} \quad \text{and} \quad b + d \in \mathbb{R}$$

$$\text{and therefore} \quad (a + c, b + d) \in \mathbb{R}^2.$$

Similarly, given that $a, b, c, d \in \mathbb{R}$, we have that

$$ac, ad, bd, bc \in \mathbb{R}, \quad \text{and therefore} \quad ac - bd \in \mathbb{R} \quad \text{and} \quad ad + bc \in \mathbb{R}$$

$$\text{and therefore} \quad (ac - bd,\ ad + bc) \in \mathbb{R}^2.$$

By the way, such a way of defining the addition

of ordered pairs is called **coordinate-wise**.

So the triple $\left(\mathbb{R}^2, +, \cdot\right)$ is of the type required in the definition of 'field', and hence, to show that it is a field, what we need to do is check that **all 9 properties/axioms** are satisfied.

In other words, we need to verify that

- **Addition is commutative**
- **Addition is associative**
- **There exists a neutral element of addition**
- **For every $(a, b) \in \mathbb{R}^2$, there exists an additive inverse**
- **Multiplication is commutative**
- **Multiplication is associative**
- **There exists a neutral element of multiplication**
- **For every $(a, b) \in \mathbb{R}^2$ which is different from the additive identity, there exists a multiplicative inverse**
- **Distributive Law** (recall that, because of the commutativity of multiplication, it suffices to check only the right distributive property or only the left distributive property)

Proof

Addition is commutative

Before we start, a comment: we need to show that a certain identity holds for every two elements $(a, b), (c, d)$ of $\mathbb{R}^2$. Clearly it's impossible to prove this just by plugging different real values for a, b, c, d and checking whether the identity holds with those specific values.

Instead, we consider (what is called) arbitrary (a, b) and (c, d) from $\mathbb{R}^2$, which means we don't make any assumption about them, and check the identity for those.
[Not making any assumption about them means that we don't e.g. say that a has to be positive, or that b has to be an integer, or that c and d are irrational numbers;
we are **NOT** allowed to make such assumptions, since we want to prove the identity for every two $(a, b), (c, d) \in \mathbb{R}^2$;

of course if any additional assumptions have been given to us
in the statement of the problem, we can use those.]

Proof

Addition is commutative

Let (a, b) and (c, d) be (arbitrary) elements of $\mathbb{R}^2$. We need to show that

$$(a, b) + (c, d) = (c, d) + (a, b).$$

We have that

$$(a, b) + (c, d) = (a + c, b + d)$$
$$\text{and} \quad (c, d) + (a, b) = (c + a, d + b).$$

But because addition in $\mathbb{R}$ is commutative, we have that $a + c = c + a$ and $b + d = d + b$. Therefore,

$$(a + c, b + d) = (c + a, d + b)$$

(since the corresponding components of the ordered pairs are equal), and this gives the desired conclusion.

Addition is associative *Left as an exercise.*

Proof (cont.)

There exists a neutral element of addition

First we need to guess what this element should look like!
Let's make the guess that the neutral element should be $(0_{\mathbb{R}}, 0_{\mathbb{R}})$, where $0_{\mathbb{R}}$ is the neutral element of addition in $\mathbb{R}$, and let's try to confirm this guess.

Again, we are asked to prove that an identity holds for every $(a, b) \in \mathbb{R}^2$, so we start our argument with:

Let (a, b) be an element of $\mathbb{R}^2$. Then we have that

$$(a, b) + (0_{\mathbb{R}}, 0_{\mathbb{R}}) = (a + 0_{\mathbb{R}}, b + 0_{\mathbb{R}}) = (a, b)$$
$$\textit{and} \quad (0_{\mathbb{R}}, 0_{\mathbb{R}}) + (a, b) = (0_{\mathbb{R}} + a, 0_{\mathbb{R}} + b) = (a, b).$$

For every $(a, b) \in \mathbb{R}^2$, there exists an additive inverse

Consider an (arbitrary) element (a, b) of $\mathbb{R}^2$. We need to show that (a, b) has an additive inverse in $\mathbb{R}^2$.

First we need to guess what this element should look like!
Plausible guess? The element $(-a, -b)$, where $-a$ is the additive inverse of a in $\mathbb{R}$ and $-b$ the additive inverse of b.

Exercise: Finish the argument here.

Proof (cont.)

Multiplication is commutative *Left as an exercise.*

Multiplication is associative

Proof (cont.)

Multiplication is associative

Let $(a, b), (c, d)$ and (e, f) be elements of $\mathbb{R}^2$. We need to show that

$$\big((a, b) \cdot (c, d)\big) \cdot (e, f) = (a, b) \cdot \big((c, d) \cdot (e, f)\big).$$

We have that

$$\begin{aligned}\big((a, b) \cdot (c, d)\big) \cdot (e, f) &= (ac - bd, ad + bc) \cdot (e, f) \\ &= \big((ac - bd)e - (ad + bc)f,\, (ac - bd)f + (ad + bc)e\big) \\ &= \big(ace - bde - adf - bcf,\, acf - bdf + ade + bce\big)\end{aligned}$$

(because in $\mathbb{R}$ we already know we can omit parentheses and we also have the distributive law), and similarly

$$\begin{aligned}(a, b) \cdot \big((c, d) \cdot (e, f)\big) &= (a, b) \cdot (ce - df, cf + de) \\ &= \big(a(ce - df) - b(cf + de),\, a(cf + de) + b(ce - df)\big) \\ &= \big(ace - adf - bcf - bde,\, acf + ade + bce - bdf\big).\end{aligned}$$

Since the addition in $\mathbb{R}$ is commutative, we finally have

$$ace - bde - adf - bcf = ace - adf - bcf - bde$$

$$\text{and} \quad acf - bdf + ade + bce = acf + ade + bce - bdf,$$

which gives the desired conclusion.

Proof (cont.)

There exists a neutral element of multiplication

First we need to guess what this element should look like!
Here the guess probably cannot be made very quickly, simply by inspection. We want an element $(m, n) \in \mathbb{R}^2$ such that,

$$\text{for all } (a, b) \in \mathbb{R}^2, \qquad (a, b) \cdot (m, n) = (a, b),$$

or in other words, given our definition for multiplication, we need

$$am - bn = a \quad \text{and} \quad an + bm = b.$$

Since we need these to hold for every $(a, b) \in \mathbb{R}^2$, we can first try plugging in some (a, b) of simple enough form to see what this gives us: e.g. try with the elements $\left(a, 0_{\mathbb{R}}\right)$ in $\mathbb{R}^2$.
In such cases, we need to have $am = a$ and $an = 0_{\mathbb{R}}$. These would be satisfied if we had

$$(m, n) = \left(1_{\mathbb{R}}, 0_{\mathbb{R}}\right).$$

Exercise: Finish the argument by verifying that $\left(1_{\mathbb{R}}, 0_{\mathbb{R}}\right)$ is the multiplicative identity.

Proof (cont.)

For every $(a, b) \in \mathbb{R}^2$ with $(a, b) \neq (0_{\mathbb{R}}, 0_{\mathbb{R}})$, there exists a multiplicative inverse

Consider an (arbitrary) element (a, b) of $\mathbb{R}^2$ which is different from $(0_{\mathbb{R}}, 0_{\mathbb{R}})$ (what does the latter mean? That at least one of the components a, b of (a, b) is $\neq 0_{\mathbb{R}}$; note that we don't need both of them to be non-zero, although we could also have both non-zero).
We need to show that this (a, b) has a multiplicative inverse in $\mathbb{R}^2$.

First we need to guess what this element should look like!
Here the guess cannot be made quickly, simply by inspection. Try to analyse what you would need c, d to satisfy in order for the equation

$$(a, b) \cdot (c, d) = (1_{\mathbb{R}}, 0_{\mathbb{R}})$$

to hold true; left as an exercise for now, will discuss it on Tuesday next week.

Finally

Distributive Law *Left as an exercise.*

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 4

Tuesday September 8

Reminder: Problem from last time

Problem 1. Consider $\mathbb{R}^2$ with operations of addition and multiplication defined as follows: for every two ordered pairs $(a, b), (c, d) \in \mathbb{R}^2$,

$$(a, b) + (c, d) \stackrel{\text{def}}{=} (a + c, b + d)$$

$$\text{and} \quad (a, b) \cdot (c, d) \stackrel{\text{def}}{=} (ac - bd,\ ad + bc).$$

Show that the triple $\left(\mathbb{R}^2, +, \cdot\right)$ is a field.

Clarification. Here the operations on the right-hand sides are the standard operations in $\mathbb{R}$. Also $ac - bd$ is the same as $ac + (-bd)$.

Last time we observed that the triple $\left(\mathbb{R}^2, +, \cdot\right)$ is of the type required in the definition of 'field', and hence, to show that it is a field, what remains to do is to check that **all 9 properties/axioms** are satisfied.

In other words, we need to verify that

- **Addition is commutative**
- **Addition is associative**
- **There exists a neutral element of addition**
- **For every $(a, b) \in \mathbb{R}^2$, there exists an additive inverse**
- **Multiplication is commutative**
- **Multiplication is associative**
- **There exists a neutral element of multiplication**
- **For every $(a, b) \in \mathbb{R}^2$ which is different from the additive identity, there exists a multiplicative inverse**
- **Distributive Law** (recall that, because of the commutativity of multiplication, it suffices to check only the right distributive property or only the left distributive property)

Proof

Addition is commutative

Let (a, b) and (c, d) be (arbitrary) elements of $\mathbb{R}^2$. We need to show that

$$(a, b) + (c, d) = (c, d) + (a, b).$$

We have that

$$(a, b) + (c, d) = (a + c, b + d)$$
$$\text{and} \quad (c, d) + (a, b) = (c + a, d + b).$$

But because addition in $\mathbb{R}$ is commutative, we have that $a + c = c + a$ and $b + d = d + b$. Therefore,

$$(a + c, b + d) = (c + a, d + b)$$

(since the corresponding components of the ordered pairs are equal), and this gives the desired conclusion.

Addition is associative *Left as an exercise.*

Proof (cont.)

There exists a neutral element of addition

First we need to guess what this element should look like!
Let's make the guess that the neutral element should be $(0_{\mathbb{R}}, 0_{\mathbb{R}})$, where $0_{\mathbb{R}}$ is the neutral element of addition in $\mathbb{R}$, and let's try to confirm this guess.

Again, we are asked to prove that an identity holds for every $(a, b) \in \mathbb{R}^2$, so we start our argument with:

Let (a, b) be an element of $\mathbb{R}^2$. Then we have that

$$(a, b) + (0_{\mathbb{R}}, 0_{\mathbb{R}}) = (a + 0_{\mathbb{R}}, b + 0_{\mathbb{R}}) = (a, b)$$
$$\textit{and} \quad (0_{\mathbb{R}}, 0_{\mathbb{R}}) + (a, b) = (0_{\mathbb{R}} + a, 0_{\mathbb{R}} + b) = (a, b).$$

For every $(a, b) \in \mathbb{R}^2$, there exists an additive inverse

Consider an (arbitrary) element (a, b) of $\mathbb{R}^2$. We need to show that (a, b) has an additive inverse in $\mathbb{R}^2$.

First we need to guess what this element should look like!
Plausible guess? The element $(-a, -b)$, where $-a$ is the additive inverse of a in $\mathbb{R}$ and $-b$ the additive inverse of b.

Exercise: Finish the argument here.

Proof (cont.)

Multiplication is commutative *Left as an exercise.*

Multiplication is associative

Proof (cont.)

Multiplication is associative

Let $(a, b), (c, d)$ and (e, f) be elements of $\mathbb{R}^2$. We need to show that

$$\big((a, b) \cdot (c, d)\big) \cdot (e, f) = (a, b) \cdot \big((c, d) \cdot (e, f)\big).$$

We have that

$$\begin{aligned}\big((a, b) \cdot (c, d)\big) \cdot (e, f) &= (ac - bd, ad + bc) \cdot (e, f) \\ &= \big((ac - bd)e - (ad + bc)f,\, (ac - bd)f + (ad + bc)e\big) \\ &= \big(ace - bde - adf - bcf,\, acf - bdf + ade + bce\big)\end{aligned}$$

(because in $\mathbb{R}$ we already know we can omit parentheses and we also have the distributive law), and similarly

$$\begin{aligned}(a, b) \cdot \big((c, d) \cdot (e, f)\big) &= (a, b) \cdot (ce - df, cf + de) \\ &= \big(a(ce - df) - b(cf + de),\, a(cf + de) + b(ce - df)\big) \\ &= \big(ace - adf - bcf - bde,\, acf + ade + bce - bdf\big).\end{aligned}$$

Since the addition in $\mathbb{R}$ is commutative, we finally have

$$ace - bde - adf - bcf = ace - adf - bcf - bde$$

$$\text{and} \quad acf - bdf + ade + bce = acf + ade + bce - bdf,$$

which gives the desired conclusion.

Proof (cont.)

There exists a neutral element of multiplication

First we need to guess what this element should look like!
Here the guess probably cannot be made very quickly, simply by inspection. We want an element $(m, n) \in \mathbb{R}^2$ such that,

$$\text{for all } (a, b) \in \mathbb{R}^2, \qquad (a, b) \cdot (m, n) = (a, b),$$

or in other words, given our definition for multiplication, we need

$$am - bn = a \quad \text{and} \quad an + bm = b.$$

Since we need these to hold for every $(a, b) \in \mathbb{R}^2$, we can first try plugging in some (a, b) of simple enough form to see what this gives us: e.g. try with the elements $\left(a, 0_{\mathbb{R}}\right)$ in $\mathbb{R}^2$.
In such cases, we need to have $am = a$ and $an = 0_{\mathbb{R}}$. These would be satisfied if we had

$$(m, n) = \left(1_{\mathbb{R}}, 0_{\mathbb{R}}\right).$$

Exercise: Finish the argument by verifying that $\left(1_{\mathbb{R}}, 0_{\mathbb{R}}\right)$ is the multiplicative identity.

Proof (cont.)

For every $(a, b) \in \mathbb{R}^2$ with $(a, b) \neq (0_{\mathbb{R}}, 0_{\mathbb{R}})$, there exists a multiplicative inverse

Consider an (arbitrary) element (a, b) of $\mathbb{R}^2$ which is different from $(0_{\mathbb{R}}, 0_{\mathbb{R}})$ (what does the latter mean? That at least one of the components a, b of (a, b) is $\neq 0_{\mathbb{R}}$; note that we don't need both of them to be non-zero, although we could also have both non-zero).
We need to show that this (a, b) has a multiplicative inverse in $\mathbb{R}^2$.

First we need to guess what this element should look like!
Here the guess cannot be made quickly, simply by inspection. Try to analyse what you would need c, d to satisfy in order for the equation

$$(a, b) \cdot (c, d) = (1_{\mathbb{R}}, 0_{\mathbb{R}})$$

to hold true; it would lead you to the following system of two linear equations:

$$\left\{ \begin{array}{rcl} ac - bc & = & 1 \\ ad + bc & = & 0 \end{array} \right\} \quad \Leftrightarrow \quad \left\{ \begin{array}{rcl} ac - bd & = & 1 \\ bc + ad & = & 0 \end{array} \right\}$$

in the unknown c, d (where the coefficients are the real numbers a, b, which we assume we have fixed, and thus they are considered known numbers to us).

Proof (cont.)

For every $(a, b) \in \mathbb{R}^2$ with $(a, b) \neq (0_{\mathbb{R}}, 0_{\mathbb{R}})$, there exists a multiplicative inverse

We've just seen that

$$(a, b) \cdot (c, d) = (1_{\mathbb{R}}, 0_{\mathbb{R}})$$

is equivalent to

$$\left\{ \begin{array}{rcl} ac - bc & = & 1 \\ ad + bc & = & 0 \end{array} \right\} \quad \Leftrightarrow \quad \left\{ \begin{array}{rcl} ac - bd & = & 1 \\ bc + ad & = & 0 \end{array} \right\}$$

Note also that the key assumption we have made about a, b is that at least one of them is non-zero; this is the assumption that allows us to find a solution to the above system, that is, to find expressions for c and d in terms of the numbers a, b that will make both equations of the system true at the same time (how do we do this?).

Proof (cont.)

We can consider two main cases:

Case 1: $a \neq 0$. Then we can solve for d in the second equation, and express d as a function of c (and of course in terms of the 'coefficients' a, b): we get that

$$d = -\frac{bc}{a}\,.$$

We can then plug this into the first equation to obtain that

$$ac - b \cdot (-\frac{bc}{a}) = 1 \quad \Rightarrow \quad ac + \frac{b^2 c}{a} = 1$$

$$\Rightarrow \quad \frac{a^2 c + b^2 c}{a} = 1 \quad \Rightarrow \quad c = \frac{a}{a^2 + b^2}\,,$$

with the last implication making sense because $\frac{a^2+b^2}{a} \neq 0$ since $a \neq 0$, and hence $a^2 + b^2 \neq 0$ too, and so we can divide by it.

Plugging this back into $d = -(bc)/a$, we also get

$$d = -\frac{b}{a^2 + b^2}\,.$$

We conclude that, in this case, a candidate for the multiplicative inverse of (a, b) is the ordered pair

$$(c, d) = \left(\frac{a}{a^2 + b^2},\ -\frac{b}{a^2 + b^2}\right)\,.$$

Proof (cont.)

We can consider two main cases:

Case 2: $a = 0$. Then the linear system we ended up with becomes simpler:

$$\left\{ \begin{array}{rcl} -bd & = & 1 \\ bc & = & 0 \end{array} \right\}.$$

Also, in this case, we certainly have $b \neq 0$, and hence we can solve for c and d in the above equations: we obtain that

$$d = -\frac{1}{b} \quad \text{which can be rewritten in this case as} \quad -\frac{b}{b^2} = -\frac{b}{a^2+b^2},$$

$$\text{and also } c = 0 \quad \text{which can be rewritten in this case as} \quad \frac{a}{a^2+b^2}.$$

Proof (cont.)

We can consider two main cases:

Case 2: $a = 0$. Then the linear system we ended up with becomes simpler:

$$\left\{ \begin{array}{rcl} -bd & = & 1 \\ bc & = & 0 \end{array} \right\}.$$

Also, in this case, we certainly have $b \neq 0$, and hence we can solve for c and d in the above equations: we obtain that

$$d = -\frac{1}{b} \quad \text{which can be rewritten in this case as} \quad -\frac{b}{b^2} = -\frac{b}{a^2+b^2},$$

$$\text{and also } c = 0 \quad \text{which can be rewritten in this case as} \quad \frac{a}{a^2+b^2}.$$

We conclude that, in all cases, a candidate for the multiplicative inverse of (a, b) is the ordered pair

$$(c, d) = \left(\frac{a}{a^2+b^2}, \; -\frac{b}{a^2+b^2} \right).$$

Exercise: Finish the argument by verifying that this is indeed the multiplicative inverse of (a, b).

Proof (cont.)

Finally, to finish the proof, it remains to verify the

Distributive Law *Left as an exercise.*

The equation $ax = b$ when the coefficients a, b come from a field

One of the advantages of working in a field $\mathbb{F}$, that is, doing computations and trying to solve different equations with coefficients coming from $\mathbb{F}$, is that we can develop very general criteria about when we will be able to successfully find solutions: in a certain sense, we can say that, **unless there's a trivial reason why substituting in the equation would lead to absurdities**, we should be able to find solutions (note that there isn't such a nice "dichotomy" in other structures).

Focusing for now on the particular case of

a linear equation in one unknown x with coefficients from $\mathbb{F}$, which has the general form $ax = b$,

what is understood here is that x is a variable that can take different values in $\mathbb{F}$, with this leading to different mathematical statements of equality which are either true or false; then "solving the equation" means finding precisely all the values that would make the equality true.

The equation $ax = b$ when the coefficients a, b come from a field

In trying to solve this, let us distinguish three main cases:

Case 1: $a = b = 0$. Then (based on the Proposition we proved last time), **no matter what element of $\mathbb{F}$ we set x equal to**, the equality will hold true.

Case 2: $a = 0,\ b \neq 0$. Then (again based on the Proposition we proved last time), no matter what element of $\mathbb{F}$ we set x equal to, the LHS will equal 0, while the RHS will be non-zero $\rightsquigarrow$ **absurd!** thus we have no solutions

Case 3: $a \neq 0$. Then, no matter what the exact value of a is, and no matter what b is, we can find a **unique solution**. How? Note that $a \neq 0$ implies that a^{-1} exists. But then, we can multiply both sides of the equation by a^{-1}, and also use the associativity of the multiplication, to obtain that

$$x = (a^{-1}a)x = a^{-1} \cdot (ax) = a^{-1} \cdot b.$$

Note that this tells us that the only value that would make the equality true is the element $a^{-1} \cdot b$ (and also that this value is a solution, since $a \cdot (a^{-1}b) = b$ as we wanted).

What next?

One of the main topics of this term is to generalise the approach here, namely the idea to distinguish into three main cases (and also the type of main cases that we chose to consider), in order to figure out when we can find solutions to multiple linear equations that we would like to have satisfied simultaneously.

We call such a collection of linear equations (which can be in multiple unknowns as well) *a system of linear equations:*

$$\text{e.g.} \qquad \left\{ \begin{array}{rcl} 3x_1 - 6x_2 + 7x_3 & = & 0 \\ -x_1 + 4x_2 - 3x_3 & = & -2 \\ 2x_1 + 0x_2 + 4x_3 & = & 1 \\ 7x_1 - 6x_2 - 12x_3 & = & \frac{8}{3} \end{array} \right\},$$

and again we will be studying examples where the coefficients belong to a field $\mathbb{F}$ (for reasons very similar to what we saw above; in this example, let's say the coefficients come from $\mathbb{R}$).

What next? (cont.)

Instead of trying to satisfy multiple linear equations simultaneously, we could also try solving **polynomial equations of higher order**.

Simplest case to start with: stick to considering one unknown, but allow the LHS of the equation to be higher-degree polynomials too (that is, not just linear polynomials).

E.g. $3x^6 - 7x^3 + 2 = -3$, or $5x^{11} = -4$, and so on.

Here we can get stuck very quickly: recall that e.g. in $\mathbb{R}$ the equations $x^2 = b$ have no solution if b is a negative real number.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 5

Wednesday September 9

The equation $ax = b$ when the coefficients a, b come from a field

One of the main topics of this course, and one of the most useful applications of Linear Algebra (and motivation for its development), is studying how we can solve systems of equations, and in particular **linear equations**, whose coefficients belong to a field $\mathbb{F}$. We will see that we can give very general criteria about when we will be able to successfully find solutions (and also that there is an (algorithmic!) approach to finding these solutions when they exist).

Simplest case, and the case we have to start with:

one linear equation in one unknown x with coefficients from $\mathbb{F}$, which has the general form $ax = b$.

What is understood here is that x is a variable that can take different values in $\mathbb{F}$, with this leading to different mathematical statements of equality which are either true or false.

'Solving the equation' means finding precisely all the values that would make the equality true.

The equation $ax = b$ when the coefficients a, b come from a field

In trying to solve this, let us distinguish three main cases:

Case 1: $a = b = 0$. Then (based on the Proposition we proved last time), **no matter what element of $\mathbb{F}$ we set x equal to**, the equality will hold true.

Case 2: $a = 0,\ b \neq 0$. Then (again based on the Proposition we proved last time), no matter what element of $\mathbb{F}$ we set x equal to, the LHS will equal 0, while the RHS will be non-zero $\rightsquigarrow$ **absurd!** thus we have no solutions

Case 3: $a \neq 0$. Then, no matter what the exact value of a is, and no matter what b is, we can find a **unique solution**. How? Note that $a \neq 0$ implies that a^{-1} exists. But then, we can multiply both sides of the equation by a^{-1}, and also use the associativity of the multiplication, to obtain that

$$x = (a^{-1}a)x = a^{-1} \cdot (ax) = a^{-1} \cdot b.$$

Note that this tells us that the only value that would make the equality true is the element $a^{-1} \cdot b$ (and also that this value is a solution, since $a \cdot (a^{-1}b) = b$ as we wanted).

What next?

We will see that this approach generalises very nicely when we try to find solutions to multiple linear equations (in multiple unknowns) that we would like to have satisfied simultaneously.

As mentioned before, such a collection of linear equations is called *a system of linear equations:*

e.g. $$\left\{\begin{array}{rcl} 3x_1 - 6x_2 + 7x_3 &=& 0 \\ -x_1 + 4x_2 - 3x_3 &=& -2 \\ 2x_1 + 0x_2 + 4x_3 &=& 1 \\ 7x_1 - 6x_2 - 12x_3 &=& \frac{8}{3} \end{array}\right\}, \quad \text{or} \quad \left\{\begin{array}{rcl} ax_1 - bx_2 &=& 1 \\ bx_1 + ax_2 &=& 0 \end{array}\right\},$$

and again we will be studying examples where the coefficients belong to a field $\mathbb{F}$ (for reasons very similar to what we saw above; in these examples, let's say the coefficients come from $\mathbb{R}$).

What about polynomial equations?

Instead of trying to satisfy multiple linear equations simultaneously, we could also try solving **polynomial equations of higher order**.

Simplest case to start with: stick to considering one unknown, but allow the LHS of the equation to be higher-degree polynomials too (that is, not just linear polynomials).

E.g. $3x^6 - 7x^3 + 2 = -3$, or $5x^{11} = -4$, and so on.

Here we can get stuck very quickly: recall that e.g. in $\mathbb{R}$ the equations $x^2 = b$ have no solution if b is a negative real number.

Reminders about polynomials

For now we will discuss real polynomials.

A **polynomial function from $\mathbb{R}$ to $\mathbb{R}$** is a function p that can be written in the following way: there is an integer $m \geqslant 0$, and real numbers $a_0, a_1, \ldots, a_m$ such that

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m.$$

For any k between 0 and m, we call the term $a_k x^k$ the k-th order term of the polynomial p, while the numbers $a_0, a_1, \ldots, a_m$ are called the coefficients of the polynomial.

- If $a_0 = a_1 = \cdots = a_m = 0$, then p is the constant zero function, the zero polynomial.
- On the other hand, if at least one of the coefficients is non-zero, then we can show that the polynomial p has only finitely many **roots**, that is, inputs u from $\mathbb{R}$ such that $p(u) = 0$. In fact, we can further see that p can have at most m roots.

Reminders about polynomials

Definition: Degree of a real polynomial

If $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a non-zero real polynomial, that is, if at least one of the coefficients of p is non-zero, then we can also find the **maximum** k between 0 and m such that $a_k \neq 0$. We then say that the **degree** of p is equal to k.

e.g. $\deg(x+1) = 1$, $\deg(x^{10} - 3x^2 + 2x) = 10$ and $\deg(2) = 0$.

By convention, the degree of the zero polynomial is set to be $-\infty$: $\deg(0) = -\infty$.

So

- Non-zero constant polynomials have degree 0.
- **A non-zero real polynomial p has at most $\deg(p)$ roots in $\mathbb{R}$.**
- Polynomials of degree 1 are also called linear polynomials, those of degree 2 are called quadratic polynomials, and those of degree 3 are called cubic polynomials (clearly we can't go on like that for all higher degrees: we do also have quartic $\leftrightarrow$ degree 4, and quintic $\leftrightarrow$ degree 5, and a few more terms like that, but most of the time we just say "a degree m polynomial", or "a polynomial equation of degree (or of order) m").

Back to the quadratic equation $x^2 = b$

Consider the case where b is a negative real number. Clearly this equation has no solutions in $\mathbb{R}$ (equivalently we can say that the polynomial $p(x) = x^2 - b$ has no roots in $\mathbb{R}$.

Given that we can write $b = -|b| = (-1) \cdot |b|$, where $|b|$ is the absolute value of b,

and given also that the equation $w^2 = |b|$ does have solutions in $\mathbb{R}$ (this is something that you prove in a Calculus/Analysis course),

we conclude that, if we managed to solve the equation $x^2 = -1$, then we would also find solutions for the more general $x^2 = -|b|$.

Let's "artificially" introduce a solution to the equation $x^2 = -1$. That is, we create a new object, which we will denote by $\boldsymbol{i}$, that will satisfy

$$\boldsymbol{i} \cdot \boldsymbol{i} = -1.$$

The set $\mathbb{R} \cup \{\boldsymbol{i}\}$

Now, in this set the equation $x^2 = -1$ has a solution. However we have ruined the field structure of $\mathbb{R}$ (which allowed us to do algebra).

The idea is to try to extend our operations $+$ and $\cdot$ too,

- so we start including every element of the form $b \cdot \boldsymbol{i}$ in our set for all $b \in \mathbb{R}$, thus making the set bigger, (and we consider any such element the same as $\boldsymbol{i} \cdot b$ or more simply $\boldsymbol{i}b$),
- and then we also include elements of the form $a + b\boldsymbol{i}$ (which we consider the same as $b\boldsymbol{i} + a$, or as $a + \boldsymbol{i}b$, and so on) for all $a, b \in \mathbb{R}$.

But all of these (except when $b = 0$) are new elements, so how will addition and multiplication be defined when we want to involve one or more of these new elements?

Operations on the set $\{a + bi : a, b \in \mathbb{R}\}$

Consider two elements from this set, say $a + bi$ and $c + di$. Note that we can also think here that we have **four** elements, the elements a, bi, c and di (since e.g. we included the element $a + bi$ to "play the role" of the result of adding the elements a and bi).

Note also that we would very much like to preserve the nice properties that $+$ has in $\mathbb{R}$!

So we write (the equalities below are what we would like to have, and motivate the definition of $+$ that we will give in the end):

$$\begin{aligned}(a + bi) + (c + di) &= \big((a + bi) + c\big) + di \\ &= \big(a + (bi + c)\big) + di \\ &= \big(a + (c + bi)\big) + di \\ &= \big((a + c) + bi\big) + di \\ &= (a + c) + (bi + di) \\ &= (a + c) + (b + d)i,\end{aligned}$$

with the last expression being again an element of the set $\{a + bi : a, b \in \mathbb{R}\}$.

Operations on the set $\{a + b\boldsymbol{i} : a, b \in \mathbb{R}\}$

This motivates the definition

$(a + b\boldsymbol{i}) + (c + d\boldsymbol{i}) := (a + c) + (b + d)\boldsymbol{i}$.

Operations on the set $\{a + b\boldsymbol{i} : a, b \in \mathbb{R}\}$

Similarly, we would like to be able to write:

$$\begin{aligned}(a+b\boldsymbol{i})\cdot(c+d\boldsymbol{i}) &= (a+b\boldsymbol{i})\cdot c + (a+b\boldsymbol{i})\cdot d\boldsymbol{i}\\ &= \big(a\cdot c + b\boldsymbol{i}\cdot c\big) + \big(a\cdot d\boldsymbol{i} + b\boldsymbol{i}\cdot d\boldsymbol{i}\big)\\ &= \big(ac + b(\boldsymbol{i}c)\big) + \big((ad)\boldsymbol{i} + b\boldsymbol{i}\cdot \boldsymbol{i}d\big)\\ &= \big(ac + b(c\boldsymbol{i})\big) + \big((ad)\boldsymbol{i} + b(\boldsymbol{i}\cdot\boldsymbol{i})d\big)\\ &= \big(ac + (bc)\boldsymbol{i}\big) + \big((ad)\boldsymbol{i} + b(-1)d\big)\end{aligned}$$

(since $\boldsymbol{i}\cdot\boldsymbol{i} = -1$ by how we defined/'created' this new element)

$$\begin{aligned}&= \big(ac + (bc)\boldsymbol{i}\big) + \big((ad)\boldsymbol{i} + (-bd)\big)\\ &= (ac - bd) + (bc + ad)\boldsymbol{i}.\end{aligned}$$

(based on our definition of $+$)

This motivates the definition

$$(a+b\boldsymbol{i})\cdot(c+d\boldsymbol{i}) := (ac - bd) + (ad + bc)\boldsymbol{i}.$$

The set $\mathbb{C}$ of complex numbers

We denote this new set $\{a + b\boldsymbol{i} : a, b \in \mathbb{R}\}$ (which is an extension of $\mathbb{R}$) by $\mathbb{C}$.

Our discussion in the previous lectures shows that $\mathbb{C}$ with the operations of $+$ and $\cdot$ defined as above is a **field**. Also the operations we defined coincide with the standard operations in $\mathbb{R}$ when we only consider real numbers (that is, when we consider elements $a + b\boldsymbol{i}$ with $b = 0$).

Thus, $\mathbb{R}$ is a subfield of $\mathbb{C}$.

Every complex number z can be uniquely written in the form $a + b\boldsymbol{i}$ with $a, b \in \mathbb{R}$. We call the real number a in this expression the **real part** of z, $\mathrm{Re}(z) = a$, and we call the real number b the **imaginary part** of z, $\mathrm{Im}(z) = b$.

We also call the element $\boldsymbol{i}$ the **imaginary unit**.

Some more terminology

- If $z = a + b\boldsymbol{i}$ with $a, b \in \mathbb{R}$, then the element $a + (-b)\boldsymbol{i} = a - b\boldsymbol{i}$, or in other words, the element $\mathrm{Re}(z) - \mathrm{Im}(z)\boldsymbol{i}$, is called the **conjugate** of z, and is denoted by $\overline{z}$.

- We define the **modulus** or **absolute value** of $z = a + b\boldsymbol{i}$ to be the **non-negative real number**

$$\sqrt{a^2 + b^2} = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}\,.$$

  We denote it by $|z|$.

Remarks

1. If we identify the complex number $z = a + b\boldsymbol{i}$ with the point (a, b) on the Cartesian plane representing $\mathbb{R}^2$, then $|z|$ takes a geometric meaning: it is the **distance** of z from the origin.

2. We have that

$$z \cdot \overline{z} = |z|^2 = \mathrm{Re}(z)^2 + \mathrm{Im}(z)^2\,.$$

   Indeed,

$$z \cdot \overline{z} = (a + b\boldsymbol{i}) \cdot (a - b\boldsymbol{i}) = \Big(a \cdot a - b \cdot (-b)\Big) + \Big(a \cdot (-b) + b \cdot a\Big)\boldsymbol{i} = a^2 + b^2.$$

Field structure of $\mathbb{C}$

Finally, let us gather here some of the conclusions of our discussion in the last lecture:

- The additive identity in $\mathbb{C}$ is the number $0 = 0 + 0\boldsymbol{i}$, while the multiplicative identity is the number $1 = 1 + 0\boldsymbol{i}$. Also, the imaginary unit $\boldsymbol{i}$ coincides with the number $0 + 1\boldsymbol{i}$ (and thus it is identified with the point $(0, 1)$ in $\mathbb{R}^2$).
- Every complex number $z = \mathrm{Re}(z) + \mathrm{Im}(z)\boldsymbol{i}$ has an additive inverse, which is given by
$$-z = (-\mathrm{Re}(z)) + (-\mathrm{Im}(z))\boldsymbol{i}$$
$\big($or more simply $-z = -\mathrm{Re}(z) - \mathrm{Im}(z)\boldsymbol{i}\,\big)$.
- For every complex number z, if z is non-zero (equivalently, if $|z| = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2} > 0$), then there exists a multiplicative inverse, which is given by
$$z^{-1} = \frac{\mathrm{Re}(z)}{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2} + \frac{-\mathrm{Im}(z)}{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}\boldsymbol{i}$$
$$= \frac{\mathrm{Re}(z)}{|z|^2} + \frac{-\mathrm{Im}(z)}{|z|^2}\boldsymbol{i} = \frac{1}{|z|^2}\,\overline{z}\,.$$
Indeed, observe that
$$z \cdot \left(\frac{1}{|z|^2}\,\overline{z}\right) = \frac{1}{|z|^2}\,z \cdot \overline{z} = \frac{1}{|z|^2}|z|^2 = 1\,.$$

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 6

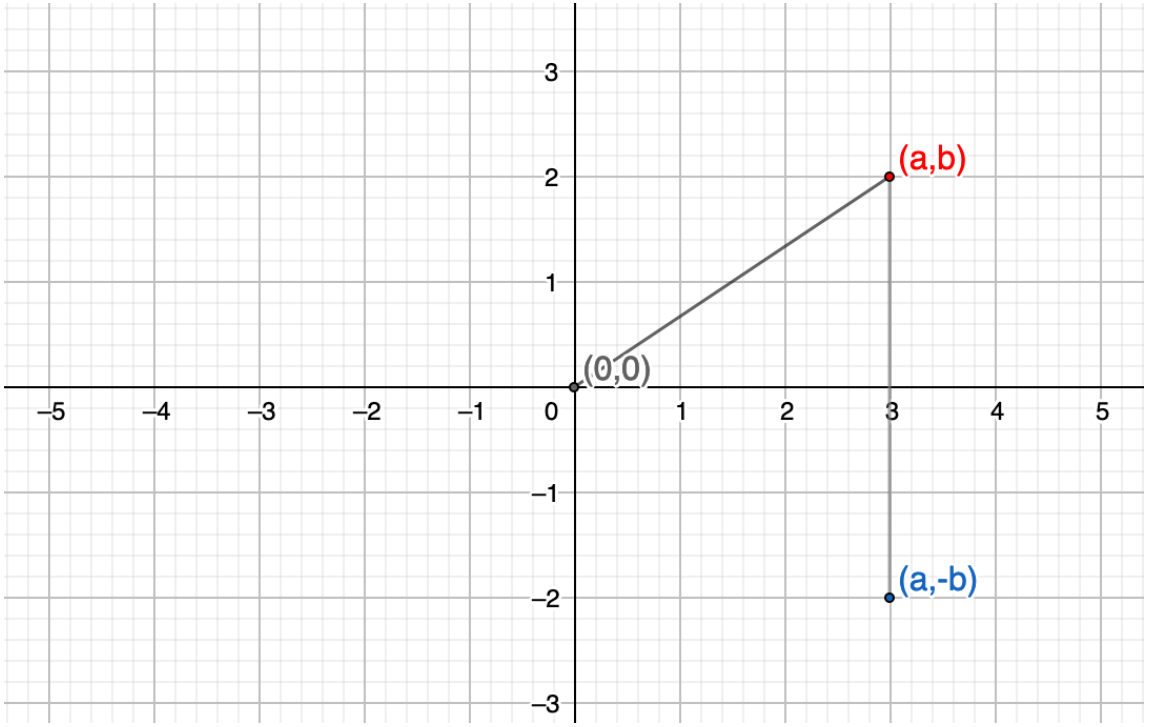
Friday September 11

From last time: the set $\mathbb{C}$ of complex numbers

Recall that every element of $\mathbb{C}$ can be uniquely written in the form $a + b\boldsymbol{i}$ with $a, b \in \mathbb{R}$, and we then say that a is the real part of the number $a + b\boldsymbol{i}$ and b is the imaginary part:

$$\mathrm{Re}(a + b\boldsymbol{i}) = a \quad \text{and} \quad \mathrm{Im}(a + b\boldsymbol{i}) = b.$$

We can then identify the number $a + b\boldsymbol{i}$ with the point (a, b) on the standard Cartesian plane, and many of the notions we define take a geometric meaning:



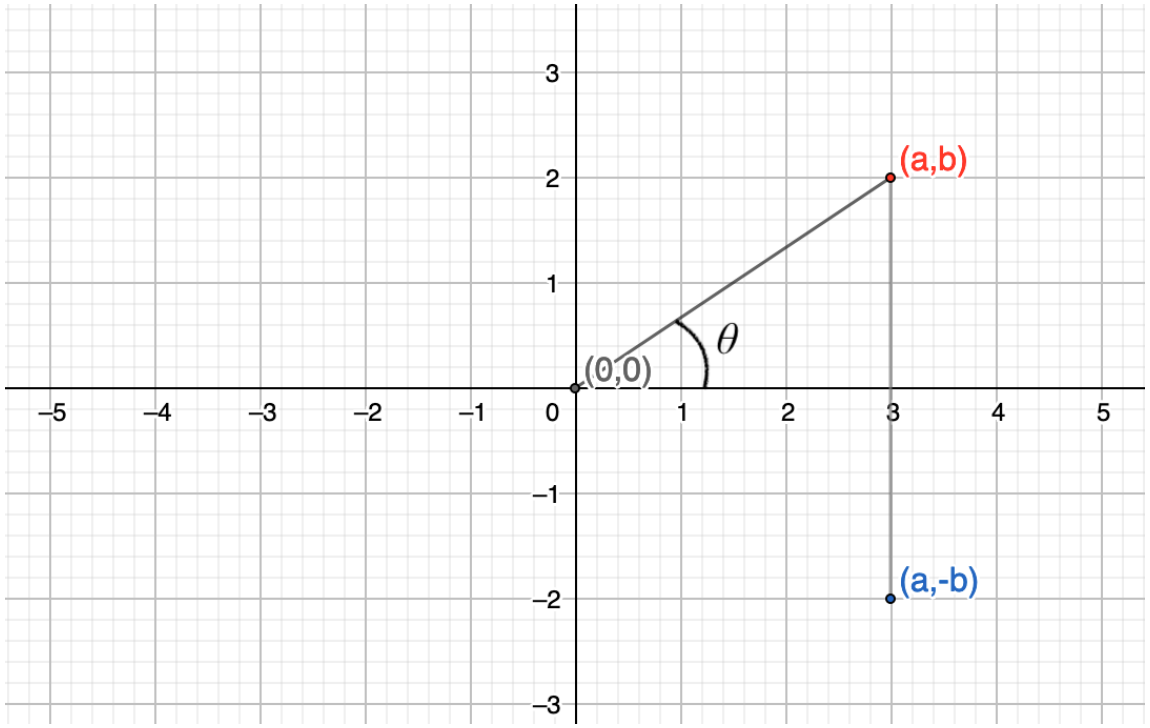
From last time: the set $\mathbb{C}$ of complex numbers

E.g. the *modulus* or *absolute value* $|z|$ of $z = a + b\boldsymbol{i}$ coincides with the Euclidean distance of the point (a, b) from the origin $(0, 0)$.

The *conjugate* $\overline{z}$ of z corresponds to the reflection of the point (a, b) across the x-axis.

Another definition of geometric meaning

We could consider the angle which the directed line segment $\overrightarrow{O\,P_z}$ (with initial point the origin O and terminal point $P_z(a,b)$) forms with the **positive** x-semiaxis.



There are several ways to choose the range for this angle: here we will adopt the convention that $\theta \in [0, \pi]$ if the point P_z is above or on the x-axis, and $\theta \in (-\pi, 0)$ if P_z is below the x-axis.

We won't specify an angle for the origin.

Another definition of geometric meaning

Thus, for non-zero complex numbers $z = a + b\boldsymbol{i}$, we define the *argument* $\arg(z)$ of z to be the unique angle $\theta \in (-\pi, \pi]$ (according to the above convention) that the directed line segment $\overrightarrow{O\,P_z}$ forms with the positive x-semiaxis.

We can also give a formula for $\arg(z)$ (although it might not be easy to remember, and there's no need to memorise it at this point):

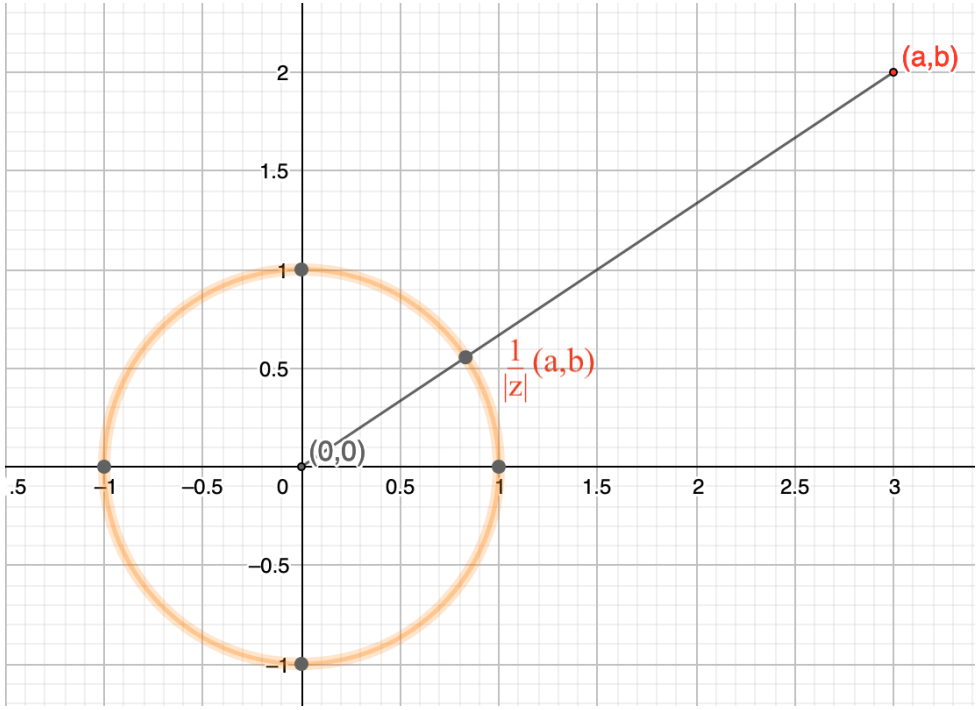
$$\arg(z) = \arg(a + b\boldsymbol{i}) = \begin{cases} \arctan(b/a) & \text{if } a > 0 \\ \arctan(b/a) + \pi & \text{if } a < 0 \text{ and } b \geqslant 0 \\ \arctan(b/a) - \pi & \text{if } a < 0 \text{ and } b < 0 \\ +\frac{\pi}{2} & \text{if } a = 0 \text{ and } b > 0 \\ -\frac{\pi}{2} & \text{if } a = 0 \text{ and } b < 0 \\ \text{undefined} & \text{if } a = b = 0 \end{cases}.$$

Importance of $\arg(z)$

Given a non-zero complex number $z = a + b\boldsymbol{i}$, consider the point $\frac{1}{|z|}z = \frac{a}{|z|} + \frac{b}{|z|}\boldsymbol{i}$. Then this point has modulus 1: indeed,

$$\left|\frac{1}{|z|}z\right| = \sqrt{\left(\frac{a}{|z|}\right)^2 + \left(\frac{b}{|z|}\right)^2} = \sqrt{\frac{a^2+b^2}{|z|^2}} = 1.$$

Thus the point $\frac{1}{|z|}z$ must be found somewhere on the **unit circle** (with centre the origin) of the Cartesian plane:



Importance of $\arg(z)$

But then $\frac{1}{|z|}z$ can be written in the form

$$\cos(\phi) + \sin(\phi)\boldsymbol{i}$$

for some angle $\phi \in \mathbb{R}$.

What should the value of this angle be? It should be equal to $\arg(z)$, or equal to $\arg z + k \cdot 2\pi$ for some $k \in \mathbb{Z}$.

We conclude that we can write

$$z = |z| \cdot \frac{1}{|z|}z = |z| \cdot \Big(\cos\big(\arg(z)\big) + \sin\big(\arg(z)\big)\boldsymbol{i}\Big) \qquad \text{(polar form of } z\text{)}$$

$$\text{and} \quad = |z| \cdot \Big(\cos\big(\arg(z) + 2\pi k\big) + \sin\big(\arg(z) + 2\pi k\big)\boldsymbol{i}\Big)$$

for any $k \in \mathbb{Z}$.

Euler's formula

Because of the field structure of $\mathbb{C}$, we can now extend real polynomial functions so that their domain becomes the entire $\mathbb{C}$: that is, if $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a polynomial function from $\mathbb{R}$ to $\mathbb{R}$, then

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

makes sense for every $z \in \mathbb{C}$ as well.

Moreover, we can now consider polynomials with coefficients from $\mathbb{C}$ as well.

Interestingly, it turns out that we can *extend* other important real functions too:

we can define the exponential function $\exp(\cdot)$*, the cosine function* $\cos(\cdot)$ *and the sine function* $\sin(\cdot)$ *for every complex input* z*, in such a way that, whenever we plug a real input* $r \in \mathbb{R}$*, the value we get is as before.*

Euler's formula

In addition, the extended *exponential function* continues to have some of its very nice properties:

e.g. for every $z, w \in \mathbb{C}$, we have that $e^{z+w} = e^z \cdot e^w$.

Another very important identity we get is Euler's formula:

for every $z \in \mathbb{C}$, $e^{\boldsymbol{i}\, z} = \cos(z) + \sin(z)\boldsymbol{i}$.

This gives us that, for every non-zero complex number z,

$$\begin{aligned} z &= |z| \cdot \Big(\cos\big(\arg(z)\big) + \sin\big(\arg(z)\big)\boldsymbol{i}\Big) \\ &= |z| \cdot e^{\boldsymbol{i}\arg(z)} \qquad \text{(exponential form of } z\text{)} \\ &= e^{\log(|z|)} \cdot e^{\boldsymbol{i}\arg(z)} = e^{\log(|z|)+\boldsymbol{i}\arg(z)}. \end{aligned}$$

Back to solving polynomial equations with coefficients from $\mathbb{R}$ or $\mathbb{C}$

What we have discussed so far gives us the following:

- The equation $x^2 = -1$ has **exactly two** solutions in $\mathbb{C}$, the numbers $\boldsymbol{i}$ and $-\boldsymbol{i}$.
- If $b \in \mathbb{R},\ b < 0$, then the equation $x^2 = b$ has **exactly two** solutions in $\mathbb{C}$, the numbers $\sqrt{|b|}\,\boldsymbol{i}$ and $-\sqrt{|b|}\,\boldsymbol{i}$.
- We conclude that, for every $\alpha \in \mathbb{R}$, the equation $x^2 = \alpha$ can be solved in $\mathbb{C}$ (and in fact, we can find all the possible solutions).
- It turns out that, given any non-zero polynomial p with coefficients from $\mathbb{R}$, or even with coefficients from $\mathbb{C}$, we can now also solve the equation

$$p(x) = 0,$$

and that it will have **exactly $\deg(\boldsymbol{p})$** solutions in $\mathbb{C}$ (of course some of these solutions may be repeated, e.g. we say that the equation $x^{10} = 0$ has 10 solutions, all of which are equal to 0).

This last statement is called the Fundamental Theorem of Algebra.

The equations $x^n = z_0$ and $x^n = 1$

With more advanced tools from Calculus/Analysis, we could give a proof of the Fundamental Theorem of Algebra, as soon as we manage to show that we can solve the equations $x^n = z_0$ for every $z_0 \in \mathbb{C}$.

Note that, if $z_0 = 0$, all the solutions to the equation should be 0 (why?). So consider now the case that $z_0 \neq 0$.
Then we saw that we can write

$$z_0 = |z_0| \cdot e^{\boldsymbol{i}\arg(z_0)}.$$

But because of the rule $e^{z+w} = e^z \cdot e^w$, we can verify that the number

$$w_0 = |z_0|^{1/n} \cdot e^{\boldsymbol{i}\frac{\arg(z_0)}{n}}$$

satisfies the equation $x^n = z_0$ (check what $w_0^n = w_0 \cdot w_0 \cdots w_0$ should give you).
Similarly the numbers

$$w_1 = |z_0|^{1/n} \cdot e^{\boldsymbol{i}\left(\frac{\arg(z_0)}{n} + 2\pi\frac{1}{n}\right)}, \quad w_2 = |z_0|^{1/n} \cdot e^{\boldsymbol{i}\left(\frac{\arg(z_0)}{n} + 2\pi\frac{2}{n}\right)}, \ldots,$$

$$\ldots, \quad w_{n-2} = |z_0|^{1/n} \cdot e^{\boldsymbol{i}\left(\frac{\arg(z_0)}{n} + 2\pi\frac{n-2}{n}\right)} \quad \text{and} \quad w_{n-1} = |z_0|^{1/n} \cdot e^{\boldsymbol{i}\left(\frac{\arg(z_0)}{n} + 2\pi\frac{n-1}{n}\right)}$$

all satisfy the equation, and they are all different.
Thus, these numbers are the n solutions to the equation $x^n = z_0$.

n-th Roots of Unity

In the special case that $z_0 = 1$, the solutions to the equation $x^n = 1$ are called the *n-th roots of unity*. In particular, these solutions are the numbers

$$\rho_0 = e^{i\frac{\arg(1)}{n}} = e^{i0} = 1,$$

$$\rho_1 = e^{i2\pi\frac{1}{n}},$$

$$\rho_2 = e^{i2\pi\frac{2}{n}},$$

$$\ddots \quad \ddots \quad \ddots,$$

$$\text{and} \quad \rho_{n-1} = e^{i2\pi\frac{n-1}{n}}.$$

Example. The numbers 1, $e^{i2\pi/3}$ and $e^{i4\pi/3}$ are all three solutions to the equation $x^3 = 1$, or in other words the 3-rd roots of unity.

Practice Exercise. Find the solutions to the equation $x^2 = 1$, as well as the solutions to the equation $x^4 = 1$ (and try to write each of these numbers in the form $a + bi$ with $a, b \in \mathbb{R}$).

Worth Noting: the solution $\rho_1 = e^{i2\pi\frac{1}{n}}$ of the equation $x^n = 1$ is always particularly important, as it can give us all other n-th roots of unity too: we have that $\rho_2 = \rho_1^2 = \rho_1 \cdot \rho_1$, $\rho_3 = \rho_1^3 = \rho_1 \cdot \rho_1 \cdot \rho_1$, and so on all the way to $\rho_{n-2} = \rho_1^{n-2}$ and $\rho_{n-1} = \rho_1^{n-1}$, and finally $\rho_1^n = 1 = \rho_0$.

Finite fields
and the notion of ‘commutative ring’

Definition of the notion of 'commutative ring'

Definition. A *commutative ring* $\mathcal{R}$ is a set of elements together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x,y) \in \mathcal{R}\times\mathcal{R} \;\mapsto\; x+y \in \mathcal{R} \\ \textit{multiplication} \qquad & (x,y) \in \mathcal{R}\times\mathcal{R} \;\mapsto\; x\cdot y \in \mathcal{R} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathcal{R}$, $\quad x+y = y+x$ *(commutativity)*

(ii) for all $x, y, z \in \mathcal{R}$, $\quad (x+y)+z = x+(y+z)$ *(associativity)*

(iii) there exists an element 0 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 0+x = x+0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathcal{R}$, there exists an element $w = w_x$ such that

$$x+w = w+x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

Definition of the notion of 'commutative ring' (cont.)

(i') for all $x, y \in \mathcal{R}$, $\quad x \cdot y = y \cdot x$ *(commutativity)*

(ii') for all $x, y, z \in \mathcal{R}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

and finally

(viii) for all $x, y, z \in \mathcal{R}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

The last property, called the Distributive Law, relates the operations of addition and multiplication in $\mathcal{R}$.

Examples of commutative rings

- Every field is a commutative ring too (why?). So all the examples of fields that we have seen so far ($\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) are commutative rings.
- The converse is not true (in this generality): that is, there are commutative rings which are not fields (see next bullet point, but also HW1, where you have to verify one more such example).
- The set $\mathbb{Z}$ of integers, with standard addition and multiplication, is a commutative ring (recall that $\mathbb{Z}$ is not a field).
- The set $\mathbb{N}_0$ of non-negative integers, with standard addition and multiplication, is **neither** a field, **nor** a commutative ring.

A structure with two elements

Consider the set $\{Green, Yellow\}$.

Let us define operations of 'addition' and 'multiplication' on this set according to the following tables:

$+$	*Green*	*Yellow*
Green	*Yellow*	*Green*
Yellow	*Green*	*Yellow*

$\cdot$	*Green*	*Yellow*
Green	*Green*	*Yellow*
Yellow	*Yellow*	*Yellow*

What properties can you say these operations satisfy?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 7

Monday September 14

From last time: Definition of 'Commutative Ring'

Definition. A *commutative ring* $\mathcal{R}$ is a set of elements together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x,y) \in \mathcal{R}\times\mathcal{R} \;\mapsto\; x+y \in \mathcal{R} \\ \textit{multiplication} \qquad & (x,y) \in \mathcal{R}\times\mathcal{R} \;\mapsto\; x\cdot y \in \mathcal{R} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathcal{R}$, $\quad x+y = y+x$ *(commutativity)*

(ii) for all $x, y, z \in \mathcal{R}$, $\quad (x+y)+z = x+(y+z)$ *(associativity)*

(iii) there exists an element 0 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 0+x = x+0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathcal{R}$, there exists an element $w = w_x$ such that

$$x+w = w+x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

From last time: Definition of 'Commutative Ring'

(i') for all $x, y \in \mathcal{R}$, $\quad x \cdot y = y \cdot x$ *(commutativity)*

(ii') for all $x, y, z \in \mathcal{R}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

and finally

(viii) for all $x, y, z \in \mathcal{R}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

The last property, called the Distributive Law, relates the operations of addition and multiplication in $\mathcal{R}$.

Testing Our Understanding

Assume $(S, \oplus, \odot)$ is a triple of a set S and of two operations of the following form:

$$(x, y) \in S \times S \ \mapsto \ x \oplus y \in S$$
$$(x, y) \in S \times S \ \mapsto \ x \odot y \in S$$

(with the first operation considered the 'addition', the second operation the 'multiplication').

Question. How do we check whether $(S, \oplus, \odot)$ is a commutative ring?

Answer. We need to check that

- **$\oplus$ is commutative**
- **$\oplus$ is associative**
- **There exists a neutral element 0 for $\oplus$**
- **For every $x \in S$, there exists $y \in S$ such that $x \oplus y = 0$**
- **$\odot$ is commutative**
- **$\odot$ is associative**
- **There exists a neutral element 1 for $\odot$**
- **The Distributive Law holds**

Observe that, in a commutative ring, it is not guaranteed that every non-zero element has a multiplicative inverse.

This doesn't preclude the possibility that some elements will have multiplicative inverses (for example, the multiplicative identity 1 is always its own multiplicative inverse). It is useful to introduce a special term for elements that have a multiplicative inverse.

Terminology

Let $\mathcal{R}$ be a commutative ring, and let $x \in \mathcal{R}$. We say that x is *multiplicatively invertible*, or more simply *invertible*, if there exists $y \in \mathcal{R}$ so that

$$x \cdot y = y \cdot x = 1_{\mathcal{R}} .$$

Examples of structures with a finite number of elements

A structure with two elements

Consider the set $\{Green, Yellow\}$.

Let us define operations of 'addition' and 'multiplication' on this set according to the following tables:

$+$	*Green*	*Yellow*
Green	*Yellow*	*Green*
Yellow	*Green*	*Yellow*

$\cdot$	*Green*	*Yellow*
Green	*Green*	*Yellow*
Yellow	*Yellow*	*Yellow*

What properties can you say these operations satisfy?

We can immediately check from the tables that: *addition is commutative*, *there exists an additive identity*, *every element has an additive inverse*, *multiplication is commutative*, *there exists a multiplicative identity* and *every non-zero element has a multiplicative inverse.*

How?

What about the associativity of addition, the associativity of multiplication and the distributive law?

To check these, we should theoretically check all possible cases: e.g. for the associativity of addition, we should confirm that

$$(\mathit{Yellow} + \mathit{Yellow}) + \mathit{Green} = \mathit{Green} = \mathit{Yellow} + (\mathit{Yellow} + \mathit{Green})$$

which, by the already verified commutativity of addition, are also equal to

$$(\mathit{Yellow} + \mathit{Green}) + \mathit{Yellow} = (\mathit{Green} + \mathit{Yellow}) + \mathit{Yellow}$$
$$= \mathit{Yellow} + (\mathit{Green} + \mathit{Yellow}) = \mathit{Green} + (\mathit{Yellow} + \mathit{Yellow}).$$

Similarly, we have

$$(\mathit{Green} + \mathit{Green}) + \mathit{Yellow} = \mathit{Yellow} = \mathit{Green} + (\mathit{Green} + \mathit{Yellow})$$
$$= (\mathit{Green} + \mathit{Yellow}) + \mathit{Green} = (\mathit{Yellow} + \mathit{Green}) + \mathit{Green}$$
$$= \mathit{Green} + (\mathit{Yellow} + \mathit{Green}) = \mathit{Yellow} + (\mathit{Green} + \mathit{Green}).$$

Finally, for every $x \in \{\mathit{Green}, \mathit{Yellow}\}$, we obviously have

$$(x + x) + x = x + (x + x).$$

Next, checking the associativity of multiplication is much more immediate in this structure, given that for all $x, y, z \in \{\mathit{Green}, \mathit{Yellow}\}$, we have that, if at least one of them is equal to Yellow, then

$$(x \cdot y) \cdot z = \mathit{Yellow} = x \cdot (y \cdot z).$$

So the only remaining case is when $x = y = z = \mathit{Green}$, where we clearly have $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.

Finally, checking the right distributive property in this structure is also not hard: for every $x, y \in \{\mathit{Green}, \mathit{Yellow}\}$ we have

$$(x + y) \cdot \mathit{Yellow} = \mathit{Yellow} = x \cdot \mathit{Yellow} + y \cdot \mathit{Yellow},$$
$$\text{while similarly } \; (x + y) \cdot \mathit{Green} = x + y = x \cdot \mathit{Green} + y \cdot \mathit{Green}.$$

Moreover, the left distributive property follows from this, if we combine it with the previously verified commutativity of addition.

Is there an idea behind the rules we gave for addition and multiplication on $\{\textit{Green}, \textit{Yellow}\}$?

YES! Remember the rules we see from early on when being taught math:

$$\begin{aligned}
\textsc{Even} + \textsc{Even} = \textsc{Even} = \textsc{Odd} + \textsc{Odd},\\
\textsc{Even} + \textsc{Odd} = \textsc{Odd} + \textsc{Even} = \textsc{Odd},\\
\textsc{Even} \cdot \textsc{Even} = \textsc{Even} \cdot \textsc{Odd} = \textsc{Odd} \cdot \textsc{Even} = \textsc{Even},\\
\textsc{Odd} \cdot \textsc{Odd} = \textsc{Odd}.
\end{aligned}$$

Based on our tables, we have chosen

Yellow to correspond to Even, Green to correspond to Odd.

Reminder. When do we call an integer m even? When we can write $m = 2k$ for some other integer k, or, in other words, when we get remainder 0 when dividing m by 2.

When do we call an integer n odd? When we can write $n = 2l + 1$ for some other integer l, or, in other words, when we get remainder 1 when dividing n by 2.

We can now try to generalise this idea: consider a set containing the following elements:

When divided by 3, leaves remainder 0,
When divided by 3, leaves remainder 1,
When divided by 3, leaves remainder 2.

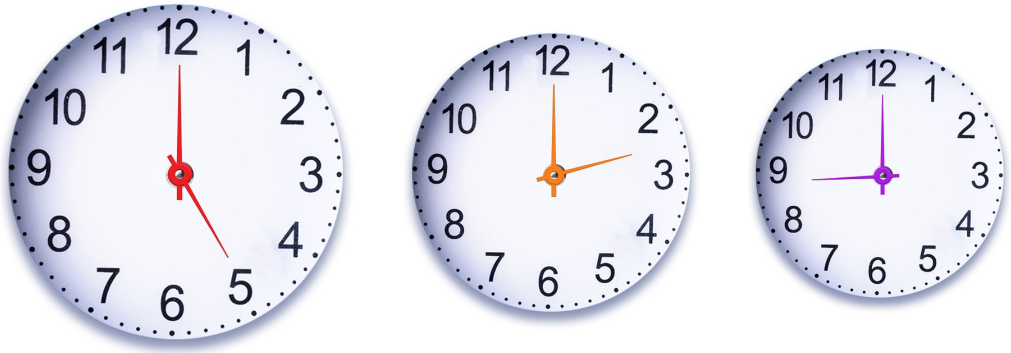
Let us denote these elements by [0], [1] and [2] respectively.

What should $[0] + [1]$ be in the structure we are trying to define (assuming we want our structure to be "nice")? It looks like it should be [1].

What should $[1] + [1]$ be? [2] What about $[2] + [1]$? [0] $[2] + [2]$? [1]

What about $[0] \cdot [2]$? It looks like it should be [0]. $[2] \cdot [1]$? [2] $[2] \cdot [2]$? [1]

Examples from real life?



- Our class started at 9am. What time will it be 13 hours after the beginning of our class?

- Today is Monday. What day will it be after 18 days have passed?

Some basic facts from Number Theory

Euclidean Division

Let m be a positive integer, and let $n \in \mathbb{Z}$. We can find **unique** $q, r \in \mathbb{Z}$ such that

- $0 \leqslant r < m$,
- $n = m \cdot q + r$.

Here we call q the *quotient* of the division process, and r the *remainder*.

- If the remainder r in this process is equal to 0, then we say that m *divides* n, or that m is a *divisor* of n, or finally that n is a *multiple* of m.
- Let $n_1, n_2 \in \mathbb{Z}$ with at least one of them non-zero. The largest positive integer that divides both n_1 and n_2 is called their *greatest common divisor*.

  We denote it by $\gcd(n_1, n_2)$.
- A positive integer $m > 1$ is called a *prime* integer if its only positive divisors are 1 and m itself.
- If an integer $m > 1$ is not a prime, we call it a *composite* integer.
- Two integers n_1, n_2 are called *relatively prime* or *coprime* if $\gcd(n_1, n_2) = 1$.

A beautiful result

Bézout's identity

Let $n_1, n_2 \in \mathbb{Z}$, and set $d = \gcd(n_1, n_2)$.

Then we can find $\kappa, \ell \in \mathbb{Z}$ *(not in a unique way)* such that

$$d = \kappa \cdot n_1 + \ell \cdot n_2 \, .$$

Examples. 1) Note that $\gcd(4, 15) = 1$. (Why? List the (positive) divisors of 4, and similarly list the divisors of 15, and note that their only common positive divisor is 1.) Recall that we say in this case that 4 and 15 are relatively prime.

Now, observe that we can write $1 = 4 \cdot 4 + (-1) \cdot 15 = (-11) \cdot 4 + 3 \cdot 15$.

2) Note that $\gcd(12, 20) = 4$.

Observe that we can write $4 = (-3) \cdot 12 + 2 \cdot 20 = 7 \cdot 12 + (-4) \cdot 20$.

The structure $\mathbb{Z}_m$

Given a positive integer $m > 1$, denote by $[n]$ (or by $[n]_m$ when the setting is not already clear) the set of all integers which, when divided by m, leave the same remainder as when n is divided by m: in other words,

$$\begin{aligned}[n] &:= \left\{ n' \in \mathbb{Z} : n' - n \text{ is a multiple of } m \right\} \\ &= \left\{ \ell \cdot m + n \, : \, \ell \in \mathbb{Z} \right\}.\end{aligned}$$

Based on this notation, set $\mathbb{Z}_m$ to be the set

$$\left\{ [0], [1], [2], \ldots, [m-2], [m-1] \right\}$$

(note that the elements of $\mathbb{Z}_m$ are all sets themselves, they are subsets of the integers).

Define operations of addition and multiplication on $\mathbb{Z}_m$ according to the following rule: if $[n], [k] \in \mathbb{Z}_m$, **then choose a representative n_1 of the class $[n]$ (that is, an element n_1 of the subset $[n]$), and similarly choose a representative k_1 of the class $[k]$,** and set

$$[n] + [k] := [n_1 + k_1], \qquad \text{(that is, the class containing the integer } n_1 + k_1\text{)}$$

$$[n] \cdot [k] := [n_1 \cdot k_1]. \qquad \text{(that is, the class containing the integer } n_1 \cdot k_1\text{)}$$

Theorem

The operations are **well-defined.**
That is, for every $[n], [k] \in \mathbb{Z}_m$,
if n_1, n_2 are both representatives of the class $[n]$,
and k_1, k_2 are both representatives of the class $[k]$,
then

$$[n_1 + k_1] = [n_2 + k_2] \qquad \text{and} \qquad [n_1 \cdot k_1] = [n_2 \cdot k_2] \,.$$

For the proof, see next lecture.

An example

Let's write down the tables of addition and multiplication for $\mathbb{Z}_7$
(left as an exercise until the next lecture):

$+$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]							
[1]							
[2]							
[3]							
[4]							
[5]							
[6]							

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]							
[1]							
[2]							
[3]							
[4]							
[5]							
[6]							

Important Remarks

- For every natural number $m > 1$, the structure $\mathbb{Z}_m$ with the operations we defined is a commutative ring.
- For every prime number $p > 1$, the structure $\mathbb{Z}_p$ with the operations we defined is a field!

To prove the latter, we rely on Bézout's identity *(we'll explain this next time)*.

What about when m is not a prime?

We have already concluded that $\mathbb{Z}_m$ is a commutative ring.
But is it a field?

The answer here is **NO**.

To justify it, recall one of the propositions we have already proven in class (see Lecture 3):

If $\mathbb{F}$ is a field,
then, for every $x, y \in \mathbb{F}$, if $x \cdot y = 0$, we have that $x = 0$ or $y = 0$.

Instinctively, we can say that this is equivalent to:

If can we find $x, y \in \mathbb{F}$ such that $x \cdot y = 0$ and $x \neq 0$, $y \neq 0$,
then $\mathbb{F}$ is **NOT** a field.

In Mathematics, the latter statement is called the **contrapositive** of the former statement *(and indeed, no matter what statement we start with, either both that statement and its contrapositive are true, or both of them are false)*.

How could we use this to show e.g. that $\mathbb{Z}_6$ is NOT a field?

Let's look at the table of multiplication for $\mathbb{Z}_6$:

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

Conclusion. Given that e.g. $[2] \cdot [3] = [0]$, while neither [2] nor [3] are equal to the additive identity, we get that $\mathbb{Z}_6$ **cannot be a field.**

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 8

Tuesday September 15

Reminder

Euclidean Division

Let m be a positive integer, and let $n \in \mathbb{Z}$. We can find **unique** $q, r \in \mathbb{Z}$ such that

- $0 \leqslant r < m$,
- $n = m \cdot q + r$.

Here we call q the *quotient* of the division process, and r the *remainder*.

Reminder: The structure $\mathbb{Z}_m$

Given a positive integer $m > 1$, and given $n \in \mathbb{Z}$, denote by $[n]$

(or by $[n]_m$ when the setting is not already clear)

the set of all integers which, when divided by m, leave the same remainder as when n is divided by m: in other words,

$$\begin{aligned}[n] &:= \{m \cdot \ell + n \,:\, \ell \in \mathbb{Z}\} \\ &= \{n' \in \mathbb{Z} : n' - n \text{ is a multiple of } m\}\,.\end{aligned}$$

Based on this notation, set $\mathbb{Z}_m$ to be the set

$$\{[0], [1], [2], \ldots, [m-2], [m-1]\}$$

(note that the elements of $\mathbb{Z}_m$ are all sets themselves, they are subsets of the integers).

Define operations of addition and multiplication on $\mathbb{Z}_m$ according to the following rule: if $[n], [k] \in \mathbb{Z}_m$, **then choose a representative n_1 of the class $[n]$ (that is, an element n_1 of the subset $[n]$), and similarly choose a representative k_1 of the class $[k]$,** and set

$$[n] + [k] := [n_1 + k_1]\,, \qquad \text{(that is, the class containing the integer } n_1 + k_1\text{)}$$

$$[n] \cdot [k] := [n_1 \cdot k_1]\,. \qquad \text{(that is, the class containing the integer } n_1 \cdot k_1\text{)}$$

Modular Arithmetic

When we start 'identifying' different integers like that, and placing some of them in one subset / 'class', some others in a different subset / 'class', and so on, based on whether their difference is divided by a fixed integer $m > 1$, and then we do the operations of addition and multiplication as above, in a way that only depends on which classes our inputs are taken from,

we say that we do *modular arithmetic*.

If two integers n_1, n_2 are placed in the same class, we say that they are *congruent modulo m,* and we write

$$n_1 \equiv n_2 \pmod{m}.$$

We also call the class $[n_1] = [n_2]$, which contains both n_1 and n_2 in this case, one of the *congruence classes modulo m.* Recall that we defined the elements of $\mathbb{Z}_m$ to be

the *m* different congruence classes modulo *m* that we can have.

Modular Arithmetic

Finally, as we may have already realised, it's often much easier to list the different congruence classes by picking a representative from each class (and 'almost' identifying the representative with the class, in terms of notation at least).

For instance, in the slide introducing the structure $\mathbb{Z}_m$, we picked as representatives the different remainders we could have when dividing by m, that is, the integers $0, 1, 2, \ldots, m-2$ and $m-1$.

However, as we will see, we could also pick other representatives.

Terminology. Any set $\{k_1, k_2, \ldots, k_{m-1}, k_m\}$ of m integers with the property that **no two** of these integers are congruent modulo m will be called a *complete set of representatives modulo m*. We will then have that

$$\mathbb{Z}_m = \{[k_1], [k_2], \ldots, [k_{m-1}], [k_m]\}.$$

What these classes look like in specific examples

If $m = 3$, we have that

$$\begin{aligned}\mathbb{Z}_3 = \Big\{ &\{\ldots, -9, -6, -3, 0, 3, 6, 9, 12, \ldots\}, \\ &\{\ldots, -8, -5, -2, 1, 4, 7, 10, 13, \ldots\}, \\ &\{\ldots, -7, -4, -1, 2, 5, 8, 11, 14, \ldots\} \Big\} \\ = \Big\{ &\{3k : k \in \mathbb{Z}\}, \{3k+1 : k \in \mathbb{Z}\}, \{3k+2 : k \in \mathbb{Z}\} \Big\}.\end{aligned}$$

Moreover, a complete set of representatives, as we have already seen, is the set $\{0, 1, 2\}$, but also the set $\{-15, -8, 5\}$. In other words, we have

$$\mathbb{Z}_3 = \{[0], [1], [2]\} = \{[-15], [-8], [5]\}.$$

What these classes look like in specific examples

If $m = 4$, then we have that

$$\begin{aligned}\mathbb{Z}_4 = \Big\{ &\{\ldots, -8, -4, 0, 4, 8, 12, \ldots\}, \\ &\{\ldots, -7, -3, 1, 5, 9, 13, \ldots\}, \\ &\{\ldots, -6, -2, 2, 6, 10, 14, \ldots\}, \\ &\{\ldots, -5, -1, 3, 7, 11, 15, \ldots\} \Big\} \\ = \Big\{ &\{4k : k \in \mathbb{Z}\}, \{4k + 9 : k \in \mathbb{Z}\}, \\ &\{4k - 6 : k \in \mathbb{Z}\}, \{4k + 23 : k \in \mathbb{Z}\} \Big\}.\end{aligned}$$

As can be seen from the last line, a complete set of representatives modulo 4 is the set $\{0, 9, -6, 23\}$.

Theorem

The operations are **well-defined.**
That is, for every $[n], [k] \in \mathbb{Z}_m$,
if n_1, n_2 are both representatives of the class $[n]$,
and k_1, k_2 are both representatives of the class $[k]$,
then

$$[n_1 + k_1] = [n_2 + k_2] \qquad \text{and} \qquad [n_1 \cdot k_1] = [n_2 \cdot k_2] \,.$$

Proof of the theorem

We first show that $[n_1 + k_1] = [n_2 + k_2]$. Recall that, to show this, we have to check that $n_1 + k_1$ and $n_2 + k_2$ are congruent modulo m, or in other words that $(n_1 + k_1) - (n_2 + k_2)$ is divided by m.

Since n_1, n_2 are both representatives of the class $[n]$, we have that n_1 and n_2 are congruent modulo m, or equivalently that $n_1 - n_2$ is a multiple of m. Similarly, since k_1, k_2 are both representatives of the class $[k]$, we have that $k_1 - k_2$ is a multiple of m.

But then

$$(n_1 + k_1) - (n_2 + k_2) = (n_1 - n_2) + (k_1 - k_2)$$

is the sum of two multiples of m, and hence a multiple of m too.

Analogously, we check that $[n_1 \cdot k_1] = [n_2 \cdot k_2]$ by showing that $n_1 \cdot k_1 - n_2 \cdot k_2$ is divided by m.

We can write

$$\begin{aligned} n_1 \cdot k_1 - n_2 \cdot k_2 &= n_1 \cdot k_1 - n_2 \cdot k_1 + n_2 \cdot k_1 - n_2 \cdot k_2 \\ &= (n_1 - n_2) \cdot k_1 + n_2 \cdot (k_1 - k_2). \end{aligned}$$

Observe that the last expression is the sum of two multiples of m *(indeed, given that e.g.* $n_1 - n_2 = m \cdot q$ *for some* $q \in \mathbb{Z}$*, we also have that* $(n_1 - n_2) \cdot k_1 = m \cdot q \cdot k_1$*)*, hence it is a multiple of m too.

An example

Let's write down the tables of addition and multiplication for $\mathbb{Z}_7$:

$+$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[1]	[1]	[2]	[3]	[4]	[5]	[6]	[0]
[2]	[2]	[3]	[4]	[5]	[6]	[0]	[1]
[3]	[3]	[4]	[5]	[6]	[0]	[1]	[2]
[4]	[4]	[5]	[6]	[0]	[1]	[2]	[3]
[5]	[5]	[6]	[0]	[1]	[2]	[3]	[4]
[6]	[6]	[0]	[1]	[2]	[3]	[4]	[5]

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[2]	[0]	[2]	[4]	[6]	[1]	[3]	[5]
[3]	[0]	[3]	[6]	[2]	[5]	[1]	[4]
[4]	[0]	[4]	[1]	[5]	[2]	[6]	[3]
[5]	[0]	[5]	[3]	[1]	[6]	[4]	[2]
[6]	[0]	[6]	[5]	[4]	[3]	[2]	[1]

Important Remarks

- For every natural number $m > 1$, the structure $\mathbb{Z}_m$ with the operations we defined is a commutative ring.

We have to check that all the axioms of a commutative ring are satisfied.

But based on the theorem we just proved, which allows us to add or multiply two classes in $\mathbb{Z}_m$ by first adding or multiplying any two representatives of those classes that we have considered, we can verify those axioms by relying on the corresponding properties of the addition and multiplication in $\mathbb{Z}$.

E.g., to show that addition in $\mathbb{Z}_m$ is associative, we consider arbitrary classes $[n], [k]$ and $[l]$ of $\mathbb{Z}_m$, as well as arbitrary representatives n_1, k_1 and l_1 of these classes respectively, and note that

$$\begin{aligned}
\big([n] + [k]\big) + [l] &= [n_1 + k_1] + [l_1] \\
&= [(n_1 + k_1) + l_1] \\
&= [n_1 + (k_1 + l_1)] \qquad \textit{(because addition in } \mathbb{Z} \textit{ is associative)} \\
&= [n_1] + [k_1 + l_1] \\
&= [n] + \big([k] + [l]\big)\,.
\end{aligned}$$

Similarly, to show e.g. the right distributive property in $\mathbb{Z}_m$, we can write

$$\begin{aligned}
\big([n] + [k]\big) \cdot [l] &= [n_1 + k_1] \cdot [l_1] \\
&= [(n_1 + k_1) \cdot l_1] \\
&= [n_1 \cdot l_1 + k_1 \cdot l_1] \qquad \textit{(due to right distributive property in } \mathbb{Z}\textit{)} \\
&= [n_1 \cdot l_1] + [k_1 \cdot l_1] \\
&= [n] \cdot [l] + [k] \cdot [l]\,.
\end{aligned}$$

We also note that $[0]$ is the additive identity in $\mathbb{Z}_m$ given that 0 is the additive identity in $\mathbb{Z}$: indeed, for every class $[n]$ in $\mathbb{Z}_m$, and any representative n_1 of it, we have

$$\begin{aligned}[n] + [0] = [n_1 + 0] &= [n_1] = [n] \\ &= [0 + n_1] = [0] + [n].\end{aligned}$$

Also, for every class $[n]$ in $\mathbb{Z}_m$, the class $[-n]$ which contains $-n$ is the additive inverse of $[n]$:

$$[n] + [-n] = [n + (-n)] = [0].$$

Similarly, we observe that $[1]$ is the multiplicative identity in $\mathbb{Z}_m$ given that 1 is the multiplicative identity in $\mathbb{Z}$:

$$\begin{aligned}[n] \cdot [1] = [n_1 \cdot 1] &= [n_1] = [n] \\ &= [1 \cdot n_1] = [1] \cdot [n].\end{aligned}$$

Important Remarks (cont.)

- For every natural number $m > 1$, the structure $\mathbb{Z}_m$ with the operations we defined is a commutative ring.

- For every prime number $p > 1$, the structure $\mathbb{Z}_p$ with the operations we defined is a field!

To prove the latter, we rely on Bézout's identity.

Reminder

Bézout's identity

Let $n_1, n_2 \in \mathbb{Z}$, and set $d = \gcd(n_1, n_2)$.

Then we can find $\kappa, \ell \in \mathbb{Z}$ *(not in a unique way)* such that

$$d = \kappa \cdot n_1 + \ell \cdot n_2 \, .$$

$\mathbb{Z}_p$ is a field when p is a prime

Given that we already concluded that Z_p with the operations we defined is a commutative ring, what still remains to be checked is that

every element of $\mathbb{Z}_p$ different from $[0]_p$ has a multiplicative inverse.

Consider an element $[n] \in \mathbb{Z}_p$ with $[n] \neq [0]$. Recall that the latter means that n and 0 are not congruent modulo p, or in other words that $n = n - 0$ is **not** a multiple of p.

Consider now the positive divisors of p: given that p is prime, its positive divisors are only 1 and p.
Also, we just said that p is **not** a divisor of n, therefore the only common positive divisor of p and n is 1. This shows that $\gcd(p, n) = 1$.

We now make use of Bézout's identity: it guarantees the existence of some $\kappa, \ell \in \mathbb{Z}$ such that

$$1 = \kappa \cdot p + \ell \cdot n\,.$$

But then we can write

$$\begin{aligned}[1] &= [\kappa \cdot p + \ell \cdot n] \\ &= [\kappa \cdot p] + [\ell \cdot n] \\ &= [0] + [\ell] \cdot [n] = [\ell] \cdot [n],\end{aligned}$$

which shows that $[\ell]$ is the multiplicative inverse of $[n]$.

What about when m is not a prime?

We have already concluded that $\mathbb{Z}_m$ is a commutative ring.
But is it a field?

The answer here is **NO**.

To justify it, recall one of the propositions we have already proven in class (see Lecture 3):

If $\mathbb{F}$ is a field,
then, for every $x, y \in \mathbb{F}$, if $x \cdot y = 0$, we have that $x = 0$ or $y = 0$.

Instinctively, we can say that this is equivalent to:

If can we find $x, y \in \mathbb{F}$ such that $x \cdot y = 0$ and $x \neq 0$, $y \neq 0$,
then $\mathbb{F}$ is **NOT** a field.

In Mathematics, the latter statement is called the **contrapositive** of the former statement *(and indeed, no matter what statement we start with, either both that statement and its contrapositive are true, or both of them are false)*.

How could we use this to show e.g. that $\mathbb{Z}_6$ is NOT a field?

Let's look at the table of multiplication for $\mathbb{Z}_6$:

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

Conclusion. Given that e.g. $[2] \cdot [3] = [0]$, while neither [2] nor [3] are equal to the additive identity, we get that $\mathbb{Z}_6$ **cannot be a field.**

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 9

Wednesday September 16

Reminder: The structure $\mathbb{Z}_m$

Given a positive integer $m > 1$, and given $n \in \mathbb{Z}$, denote by $[n]$

(or by $[n]_m$ when the setting is not already clear)

the set of all integers which, when divided by m, leave the same remainder as when n is divided by m: in other words,

$$\begin{aligned}[n] &:= \{m \cdot \ell + n \,:\, \ell \in \mathbb{Z}\} \\ &= \{n' \in \mathbb{Z} : n' - n \text{ is a multiple of } m\}\,.\end{aligned}$$

Based on this notation, set $\mathbb{Z}_m$ to be the set

$$\{[0], [1], [2], \ldots, [m-2], [m-1]\}$$

(note that the elements of $\mathbb{Z}_m$ are all sets themselves, they are subsets of the integers).

Define operations of addition and multiplication on $\mathbb{Z}_m$ according to the following rule: if $[n], [k] \in \mathbb{Z}_m$, **then choose a representative n_1 of the class $[n]$ (that is, an element n_1 of the subset $[n]$), and similarly choose a representative k_1 of the class $[k]$,** and set

$$[n] + [k] := [n_1 + k_1]\,, \qquad \text{(that is, the class containing the integer } n_1 + k_1\text{)}$$

$$[n] \cdot [k] := [n_1 \cdot k_1]\,. \qquad \text{(that is, the class containing the integer } n_1 \cdot k_1\text{)}$$

An example

Let's write down the tables of addition and multiplication for $\mathbb{Z}_7$:

$+$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[1]	[1]	[2]	[3]	[4]	[5]	[6]	[0]
[2]	[2]	[3]	[4]	[5]	[6]	[0]	[1]
[3]	[3]	[4]	[5]	[6]	[0]	[1]	[2]
[4]	[4]	[5]	[6]	[0]	[1]	[2]	[3]
[5]	[5]	[6]	[0]	[1]	[2]	[3]	[4]
[6]	[6]	[0]	[1]	[2]	[3]	[4]	[5]

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]	[6]
[2]	[0]	[2]	[4]	[6]	[1]	[3]	[5]
[3]	[0]	[3]	[6]	[2]	[5]	[1]	[4]
[4]	[0]	[4]	[1]	[5]	[2]	[6]	[3]
[5]	[0]	[5]	[3]	[1]	[6]	[4]	[2]
[6]	[0]	[6]	[5]	[4]	[3]	[2]	[1]

Important Remarks

- For every natural number $m > 1$, the structure $\mathbb{Z}_m$ with the operations we defined is a commutative ring.
  - $[0]_m = \{m \cdot \ell : \ell \in \mathbb{Z}\}$ is the additive identity in $\mathbb{Z}_m$,
  - while $[1]_m = \{m \cdot \ell + 1 : \ell \in \mathbb{Z}\}$ is the multiplicative identity.

- For every prime number $p > 1$, the structure $\mathbb{Z}_p$ with the operations we defined is a field!

To prove the latter, we rely on Bézout's identity.

Reminder

Bézout's identity

Let $n_1, n_2 \in \mathbb{Z}$, and set $d = \gcd(n_1, n_2)$.

Then we can find $\kappa, \ell \in \mathbb{Z}$ *(not in a unique way)* such that

$$d = \kappa \cdot n_1 + \ell \cdot n_2 \, .$$

$\mathbb{Z}_p$ is a field when p is a prime

Given that we already concluded that Z_p with the operations we defined is a commutative ring, what still remains to be checked is that

every element of $\mathbb{Z}_p$ different from $[0]_p$ has a multiplicative inverse.

Consider an element $[n] \in \mathbb{Z}_p$ with $[n] \neq [0]$. Recall that the latter means that n and 0 are not congruent modulo p, or in other words that $n = n - 0$ is **not** a multiple of p.

Consider now the positive divisors of p: given that p is prime, its positive divisors are only 1 and p.
Also, we just said that p is **not** a divisor of n, therefore the only common positive divisor of p and n is 1. This shows that $\gcd(p, n) = 1$.

We now make use of Bézout's identity: it guarantees the existence of some $\kappa, \ell \in \mathbb{Z}$ such that

$$1 = \kappa \cdot p + \ell \cdot n\,.$$

But then we can write

$$\begin{aligned}[1] &= [\kappa \cdot p + \ell \cdot n] \\ &= [\kappa \cdot p] + [\ell \cdot n] \\ &= [0] + [\ell] \cdot [n] = [\ell] \cdot [n],\end{aligned}$$

which shows that $[\ell]$ is the multiplicative inverse of $[n]$.

How to apply this in specific examples

Question. How could we find the multiplicative inverses of (some of) the non-zero elements of $\mathbb{Z}_{11}$?

Let's discuss here how we could find the multiplicative inverses of [10], [2] and [8].

- To find the multiplicative inverse of [10], the first thing we could do is complete that row of the table of multiplication of $\mathbb{Z}_{11}$: that is, consider all products of the form $[10] \cdot [k]$ with $[k] \neq [0]$, and see which one will equal [1].

  Alternatively, we could note that $[10] = [11 - 1] = [-1]$, and therefore, by a property true in every field, we have that

  $$[10] \cdot [10] = [-1] \cdot [-1] = [1] .$$

- To find the multiplicative inverse of [2], similarly we complete that row of the table of multiplication of $\mathbb{Z}_{11}$: that is, we consider all products of the form $[2] \cdot [l]$ with $[l] \neq [0]$, and see which one will equal [1].
  We have

  $$[2] \cdot [1] = [2], \quad [2] \cdot [2] = [4], \quad [2] \cdot [3] = [6], \quad [2] \cdot [4] = [8],$$
  $$[2] \cdot [5] = [10], \quad [2] \cdot [6] = [12] = [1],$$

  and here we can stop since we have found that [6] is the multiplicative inverse of [2] in $\mathbb{Z}_{11}$.

Finally, in the case of [8], we discuss one more method which is based on one way we have to prove Bézout's identity, via consecutive applications of Euclidean division. Recall that, since $\gcd(8, 11) = 1$, Bézout's identity tells us that we can find $\kappa, \ell \in \mathbb{Z}$ so that

$$1 = \kappa \cdot 11 + \ell \cdot 8.$$

Then, based on this we can conclude that the class $[\ell]$ which contains the integer ℓ is the multiplicative inverse of $[8]$ in $\mathbb{Z}_{11}$.

So, how do we find such integers κ and ℓ? We start by writing

$$11 = 1 \cdot 8 + 3.$$

We now continue by dividing 8 by 3:

$$8 = 2 \cdot 3 + 2$$

and observe that the remainder of this Euclidean division is smaller than the previous one. We continue by dividing 3 by 2:

$$3 = 1 \cdot 2 + 1,$$

and observe that we have finally found remainder 1.

We now 'reverse' this process: we can write

$$\begin{aligned} 1 &= 3 - 1 \cdot 2 \\ &= 3 - 1 \cdot (8 - 2 \cdot 3) = 3 + (-1) \cdot 8 + (-1) \cdot (-2) \cdot 3 = 3 \cdot 3 + (-1) \cdot 8 \\ &= 3 \cdot (11 - 1 \cdot 8) + (-1) \cdot 8 = 3 \cdot 11 + 3 \cdot (-1) \cdot 8 + (-1) \cdot 8 = 3 \cdot 11 + (-4) \cdot 8. \end{aligned}$$

Thus $[8]^{-1} = [-4] = [11 - 4] = [7]$.

What about when m is not a prime?

We have already concluded that $\mathbb{Z}_m$ is a commutative ring.
But is it a field?

The answer here is **NO**.

To justify it, recall one of the propositions we have already proven in class (see Lecture 3):

If $\mathbb{F}$ is a field,
then, for every $x, y \in \mathbb{F}$, if $x \cdot y = 0$, we have that $x = 0$ or $y = 0$.

Instinctively, we can say that this is equivalent to:

If can we find $x, y \in \mathbb{F}$ such that $x \cdot y = 0$ and $x \neq 0$, $y \neq 0$,
then $\mathbb{F}$ is **NOT** a field.

In Mathematics, the latter statement is called the **contrapositive** of the former statement *(and indeed, no matter what statement we start with, either both that statement and its contrapositive are true, or both of them are false)*.

How could we use this to show e.g. that $\mathbb{Z}_6$ is NOT a field?

Let's look at the table of multiplication for $\mathbb{Z}_6$:

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

Conclusion. Given that e.g. $[2] \cdot [3] = [0]$, while neither [2] nor [3] are equal to the additive identity, we get that $\mathbb{Z}_6$ **cannot be a field.**

Solving Systems of Linear Equations

Consider the following *system of linear equations*, or *linear system*, in 3 unknowns x_1, x_2, x_3 and with coefficients from $\mathbb{R}$:

$$\left\{ \begin{array}{rcl} 3x_1 - 6x_2 + 7x_3 & = & 0 \\ -x_1 + 3x_2 - 3x_3 & = & -2 \\ 2x_1 + 0x_2 + 4x_3 & = & 1 \end{array} \right\}.$$

How would we solve it? That is, how would we find the values for x_1, x_2 and x_3 that make all three equations true at the same time, **if any such values in $\mathbb{R}$ exist?**

What about the system

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & & & 2x_3 & = & 9 \end{array}\right\} ?$$

(again, assume that the coefficients are taken from $\mathbb{R}$)

Or the system

$$\left\{\begin{array}{rcrcrcr} \frac{1}{2}x_1 & & & & & = & -\frac{17}{4} \\ \frac{1}{2}x_1 & + & 3x_2 & & & = & -\frac{5}{4} \\ 2x_1 & & & + & 4x_3 & = & 1 \end{array}\right\} ?$$

What about the system

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ 0x_1 & + & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\}$$

with 3 linear equations in 4 unknowns

(i) if the coefficients are taken from $\mathbb{Q}$?

(ii) if the coefficients are taken from $\mathbb{Z}_5$?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 10

Friday September 18

Solving Systems of Linear Equations

Consider the following *system of linear equations*, or *linear system*, in 3 unknowns x_1, x_2, x_3 and with coefficients from $\mathbb{R}$:

$$\left\{ \begin{array}{rcl} 3x_1 - 6x_2 + 7x_3 & = & 0 \\ -x_1 + 3x_2 - 3x_3 & = & -2 \\ 2x_1 + 0x_2 + 4x_3 & = & 1 \end{array} \right\}.$$

How would we solve it? That is, how would we find the values for x_1, x_2 and x_3 that make all three equations true at the same time, **if any such values in $\mathbb{R}$ exist?**

What about the system

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & & & 2x_3 & = & 9 \end{array}\right\} ?$$

(again, assume that the coefficients are taken from $\mathbb{R}$)

Here we can use back substitution: from the last equation, we get $x_3 = 9/2$, and then we plug this into the second equation, and we get $x_2 - \frac{2}{3}(9/2) = -2 \Rightarrow x_2 = 1$, and finally we plug these into the first equation to get

$3x_1 - 6 + 7(9/2) = 0 \Rightarrow 3x_1 = -51/2 \Rightarrow x_1 = -17/2$.

Or the system

$$\left\{\begin{array}{rcrcrcr} \frac{1}{2}x_1 & & & & & = & -\frac{17}{4} \\ \frac{1}{2}x_1 & + & 3x_2 & & & = & -\frac{5}{4} \\ 2x_1 & & & + & 4x_3 & = & 1 \end{array}\right\} ?$$

We can use forward substitution: from the first equation, we get $x_1 = -17/2$, and then we plug this into the second equation, and we get

$\frac{1}{2}(-17/2) + 3x_2 = -\frac{5}{4} \Rightarrow x_2 = 1$, and finally we plug these into the last equation to get $2(-17/2) + 4x_3 = 1 \Rightarrow x_3 = 18/4$.

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ -x_1 & + & 3x_2 & - & 3x_3 & = & -2 \\ 2x_1 & & & + & 4x_3 & = & 1 \end{array}\right\} \quad \begin{array}{c} E_2+\frac{1}{3}E_1\to E_2' \\ E_3-\frac{2}{3}E_1\to E_3' \\ \Longleftrightarrow \end{array}$$

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & 4x_2 & - & \frac{2}{3}x_3 & = & 1 \end{array}\right\} \quad \begin{array}{c} E_3-4E_2\to E_3' \\ \Longleftrightarrow \end{array}$$

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & & & 2x_3 & = & 9 \end{array}\right\}.$$

Finally, the last system, which is *equivalent* to the first one *(that is, the two systems have the same set of solutions)*, is one of the systems we discussed in the previous slide, which we can solve using back substitution.

What about the system

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ 0x_1 & + & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\}$$

with 3 linear equations in 4 unknowns

(i) if the coefficients are taken from $\mathbb{Q}$?

(ii) if the coefficients are taken from $\mathbb{Z}_5$?

Coefficients from $\mathbb{Q}$:

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \quad \overset{E_2+\frac{1}{2}E_1\to E_2'}{\Longleftrightarrow}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & \frac{9}{2}x_2 & - & \frac{5}{2}x_3 & + & \frac{9}{2}x_4 & = & \frac{7}{2} \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \quad \overset{E_3-\frac{2}{9}E_2\to E_3'}{\Longleftrightarrow}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & \frac{9}{2}x_2 & - & \frac{5}{2}x_3 & + & \frac{9}{2}x_4 & = & \frac{7}{2} \\ & & & & \frac{41}{9}x_3 & - & 2x_4 & = & \frac{2}{9} \end{array}\right\}.$$

Finally, the last system, which is equivalent to the first one, can be solved as follows: for every value $\mu \in \mathbb{Q}$ that we choose for the variable x_4, we get a linear system in the unknowns x_1, x_2 and x_3 which we can then solve via back substitution.

In particular, we will have that: if $x_4 = \mu \in \mathbb{Q}$, then by the 3rd equation we get that $x_3 = \frac{18\mu+2}{41}$, which we can plug into the 2nd equation to also get $x_2 = \frac{33-31\mu}{41}$, and finally, combining all these with the 1st equation, we also obtain that $x_1 = \frac{17\mu-30}{41}$.

Clearly, we have infinitely many different solutions here.

Coefficients from $\mathbb{Z}_5$:

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{3E_1 \to E_1'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{E_2+E_1 \to E_2'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & 2x_2 & & & + & 2x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{3E_2 \to E_2'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & & & + & x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{E_3 - E_2 \to E_3'}$$

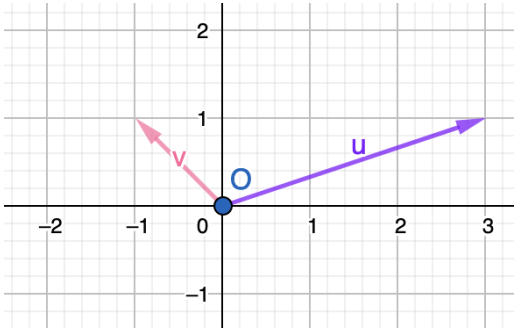
$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & & & + & x_4 & = & 3 \\ & & & & 4x_3 & - & 2x_4 & = & 3 \end{array}\right\}.$$

Looking at the the last system, it's clear that, for every value $\mu \in \mathbb{Z}_5$ that we choose for the variable x_4, we get a linear system in the unknowns x_1, x_2 and x_3 which we can then solve via back substitution. In particular, the solution set of the system is the following 4-tuples: $(x_1, x_2, x_3, x_4) = (-3\mu,\ 3-\mu,\ 2-2\mu,\ \mu), \quad \mu \in \mathbb{Z}_5$.

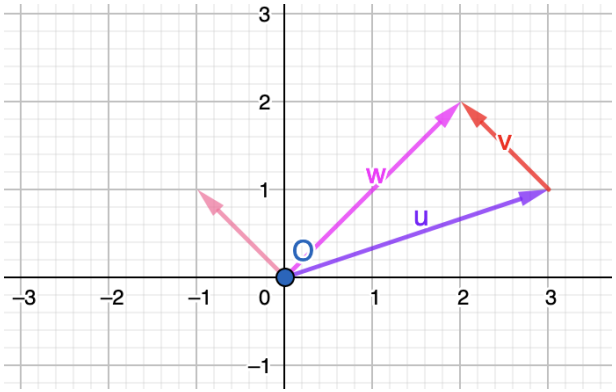
Note that here we get 5 different solutions.

Back to familiar notions: Geometry in $\mathbb{R}^2$

Consider the Cartesian plane:



If we are given two vectors on the plane with initial point the origin O, what can we do with them? We must have seen (e.g. in Physics) that we can add them:



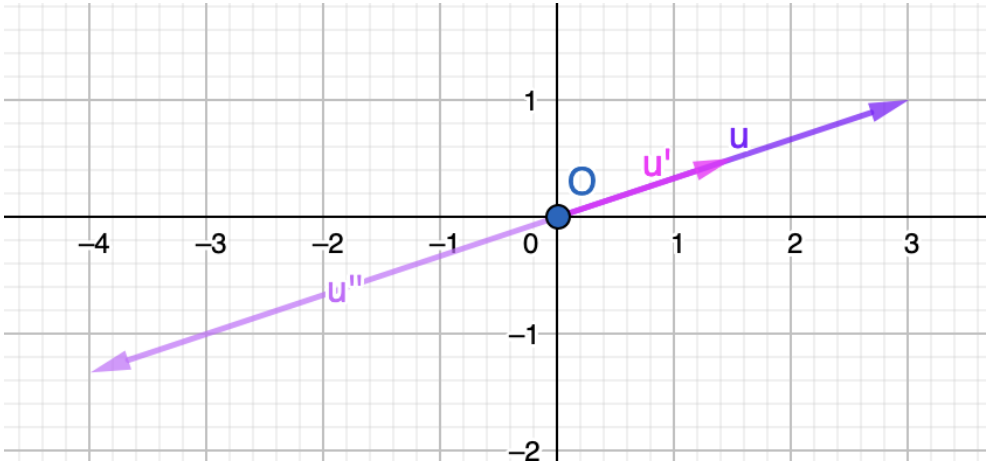
Important Observation. Note that, if we identify the vectors u and v (as they were initially drawn, to start at the origin) **with their terminal points** (and we think of each such point as its ordered pair of coordinates), then this rule for vector addition corresponds to doing coordinate-wise addition!

Back to familiar notions: Geometry in $\mathbb{R}^2$

What else can we do with vectors drawn on the plane?

Given that we identify them with ordered pairs, we could also consider doing coordinate-wise multiplication, or define some other rule for multiplication.

However, what we might have seen before (in courses such as Physics again) is not this type of operations, but another operation: *multiplication of a vector by a scalar r*, that is, multiplying **each of the coordinates** of the vector / ordered pair **by the same number *r***.



Here $u' = \frac{1}{2} \cdot u$, while $u'' = -\frac{4}{3} \cdot u$.

Definition of the notion of 'vector space'

Definition. Let $\mathbb{F}$ be a field. A set V is called a *vector space over* $\mathbb{F}$ if V is taken together with two operations/functions of the following form:

$$\textit{vector addition} \quad (\bar{x}, \bar{y}) \in V \times V \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$\textit{scalar multiplication} \quad (r, \bar{x}) \in \mathbb{F} \times V \ \mapsto \ r \cdot \bar{x} \in V$$

which satisfy the following properties:

(i) for all $\bar{x}, \bar{y} \in V$, $\quad \bar{x} + \bar{y} = \bar{y} + \bar{x}$ *(commutativity)*

(ii) for all $\bar{x}, \bar{y}, \bar{z} \in V$, $\quad (\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z})$ *(associativity)*

(iii) there exists an element $\bar{0}$ in V such that

$$\text{for all } \bar{x} \in V, \quad \bar{0} + \bar{x} = \bar{x} + \bar{0} = \bar{x}$$

(zero vector, or neutral element of vector addition)

(iv) for every $\bar{x} \in V$, there exists an element $\bar{w} = \bar{w}_x$ in V such that

$$\bar{x} + \bar{w} = \bar{w} + \bar{x} = \bar{0}$$

(negative or additive inverse of $\bar{x}$) usually denoted by $-\bar{x}$

(to be continued...)

Definition of the notion of 'vector space' (cont.)

(v) for all $\lambda, \mu \in \mathbb{F}$ and all $\bar{x} \in V$,

$$(\lambda \cdot \mu) \cdot \bar{x} = \lambda \cdot (\mu \cdot \bar{x})$$

(associativity of scalar multiplication)

(vi) for all $\bar{x} \in V$, $\quad 1 \cdot \bar{x} = \bar{x}$

(where 1 is the multiplicative identity in the field $\mathbb{F}$)

and finally

(vii) for every $\lambda \in \mathbb{F}$ and for all $\bar{x}, \bar{y} \in V$,

$$\lambda \cdot (\bar{x} + \bar{y}) = \lambda \cdot \bar{x} + \lambda \cdot \bar{y}$$

(scalar multiplication distributes over vector addition)

(viii) for all $\lambda, \mu \in \mathbb{F}$ and for every $\bar{x} \in V$,

$$(\lambda + \mu) \cdot \bar{x} = \lambda \cdot \bar{x} + \mu \cdot \bar{x}$$

(scalar multiplication distributes over the addition in $\mathbb{F}$)

The last two properties are called the Distributive Laws, and they relate the operation of vector addition or of scalar addition with the operation of scalar multiplication.

Terminology

Traditionally, the elements of a vector space are called *vectors*, while the elements of the associated field are called *scalars* (hence, often we also call the associated field $\mathbb{F}$ the *scalar field* or the *field of scalars*).

Basic Examples

Eventually we will see that, not only are the following examples the "easiest" to start with, but they do allow us to develop (almost) the entire theory while working with them, so essentially we could restrict our attention only to these.

Given an integer $n \geqslant 1$, consider the set $\mathbb{R}^n$ to be the set of all n-tuples of the form $(x_1, x_2, \ldots, x_n)$ with $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Following the standard convention, most of the time we will write the elements of $\mathbb{R}^n$ as column vectors instead of row vectors *(but we might also deviate from this sometimes)*:

$$\bar{x} \in \mathbb{R}^n, \qquad \bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} .$$

Very Important Remark. A vector is determined by (the ordered sequence of) its components $x_1, x_2, \ldots, x_n$. In other words,

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} \quad \text{and} \quad \bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} \quad \text{are equal}$$

if and only if $x_1 = y_1$ and $x_2 = y_2$ $\cdots$ and finally $x_n = y_n$.

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

1. Let $\bar{x}, \bar{y} \in \mathbb{R}^n$. Do we have

$$\bar{x} + \bar{y} = \bar{y} + \bar{x} \ ?$$

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

2. Let $\bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^n$. Do we have

$$(\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z}) \ ?$$

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

3. Is there a zero vector $\bar{0}$ in $\mathbb{R}^n$? That is, is there a neutral element for the vector addition we defined?

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

What other sub-questions should we pose and try to give an answer to in order to answer the Key Question?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 11

Monday September 21

Reminder: The notion of 'vector space'

Definition. Let $\mathbb{F}$ be a field. A set V is called a *vector space over* $\mathbb{F}$ if V is taken together with two operations/functions of the following form:

$$\textit{vector addition} \quad (\bar{x}, \bar{y}) \in V \times V \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$\textit{scalar multiplication} \quad (r, \bar{x}) \in \mathbb{F} \times V \ \mapsto \ r \cdot \bar{x} \in V$$

which satisfy the following properties:

(i) for all $\bar{x}, \bar{y} \in V$, $\quad \bar{x} + \bar{y} = \bar{y} + \bar{x}$ *(commutativity)*

(ii) for all $\bar{x}, \bar{y}, \bar{z} \in V$, $\quad (\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z})$ *(associativity)*

(iii) there exists an element $\bar{0}$ in V such that

$$\text{for all } \bar{x} \in V, \quad \bar{0} + \bar{x} = \bar{x} + \bar{0} = \bar{x}$$

(zero vector, or neutral element of vector addition)

(iv) for every $\bar{x} \in V$, there exists an element $\bar{w} = \bar{w}_x$ in V such that

$$\bar{x} + \bar{w} = \bar{w} + \bar{x} = \bar{0}$$

(negative or additive inverse of $\bar{x}$) usually denoted by $-\bar{x}$

(to be continued...)

Definition of the notion of 'vector space' (cont.)

(v) for all $\lambda, \mu \in \mathbb{F}$ and all $\bar{x} \in V$,

$$(\lambda \cdot \mu) \cdot \bar{x} = \lambda \cdot (\mu \cdot \bar{x})$$

(associativity of scalar multiplication)

(vi) for all $\bar{x} \in V$, $\quad 1 \cdot \bar{x} = \bar{x}$

(where 1 is the multiplicative identity in the field $\mathbb{F}$)

and finally

(vii) for every $\lambda \in \mathbb{F}$ and for all $\bar{x}, \bar{y} \in V$,

$$\lambda \cdot (\bar{x} + \bar{y}) = \lambda \cdot \bar{x} + \lambda \cdot \bar{y}$$

(scalar multiplication distributes over vector addition)

(viii) for all $\lambda, \mu \in \mathbb{F}$ and for every $\bar{x} \in V$,

$$(\lambda + \mu) \cdot \bar{x} = \lambda \cdot \bar{x} + \mu \cdot \bar{x}$$

(scalar multiplication distributes over the addition in $\mathbb{F}$)

The last two properties are called the Distributive Laws, and they relate the operation of vector addition or of scalar addition with the operation of scalar multiplication.

Terminology

Traditionally, the elements of a vector space are called *vectors*, while the elements of the associated field are called *scalars* (hence, often we also call the associated field $\mathbb{F}$ the *scalar field* or the *field of scalars*).

Basic Examples

Eventually we will see that, not only are the following examples the "easiest" to start with, but they do allow us to develop (almost) the entire theory while working with them, so essentially we could restrict our attention only to these.

Given an integer $n \geqslant 1$, consider the set $\mathbb{R}^n$ to be the set of all n-tuples of the form $(x_1, x_2, \ldots, x_n)$ with $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Following the standard convention, most of the time we will write the elements of $\mathbb{R}^n$ as column vectors instead of row vectors *(but we might also deviate from this sometimes)*:

$$\bar{x} \in \mathbb{R}^n, \qquad \bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} .$$

Very Important Remark. A vector is determined by (the ordered sequence of) its components $x_1, x_2, \ldots, x_n$. In other words,

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} \quad \text{and} \quad \bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix} \quad \text{are equal}$$

if and only if $x_1 = y_1$ and $x_2 = y_2$ $\cdots$ and finally $x_n = y_n$.

Basic Examples

We define addition of two elements in $\mathbb{R}^n$ as follows:

$$\bar{x} + \bar{y} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix} := \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \ddots \\ x_{n-1}+y_{n-1} \\ x_n+y_n \end{pmatrix}$$

(that is, we consider coordinate-wise addition),

and we define multiplication of an element of $\mathbb{R}^n$ by a real number r (a scalar) as follows:

$$r \cdot \bar{x} = r \cdot \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} := \begin{pmatrix} rx_1 \\ rx_2 \\ \ddots \\ rx_{n-1} \\ rx_n \end{pmatrix} .$$

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

1. Let $\bar{x}, \bar{y} \in \mathbb{R}^n$. Do we have

$$\bar{x} + \bar{y} = \bar{y} + \bar{x} \ ?$$

The answer is yes. Note that, by the definition of the vector addition that we are considering,

$$\bar{x} + \bar{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_{n-1}+y_{n-1} \\ x_n+y_n \end{pmatrix},$$

$$\text{while} \qquad \bar{y} + \bar{x} = \begin{pmatrix} y_1+x_1 \\ y_2+x_2 \\ \vdots \\ y_{n-1}+x_{n-1} \\ y_n+x_n \end{pmatrix}.$$

But because of the **commutativity of addition in** $\mathbb{R}$, we have that

$$x_1 + y_1 = y_1 + x_1, \quad x_2 + y_2 = y_2 + x_2, \ldots\ldots\ldots, \quad x_n + y_n = y_n + x_n.$$

Therefore, the vectors $\bar{x} + \bar{y}$ and $\bar{y} + \bar{x}$ have equal corresponding components, and hence they must be equal.

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

2. Let $\bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^n$. Do we have

$$(\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z}) \ ?$$

The answer is yes. Note that, by the definition of the vector addition that we are considering,

$$(\bar{x} + \bar{y}) + \bar{z} = \left(\begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix} \right) + \begin{pmatrix} z_1 \\ z_2 \\ \ddots \\ z_{n-1} \\ z_n \end{pmatrix}$$

$$= \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \ddots \\ x_{n-1}+y_{n-1} \\ x_n+y_n \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \ddots \\ z_{n-1} \\ z_n \end{pmatrix} = \begin{pmatrix} (x_1+y_1)+z_1 \\ (x_2+y_2)+z_2 \\ \ddots \\ (x_{n-1}+y_{n-1})+z_{n-1} \\ (x_n+y_n)+z_n \end{pmatrix},$$

$$\text{while} \qquad \bar{x} + (\bar{y} + \bar{z}) = \begin{pmatrix} x_1+(y_1+z_1) \\ x_2+(y_2+z_2) \\ \ddots \\ x_{n-1}+(y_{n-1}+z_{n-1}) \\ x_n+(y_n+z_n) \end{pmatrix}.$$

But because of the **associativity of addition in** $\mathbb{R}$, we have that

$$(x_1 + y_1) + z_1 = x_1 + (y_1 + z_1), \quad (x_2 + y_2) + z_2 = x_2 + (y_2 + z_2), \ \ldots\ldots,$$
$$(x_{n-1} + y_{n-1}) + z_{n-1} = x_{n-1} + (y_{n-1} + z_{n-1}), \quad (x_n + y_n) + z_n = (x_n + y_n) + z_n.$$

Therefore, the vectors $(\bar{x} + \bar{y}) + \bar{z}$ and $\bar{x} + (\bar{y} + \bar{z})$ have equal corresponding components, and hence they must be equal.

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

3. Is there a zero vector $\bar{0}$ in $\mathbb{R}^n$? That is, is there a neutral element for the vector addition we defined?

The answer is yes. Consider the vector in $\mathbb{R}^n$ all of whose components are equal to 0, the neutral element of addition in $\mathbb{R}$.

Then, for every $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} \in \mathbb{R}^n$, we have that

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x_1+0 \\ x_2+0 \\ \vdots \\ x_{n-1}+0 \\ x_n+0 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix},$$

and similarly $$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 0+x_1 \\ 0+x_2 \\ \vdots \\ 0+x_{n-1} \\ 0+x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}.$$

We have verified that the neutral element of vector addition in $\mathbb{R}^n$ is the vector

$$\bar{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

Key Question. Is $\mathbb{R}^n$ with these operations a vector space over $\mathbb{R}$?

What other sub-questions should we pose and try to give an answer to in order to answer the Key Question?

The vector spaces $\mathbb{F}^n$

Completely analogously we can check that $\mathbb{Z}_5^n$, the set of n-tuples of the form $(a_1, a_2, \ldots, a_n)$ with $a_1, a_2, \ldots, a_n \in \mathbb{Z}_5$, together with coordinate-wise addition and coordinate-wise scalar multiplication, is a vector space over $\mathbb{Z}_5$. *Exercise: confirm this yourselves.*

In fact, **no matter which field $\mathbb{F}$ we start with**, we can reach the same conclusion regarding the set $\mathbb{F}^n$ (viewed together with the analogously defined operations of vector addition '$+$' and scalar multiplication '$\cdot$'): the structure $\left(\mathbb{F}^n, +, \cdot\right)$ is a vector space over the field $\mathbb{F}$.

The vector spaces $\mathbb{F}^n$

Important Terminology. We say that $\mathbb{F}^n$ has **dimension** n over $\mathbb{F}$. *(the concept of dimension will be defined properly later in this term)*

Conventions to pay attention to. When we are given vectors in $\mathbb{F}^n$ (where $\mathbb{F}$ is some field), it is always assumed that any scalars we have to use are also from $\mathbb{F}$, unless otherwise specified.

If $\bar{x}$ is an element of $\mathbb{F}^n$, we sometimes say that $\bar{x}$ is an n-dimensional vector.

Assuming that we are given a number of vectors from one or more spaces of the type $\mathbb{F}^n$ (with $\mathbb{F}$ being a fixed field), we can only add two such vectors if they are of the same dimension.

Examples

Ex1. Consider the vectors $\bar{x} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 0 \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} 10 \\ -2 \\ 8 \\ 3 \end{pmatrix}$ in $\mathbb{R}^4$.

What is the vector $3\bar{x} + 9\bar{y}$ equal to?

Answer. We have

$$3\begin{pmatrix} 2 \\ 3 \\ 5 \\ 0 \end{pmatrix} + 9\begin{pmatrix} 10 \\ -2 \\ 8 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 9 \\ 15 \\ 0 \end{pmatrix} + \begin{pmatrix} 90 \\ -18 \\ 72 \\ 27 \end{pmatrix} = \begin{pmatrix} 96 \\ -9 \\ 87 \\ 27 \end{pmatrix}.$$

Terminology. What we just found is called a *linear combination* of the vectors $\bar{x}$ and $\bar{y}$.

Examples (cont.)

Ex2. *(Past Midterm Exam Problem)* Consider the following vectors, all of which have real coordinates:

$$\bar{a} = \begin{pmatrix} 1 \\ -3 \\ 0 \\ 4 \end{pmatrix}, \ \bar{b} = \begin{pmatrix} 2 \\ -2 \\ 3 \\ -5 \end{pmatrix}, \ \bar{c} = \begin{pmatrix} 6 \\ 3 \\ 9 \\ -1 \\ 8 \\ -1 \end{pmatrix}, \ \bar{d} = \begin{pmatrix} -7 \\ 4 \\ 5 \\ 4 \\ -1 \end{pmatrix},$$

$$\bar{u} = \begin{pmatrix} -4 \\ 3 \\ 9 \end{pmatrix}, \ \bar{v} = \begin{pmatrix} 5 \\ 4 \\ -2 \end{pmatrix}, \ \bar{w} = \begin{pmatrix} 0 \\ 3 \\ 5 \\ -2 \end{pmatrix}.$$

For every pair of two different vectors from this list, find the sum of those two vectors, **if it makes sense**.

Examples (cont.)

Answer to Ex2. We observe that the vectors $\bar{a}$, $\bar{b}$ and $\bar{w}$ are 4-dimensional, so we can add any two of those. We have

$$\bar{a} + \bar{b} = \bar{b} + \bar{a} = \begin{pmatrix} 3 \\ -5 \\ 3 \\ -1 \end{pmatrix}, \qquad \bar{a} + \bar{w} = \bar{w} + \bar{a} = \begin{pmatrix} 1 \\ 0 \\ 5 \\ 2 \end{pmatrix}$$

$$\text{and} \quad \bar{b} + \bar{w} = \bar{w} + \bar{b} = \begin{pmatrix} 2 \\ 1 \\ 8 \\ -7 \end{pmatrix}.$$

Moreover, the vectors $\bar{u}$ and $\bar{v}$ are both 3-dimensional, so we can consider their sum:

$$\bar{u} + \bar{v} = \bar{v} + \bar{u} = \begin{pmatrix} 1 \\ 7 \\ 7 \end{pmatrix}.$$

We finally note that the vector $\bar{c}$ is the only 6-dimensional vector in this list, so we cannot add it to any of the other vectors, while similarly $\bar{d}$ is the only 5-dimensional vector.

Scalar multiples, linear combinations and linear span

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over the field $\mathbb{F}$. Suppose $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ are vectors in V, and also that S is a non-empty subset of V (S may be a finite set, or an infinite set).

- *(Scalar Multiples)* Any vector of the form $\lambda \cdot \bar{x}_1$, where λ is an element of $\mathbb{F}$ (so in other words a *scalar*), is called a *scalar multiple* of $\bar{x}_1$.
- *(Linear Combinations)* Any expression / vector of the form

$$\mu_1 \bar{x}_1 + \mu_2 \bar{x}_2 + \cdots + \mu_k \bar{x}_k,$$

  where $\mu_1, \mu_2, \ldots, \mu_k$ are scalars, is called a *linear combination* of the vectors $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$.

Scalar multiples, linear combinations and linear span

Remark. Recall that vector addition is a binary operation (that is, its inputs are pairs of elements of V); thus, when we want to add k vectors in a row, formally we should understand that one application of vector addition happens first, then a third vector is added to the previous sum, and so on; of course, **because vector addition is required to be associative**, we can confirm (using mathematical induction too) that it really doesn't matter in which order we will do the operations (and so we can completely omit parentheses).

Similarly, **because vector addition is required to be commutative**, we can confirm (using mathematical induction too) that it also doesn't matter in which order we write the vectors we want to add.

We will call these very useful properties, which, as we said, can be justified from the vector space axioms, the properties of *generalised associativity* and *generalised commutativity* respectively, and we will be free to use them from now on (unless otherwise specified in a certain problem).

Note that analogous properties can be confirmed to hold in other structures too, and for other operations that we have already seen *(try to give one or two examples here)*.

Scalar multiples, linear combinations and linear span (cont.)

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over the field $\mathbb{F}$. Suppose $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ are vectors in V, and also that S is a non-empty subset of V (S may be a finite set, or an infinite set).

- *(Linear Span of k vectors)* The set of **all** linear combinations of $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ is called the *linear span* of $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, and is denoted by

  $$\operatorname{span}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k).$$

  In other words,

  $$\operatorname{span}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k) = \left\{\mu_1\bar{x}_1 + \mu_2\bar{x}_2 + \cdots + \mu_k\bar{x}_k : \mu_1, \mu_2, \ldots, \mu_k \in \mathbb{F}\right\}.$$

- *(Linear Span of a set)* Finally, given the non-empty subset S of V, we define the *linear span* of S to be the set of **all** linear combinations of any k vectors from S, for every integer $k \geqslant 1$. We denote this set by $\operatorname{span}(S)$.

  In other words,

  $$\operatorname{span}(S) = \left\{\mu_1\bar{y}_1 + \mu_2\bar{y}_2 + \cdots + \mu_k\bar{y}_k : k \geqslant 1,\ \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k \in S,\ \mu_1, \mu_2, \ldots, \mu_k \in \mathbb{F}\right\}.$$

Very Important Remark. We can only consider sums / linear combinations of **finitely many vectors at a time**, so even if we start with an infinite set S, its span contains only finite sums of (scalar multiples of) its elements (although the number of summands doesn't have to be the same in all these sums).

Examples (cont.)

Ex3. Consider the vectors $\bar{x} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ 4.5 \\ 6 \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} 2 \\ -4 \\ 7 \\ -1 \\ 3 \end{pmatrix}$ in $\mathbb{R}^5$.

Is the vector $\begin{pmatrix} 5 \\ 6 \\ 1 \\ -10 \\ -9 \end{pmatrix}$ a linear combination of $\bar{x}$ and $\bar{y}$?

Answer. We have to check whether there are $\lambda, \mu \in \mathbb{R}$ such that

$$\begin{pmatrix} 5 \\ 6 \\ 1 \\ -10 \\ -9 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2 \\ 3 \\ 4.5 \\ 6 \end{pmatrix} + \mu \begin{pmatrix} 2 \\ -4 \\ 7 \\ -1 \\ 3 \end{pmatrix} .$$

But

$$\lambda \begin{pmatrix} 1 \\ -2 \\ 3 \\ 4.5 \\ 6 \end{pmatrix} + \mu \begin{pmatrix} 2 \\ -4 \\ 7 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} \lambda \cdot 1 \\ \lambda \cdot (-2) \\ \lambda \cdot 3 \\ \lambda \cdot 4.5 \\ \lambda \cdot 6 \end{pmatrix} + \begin{pmatrix} \mu \cdot 2 \\ \mu \cdot (-4) \\ \mu \cdot 7 \\ \mu \cdot (-1) \\ \mu \cdot 3 \end{pmatrix} = \begin{pmatrix} \lambda + 2\mu \\ (-2)\lambda + (-4)\mu \\ 3\lambda + 7\mu \\ 4.5\lambda + (-1)\mu \\ 6\lambda + 3\mu \end{pmatrix} .$$

Also we recall that for the vector $\begin{pmatrix} 5 \\ 6 \\ 1 \\ -10 \\ -9 \end{pmatrix}$ to be equal to a vector of the form we just found, **we need all the corresponding components to be equal.**

Answer to Ex3 (cont.)

That is, in order to have

$$\begin{pmatrix} \lambda + 2\mu \\ (-2)\lambda + (-4)\mu \\ 3\lambda + 7\mu \\ 4.5\lambda + (-1)\mu \\ 6\lambda + 3\mu \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 1 \\ -10 \\ -9 \end{pmatrix},$$

we need all the following equations to be satisfied at the same time:

$$\begin{array}{rcrcr} \lambda & + & 2\mu & = & 5 \\ -2\lambda & - & 4\mu & = & 6 \\ 3\lambda & + & 7\mu & = & 1 \\ 4.5\lambda & - & \mu & = & -10 \\ 6\lambda & + & 3\mu & = & -9 \end{array}$$

In other words, we need to find a solution for the above system of linear equations. **We now note that the LHS of the 2nd equation is -2 times the LHS of the 1st equation, while the same is not true for their respective RHS.** This shows that we won't be able to find any values for λ and μ that will satisfy both the 1st and the 2nd equation at the same time, so clearly we won't be able to find a solution to the entire system either.

We conclude that $\begin{pmatrix} 5 \\ 6 \\ 1 \\ -10 \\ -9 \end{pmatrix}$ is **not** a linear combination of $\bar{x}$ and $\bar{y}$.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 12

Tuesday September 22

Important Remark
(based on last example of previous lecture)

Let $\mathbb{F}$ be a field, and consider the vector space $\mathbb{F}^n$ over $\mathbb{F}$.

Assume that you have been given vectors $\bar{u}$, $\bar{a}_1, \ldots, \bar{a}_k$ of $\mathbb{F}^n$, and for each $1 \leqslant i \leqslant k$ write $a_{i,j}$ for the j-th component of the vector $\bar{a}_i$; that is,

$$\bar{a}_i = \begin{pmatrix} a_{i,1} \\ a_{i,2} \\ \vdots \\ a_{i,n-1} \\ a_{i,n} \end{pmatrix} .$$

The following two questions are 'equivalent' *(that is, if the answer to the first question is affirmative, then so is the answer to the second question, and vice versa):*

Question 1. Is the vector $\bar{u}$ a linear combination of the vectors $\bar{a}_1, \ldots, \bar{a}_k$? In other words, is $\bar{u}$ contained in $\mathrm{span}(\bar{a}_1, \ldots, \bar{a}_k)$?

Question 2. Does the following system of linear equations (with coefficients from $\mathbb{F}$ and k unknowns) have a solution?

$$\left\{ \begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{2,1}x_2 & + & \cdots\cdots & + & a_{k,1}x_k & = & u_1 \\ a_{1,2}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{k,2}x_k & = & u_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{1,n-1}x_1 & + & a_{2,n-1}x_2 & + & \cdots\cdots & + & a_{k,n-1}x_k & = & u_{n-1} \\ a_{1,n}x_1 & + & a_{2,n}x_2 & + & \cdots\cdots & + & a_{k,n}x_k & = & u_n \end{array} \right\}$$

Substructures of ‘nice’ structures

Case of fields: Subfields

Definition. Let $\mathbb{F} = \big(\{\text{elements in } \mathbb{F}\},\, +,\, \cdot\big)$ be a field. Given a (non-empty) subset $\mathbb{K}$ of $\mathbb{F}$, we can consider the restrictions of the operations of addition and of multiplication in $\mathbb{F}$ to pairs of elements from $\mathbb{K}$:

$$(x, y) \in \mathbb{K} \times \mathbb{K} \;\mapsto\; x + y \in \mathbb{F}\,,$$
$$(x, y) \in \mathbb{K} \times \mathbb{K} \;\mapsto\; x \cdot y \in \mathbb{F}\,.$$

- We say that $\mathbb{K}$ is *closed under the addition $+$ in $\mathbb{F}$* if, for all $x, y \in \mathbb{K}$, we have that $x + y \in \mathbb{K}$ as well. Similarly, we say that $\mathbb{K}$ is *closed under the multiplication $\cdot$ in $\mathbb{F}$* if, for all $x, y \in \mathbb{K}$, we have that $x \cdot y \in \mathbb{K}$ as well.

- If $\mathbb{K}$ is closed under both the addition and the multiplication in $\mathbb{F}$, then we say that $\mathbb{K}$ is a subfield of $\mathbb{F}$ if $\mathbb{K}$ together with the restricted operations is a field itself.

Example. In HW1, Problem 4, you had to show that $\mathbb{Q}(\sqrt{17})$ (together with the standard operations of addition and of multiplication of real numbers) is a subfield of $\mathbb{R}$.

Question. Given a field $\mathbb{F} = \big(\{\text{elements in } \mathbb{F}\}, +, \cdot\big)$, and a subset $\mathbb{K}$ of $\mathbb{F}$, how do we check that $\mathbb{K}$ is (or rather, can be viewed as) a subfield of $\mathbb{F}$?

Case of commutative rings: Commutative Subrings

Definition. Let $\mathcal{R} = \big(\{\text{elements in } \mathcal{R}\},\, +,\, \cdot\big)$ be a commutative ring. Given a (non-empty) subset $\mathcal{S}$ of $\mathcal{R}$, we can consider the restrictions of the operations of addition and of multiplication in $\mathcal{R}$ to pairs of elements from $\mathcal{S}$:

$$(x, y) \in \mathcal{S} \times \mathcal{S} \;\mapsto\; x + y \in \mathcal{R}\,,$$
$$(x, y) \in \mathcal{S} \times \mathcal{S} \;\mapsto\; x \cdot y \in \mathcal{R}\,.$$

- We say that $\mathcal{S}$ is *closed under the addition $+$ in $\mathcal{R}$* if, for all $x, y \in \mathcal{S}$, we have that $x + y \in \mathcal{S}$ as well. Similarly, we say that $\mathcal{S}$ is *closed under the multiplication $\cdot$ in $\mathcal{R}$* if, for all $x, y \in \mathcal{S}$, we have that $x \cdot y \in \mathcal{S}$ as well.

- If $\mathcal{S}$ is closed under both the addition and the multiplication in $\mathcal{R}$, then we say that $\mathcal{S}$ is a commutative subring of $\mathcal{R}$ if $\mathcal{S}$ together with the restricted operations is a commutative ring itself.

Question. Given a commutative ring $\mathcal{R} = \big(\{\text{elements in } \mathcal{R}\},\, +,\, \cdot\big)$, and a subset $\mathcal{S}$ of $\mathcal{R}$, how do we check that $\mathcal{S}$ is (or rather, can be viewed as) a commutative subring of $\mathcal{R}$?

Case of vector spaces: Subspaces

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Given a (non-empty) subset W of V, we can consider the restrictions of the operations of vector addition and of scalar multiplication in V to elements of W as follows:

$$(\bar{x}, \bar{y}) \in W \times W \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$(r, \bar{x}) \in \mathbb{F} \times W \ \mapsto \ r \cdot \bar{x} \in V$$

- We say that W is *closed under the vector addition in V* if, for all $\bar{x}, \bar{y} \in W$, we have that $\bar{x} + \bar{y} \in W$ as well. Similarly, we say that W is *closed under the scalar multiplication in V* if, for all $\bar{x} \in W$ and for all $r \in \mathbb{F}$, we have that $r \cdot \bar{x} \in W$ as well.

- If W is closed under both the vector addition and the scalar multiplication in V, then we say that W is a subspace of V if W together with the restricted operations is a vector space over $\mathbb{F}$ itself.

Important Theorem

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

The subspaces of V coincide with the different linear spans of subsets of V.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 13

Wednesday September 23

Substructures of ‘nice’ structures

Case of fields: Subfields

Definition. Let $\mathbb{F} = \big(\{\text{elements in } \mathbb{F}\},\, +,\, \cdot\big)$ be a field. Given a (non-empty) subset $\mathbb{K}$ of $\mathbb{F}$, we can consider the restrictions of the operations of addition and of multiplication in $\mathbb{F}$ to pairs of elements from $\mathbb{K}$:

$$(x, y) \in \mathbb{K} \times \mathbb{K} \;\mapsto\; x + y \in \mathbb{F}\,,$$
$$(x, y) \in \mathbb{K} \times \mathbb{K} \;\mapsto\; x \cdot y \in \mathbb{F}\,.$$

- We say that $\mathbb{K}$ is *closed under the addition $+$ in $\mathbb{F}$* if, for all $x, y \in \mathbb{K}$, we have that $x + y \in \mathbb{K}$ as well. Similarly, we say that $\mathbb{K}$ is *closed under the multiplication $\cdot$ in $\mathbb{F}$* if, for all $x, y \in \mathbb{K}$, we have that $x \cdot y \in \mathbb{K}$ as well.

- If $\mathbb{K}$ is closed under both the addition and the multiplication in $\mathbb{F}$, then we say that $\mathbb{K}$ is a subfield of $\mathbb{F}$ if $\mathbb{K}$ together with the restricted operations is a field itself.

Example. In HW1, Problem 4, you had to show that $\mathbb{Q}(\sqrt{17})$ (together with the standard operations of addition and of multiplication of real numbers) is a subfield of $\mathbb{R}$.

Question. Given a field $\mathbb{F} = \big(\{\text{elements in } \mathbb{F}\}, +, \cdot\big)$, and a subset $\mathbb{K}$ of $\mathbb{F}$, how do we check that $\mathbb{K}$ is (or rather, can be viewed as) a subfield of $\mathbb{F}$?

Sufficient to check that

- $\mathbb{K}$ is closed under the addition in $\mathbb{F}$;
- $\mathbb{K}$ is closed under the multiplication in $\mathbb{F}$;
- $0_{\mathbb{F}}$ is contained in $\mathbb{K}$;
- $\mathbb{K}$ is closed under taking additive inverses (that is, if $x \in \mathbb{K}$, then $-x$ (the additive inverse of x in $\mathbb{F}$) is also contained in $\mathbb{K}$);
- $1_{\mathbb{F}}$ is contained in $\mathbb{K}$;
- $\mathbb{K}$ is closed under taking multiplicative inverses (that is, if $x \in \mathbb{K}$ and $x \neq 0_{\mathbb{F}}$, then x^{-1} (the multiplicative inverse of x in $\mathbb{F}$) is also in $\mathbb{K}$).

Case of commutative rings: Commutative Subrings

Definition. Let $\mathcal{R} = \big(\{\text{elements in } \mathcal{R}\},\, +,\, \cdot\big)$ be a commutative ring. Given a (non-empty) subset $\mathcal{S}$ of $\mathcal{R}$, we can consider the restrictions of the operations of addition and of multiplication in $\mathcal{R}$ to pairs of elements from $\mathcal{S}$:

$$(x, y) \in \mathcal{S} \times \mathcal{S} \;\mapsto\; x + y \in \mathcal{R}\,,$$
$$(x, y) \in \mathcal{S} \times \mathcal{S} \;\mapsto\; x \cdot y \in \mathcal{R}\,.$$

- We say that $\mathcal{S}$ is *closed under the addition $+$ in $\mathcal{R}$* if, for all $x, y \in \mathcal{S}$, we have that $x + y \in \mathcal{S}$ as well. Similarly, we say that $\mathcal{S}$ is *closed under the multiplication $\cdot$ in $\mathcal{R}$* if, for all $x, y \in \mathcal{S}$, we have that $x \cdot y \in \mathcal{S}$ as well.

- If $\mathcal{S}$ is closed under both the addition and the multiplication in $\mathcal{R}$, then we say that $\mathcal{S}$ is a commutative subring of $\mathcal{R}$ if $\mathcal{S}$ together with the restricted operations is a commutative ring itself.

Question. Given a commutative ring $\mathcal{R} = \big(\{\text{elements in } \mathcal{R}\},\, +,\, \cdot\big)$, and a subset $\mathcal{S}$ of $\mathcal{R}$, how do we check that $\mathcal{S}$ is (or rather, can be viewed as) a commutative subring of $\mathcal{R}$?

Sufficient to check?

To verify whether a subset $\mathcal{S}$ of the commutative ring $\mathcal{R} = \big(\{\text{elements in } \mathcal{R}\}, +, \cdot\big)$ can be viewed as a commutative subring of $\mathcal{R}$ with the restricted operations of $+$ and $\cdot$,

- we need to check
- and it is sufficient to check that

- $\mathcal{S}$ is closed under the addition in $\mathcal{R}$;
- $\mathcal{S}$ is closed under the multiplication in $\mathcal{R}$;
- $0_{\mathcal{R}}$ is contained in $\mathcal{S}$;
- $\mathcal{S}$ is closed under taking additive inverses (that is, if $x \in \mathcal{S}$, then $-x$ (the additive inverse of x in $\mathcal{R}$) is also contained in $\mathcal{S}$);
- $1_{\mathcal{R}}$ is contained in $\mathcal{S}$.

Case of vector spaces: Subspaces

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Given a (non-empty) subset W of V, we can consider the restrictions of the operations of vector addition and of scalar multiplication in V to elements of W as follows:

$$(\bar{x}, \bar{y}) \in W \times W \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$(r, \bar{x}) \in \mathbb{F} \times W \ \mapsto \ r \cdot \bar{x} \in V$$

- We say that W is *closed under the vector addition in V* if, for all $\bar{x}, \bar{y} \in W$, we have that $\bar{x} + \bar{y} \in W$ as well. Similarly, we say that W is *closed under the scalar multiplication in V* if, for all $\bar{x} \in W$ and for all $r \in \mathbb{F}$, we have that $r \cdot \bar{x} \in W$ as well.

- If W is closed under both the vector addition and the scalar multiplication in V, then we say that W is a subspace of V if W together with the restricted operations is a vector space over $\mathbb{F}$ itself.

Sufficient to check?

Question. Let $\mathbb{F}$ be a field, and let $V = \big(\{\text{vectors in } V\}, +, \cdot\big)$ be a vector space over $\mathbb{F}$. Given a subset W of V, how do we check that W is (or rather, can be viewed as) a subspace of V?

- We **need to check**
- and it is **sufficient to check** that

- W is closed under the vector addition in V
  (that is, for every $\bar{x}$ and $\bar{y}$ in W, we have that $\bar{x} + \bar{y} \in W$ as well);
- W is closed under the scalar multiplication in V
  (that is, for every $\bar{x} \in W$ and every $\lambda \in \mathbb{F}$, we have that $\lambda \cdot \bar{x} \in W$ as well);
- $\bar{0}_V$ is contained in W.

Question. Why are these conditions sufficient to check?

Important Theorem

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Consider a subset W of V.

Then W is a subspace of V **if and only if** W is equal to the linear span $\mathrm{span}(S)$ of some subset S of V.

Remark. To prove Theorem 1, we need to show two things:

- every subspace W of V can be written as the linear span of some subset S of V;
- for every subset T of V, the set $\mathrm{span}(T)$ is a subspace of V.

Proof of Theorem 1

We first show that, if W is a subspace of V, then W can be written as the linear span of some subset S of V.

We will show that $W = \mathrm{span}(W)$.
To show this, we need to show two inclusions (why?): that $W \subseteq \mathrm{span}(W)$ and that $\mathrm{span}(W) \subseteq W$.

- For the first inclusion, we note that $1 \cdot \bar{x} = \bar{x}$ for every $\bar{x} \in W$, where 1 is the multiplicative identity of $\mathbb{F}$ (this is guaranteed by one of the axioms of vector space), and thus every element $\bar{x}$ of W can be written as a linear combination of elements of W. In other words, every element $\bar{x}$ of W is in $\mathrm{span}(W)$, and thus $W \subseteq \mathrm{span}(W)$.

- For the second inclusion, we use the facts that W is closed under vector addition and scalar multiplication. Based on these, we see that, for every $\bar{x}, \bar{y} \in W$ and every $\lambda, \mu \in \mathbb{F}$ we must have that

$$\lambda \cdot \bar{x} \text{ and } \mu \cdot \bar{y} \in W \text{ too,}$$

$$\text{and hence also } \lambda \cdot \bar{x} + \mu \cdot \bar{y} \in W \text{ as well.}$$

We can then continue like this and argue analogously about linear combinations of k vectors from W (with $k \geqslant 3$): for every $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \in W$ and every $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, we must have that

$$\lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 + \cdots + \lambda_k \bar{x}_k \in W \text{ as well.}$$

But this shows that every element of $\mathrm{span}(W)$ is also an element of W, and hence $\mathrm{span}(W) \subseteq W$. **Combining the above, we conclude that $W = \mathrm{span}(W)$.**

Proof of Theorem 1 (cont.)

We now have to show that, if T is a (non-empty) subset of V, then $\mathrm{span}(T)$ is a subspace of V.

We will discuss this in the next lecture.

Terminology

Given a field $\mathbb{F}$, and a vector space V over $\mathbb{F}$, we call any subspace W of V which can be spanned by exactly one non-zero vector of V a *line*.

In other words, the lines in V are sets containing all the scalar multiples of some non-zero vector $\bar{x}$ of V and only those elements; that is, sets of the form

$$\{\lambda \cdot \bar{x} : \lambda \in \mathbb{F}\}$$

where $\bar{x} \neq \bar{0}$.

Note that the terminology is inspired by settings such as $\mathbb{R}^2$ and $\mathbb{R}^3$, where the common visualisation of such sets is to draw them precisely as **straight lines passing through the origin** *(with the word 'line' having a geometric meaning now).*

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 14

Friday September 25

Substructures of ‘nice’ structures

Case of vector spaces: Subspaces

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Given a (non-empty) subset W of V, we can consider the restrictions of the operations of vector addition and of scalar multiplication in V to elements of W as follows:

$$(\bar{x}, \bar{y}) \in W \times W \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$(r, \bar{x}) \in \mathbb{F} \times W \ \mapsto \ r \cdot \bar{x} \in V$$

- We say that W is *closed under the vector addition in V* if, for all $\bar{x}, \bar{y} \in W$, we have that $\bar{x} + \bar{y} \in W$ as well. Similarly, we say that W is *closed under the scalar multiplication in V* if, for all $\bar{x} \in W$ and for all $r \in \mathbb{F}$, we have that $r \cdot \bar{x} \in W$ as well.

- If W is closed under both the vector addition and the scalar multiplication in V, then we say that W is a subspace of V if W together with the restricted operations is a vector space over $\mathbb{F}$ itself.

Sufficient to check?

Question. Let $\mathbb{F}$ be a field, and let $V = \big(\{\text{vectors in } V\}, +, \cdot\big)$ be a vector space over $\mathbb{F}$. Given a subset W of V, how do we check that W is (or rather, can be viewed as) a subspace of V?

- We **need to check**
- and it is **sufficient to check** that

- W is closed under the vector addition in V
  (that is, for every $\bar{x}$ and $\bar{y}$ in W, we have that $\bar{x} + \bar{y} \in W$ as well);
- W is closed under the scalar multiplication in V
  (that is, for every $\bar{x} \in W$ and every $\lambda \in \mathbb{F}$, we have that $\lambda \cdot \bar{x} \in W$ as well);
- $\bar{0}_V$ is contained in W.

Question. Why are these conditions sufficient to check?

Important Theorem

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Consider a subset W of V.

Then W is a subspace of V **if and only if** W is equal to the linear span $\mathrm{span}(S)$ of some subset S of V.

Remark. To prove Theorem 1, we need to show two things:

- every subspace W of V can be written as the linear span of some subset S of V;
- for every subset T of V, the set $\mathrm{span}(T)$ is a subspace of V.

Proof of Theorem 1

We first show that, if W is a subspace of V, then W can be written as the linear span of some subset S of V.

We will show that $W = \mathrm{span}(W)$.
To show this, we need to show two inclusions (why?): that $W \subseteq \mathrm{span}(W)$ and that $\mathrm{span}(W) \subseteq W$.

- For the first inclusion, we note that $1 \cdot \bar{x} = \bar{x}$ for every $\bar{x} \in W$, where 1 is the multiplicative identity of $\mathbb{F}$ (this is guaranteed by one of the axioms of vector space), and thus every element $\bar{x}$ of W can be written as a linear combination of elements of W. In other words, every element $\bar{x}$ of W is in $\mathrm{span}(W)$, and thus $W \subseteq \mathrm{span}(W)$.

- For the second inclusion, we use the facts that W is closed under vector addition and scalar multiplication. Based on these, we see that, for every $\bar{x}, \bar{y} \in W$ and every $\lambda, \mu \in \mathbb{F}$ we must have that

$$\lambda \cdot \bar{x} \text{ and } \mu \cdot \bar{y} \in W \text{ too,}$$

$$\text{and hence also } \lambda \cdot \bar{x} + \mu \cdot \bar{y} \in W \text{ as well.}$$

We can then continue like this and argue analogously about linear combinations of k vectors from W (with $k \geqslant 3$): for every $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \in W$ and every $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, we must have that

$$\lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 + \cdots + \lambda_k \bar{x}_k \in W \text{ as well.}$$

But this shows that every element of $\mathrm{span}(W)$ is also an element of W, and hence $\mathrm{span}(W) \subseteq W$. **Combining the above, we conclude that $W = \mathrm{span}(W)$.**

Proof of Theorem 1 (cont.)

We now have to show that, if T is a (non-empty) subset of V, then $\mathrm{span}(T)$ is a subspace of V.

Consider such a non-empty subset T of V, and recall that, to show that $\mathrm{span}(T)$ is a subspace of V, it suffices to check that

1. $\bar{0}_V \in \mathrm{span}(T)$;
2. $\mathrm{span}(T)$ is closed under vector addition;
3. $\mathrm{span}(T)$ is closed under scalar multiplication.

To show 1., we note that, if we fix an element $\bar{x}$ of T, then, for every $\lambda \in \mathbb{F}$, we will have $\lambda \cdot \bar{x} \in \mathrm{span}(T)$ (since $\lambda \cdot \bar{x}$ is a linear combination of elements of T). In particular, $\bar{0}_V = 0_{\mathbb{F}} \cdot \bar{x}$ will be contained in $\mathrm{span}(T)$.

Proof of Theorem 1 (cont.)

To show 2., we consider arbitrary elements $\bar{u}, \bar{v}$ in $\mathrm{span}(T)$.

Given that our only assumption about $\bar{u}$ and $\bar{v}$ is that they belong to $\mathrm{span}(T)$, we can simply say that $\bar{u}$ and $\bar{v}$ are linear combinations of elements of T. In other words,

- there exist $k_1 \geqslant 1$, $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{k_1} \in T$ and $\lambda_1, \lambda_2, \ldots, \lambda_{k_1} \in \mathbb{F}$ such that

$$\bar{u} = \lambda_1 \cdot \bar{x}_1 + \lambda_2 \cdot \bar{x}_2 + \cdots + \lambda_{k_1} \cdot \bar{x}_{k_1} \,,$$

- and similarly, there exist $k_2 \geqslant 1$, $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{k_2} \in T$ and $\mu_1, \mu_2, \ldots, \mu_{k_2} \in \mathbb{F}$ such that

$$\bar{v} = \mu_1 \cdot \bar{y}_1 + \mu_2 \cdot \bar{y}_2 + \cdots + \mu_{k_2} \cdot \bar{y}_{k_2} \,.$$

But then

$$\begin{aligned} \bar{u} + \bar{v} &= \left(\lambda_1 \cdot \bar{x}_1 + \lambda_2 \cdot \bar{x}_2 + \cdots + \lambda_{k_1} \cdot \bar{x}_{k_1}\right) + \left(\mu_1 \cdot \bar{y}_1 + \mu_2 \cdot \bar{y}_2 + \cdots + \mu_{k_2} \cdot \bar{y}_{k_2}\right) \\ &= \lambda_1 \cdot \bar{x}_1 + \lambda_2 \cdot \bar{x}_2 + \cdots + \lambda_{k_1} \cdot \bar{x}_{k_1} \;+\; \mu_1 \cdot \bar{y}_1 + \mu_2 \cdot \bar{y}_2 + \cdots + \mu_{k_2} \cdot \bar{y}_{k_2} \,, \end{aligned}$$

which is a linear combination of $k_1 + k_2$ elements of T.

This shows that $\bar{u} + \bar{v} \in \mathrm{span}(T)$, as we wanted.

Proof of Theorem 1 (cont.)

Finally, to show 3., we consider again an arbitrary element $\bar{w}$ in $\mathrm{span}(T)$, as well as an arbitrary scalar $r \in \mathbb{F}$.

As before, we can find $k \geqslant 1$, $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_k \in T$ and $\nu_1, \nu_2, \ldots, \nu_k \in \mathbb{F}$ such that

$$\bar{w} = \nu_1 \cdot \bar{z}_1 + \nu_2 \cdot \bar{z}_2 + \cdots + \nu_k \cdot \bar{z}_k \, .$$

But then we can write

$$\begin{aligned} r \cdot \bar{w} &= r \cdot \big(\nu_1 \cdot \bar{z}_1 + \nu_2 \cdot \bar{z}_2 + \cdots + \nu_k \cdot \bar{z}_k\big) \\ &= r \cdot \big(\nu_1 \cdot \bar{z}_1\big) + r \cdot \big(\nu_2 \cdot \bar{z}_2\big) + \cdots + r \cdot \big(\nu_k \cdot \bar{z}_k\big) && \text{(by the distributive law)} \\ &= (r\nu_1) \cdot \bar{z}_1 + (r\nu_2) \cdot \bar{z}_2 + \cdots + (r\nu_k) \cdot \bar{z}_k \, , && \text{(by the associativity of scalar multiplication)} \end{aligned}$$

with the last expression being in the standard form of a linear combination of k elements from T.

This shows that $r \cdot \bar{w} \in \mathrm{span}(T)$, as we wanted.

Having shown 1., 2. and 3., we can conclude that $\mathrm{span}(T)$ is a subspace of V.

What about the case of T being the empty subset of V, $T = \emptyset$?

Important Convention

We are allowed to consider the *empty sum*, that is, the sum which has no summands. In this case, we agree that the result equals the neutral element of vector addition, that is, the zero vector $\bar{0}_V$ of V.

Based on this convention, we can now see that it makes sense to consider the linear span of the empty set $\emptyset$ too:

since $\emptyset$ contains no elements, the only finite sum of scalar multiples of elements of $\emptyset$ that we could consider is the empty sum.

This shows that the only element of $\operatorname{span}(\emptyset)$ is the zero vector:

$$\operatorname{span}(\emptyset) = \{\bar{0}_V\}.$$

Thus, we can rewrite the definition of the linear span of a subset T of V (including now the case where $T = \emptyset$):

$$\begin{aligned}\operatorname{span}(T) &= \{\bar{0}_V\} \cup \{\mu_1\bar{y}_1 + \mu_2\bar{y}_2 + \cdots + \mu_k\bar{y}_k : k \geqslant 1,\ \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k \in T,\ \mu_1, \mu_2, \ldots, \mu_k \in \mathbb{F}\} \\ &= \{\mu_1\bar{y}_1 + \mu_2\bar{y}_2 + \cdots + \mu_k\bar{y}_k : k \geqslant 0,\ \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k \in T,\ \mu_1, \mu_2, \ldots, \mu_k \in \mathbb{F}\}.\end{aligned}$$

The zero subspace of V

Also note that the set $\{\bar{0}_V\}$ is a subspace of V (why? what would you need to check?). It is called the *zero subspace* of V.

Thus, combined with what we showed previously, this gives us that, for every subset T of V *(regardless of whether it is empty or not)*, the set $\mathrm{span}(T)$ is a subspace of V (this finishes the proof of the 2nd part of the theorem).

Next, observe that the zero subspace is contained in any other subspace of V (why?), and thus it is the smallest subspace.

Question 1. Is there a largest subspace? If yes, which is it? Yes, it is the vector space V itself.

Question 2. Given two subspaces W_1, W_2 of V, what can you say about $W_1 \cap W_2$? Can the intersection of W_1 and W_2 be empty? No, it cannot be empty; the zero vector $\bar{0}_V$ is certainly contained in $W_1 \cap W_2$.

Terminology. The intersection of W_1 and W_2, denoted by $W_1 \cap W_2$, is the set that contains the common elements of W_1 and W_2 and only those, that is, the elements that are contained both in W_1 and in W_2.

Terminology from Set Theory: Intersections and Unions

Let A, B be two sets. As we've already said,

the intersection of A and B, denoted by $A \cap B$, is the set that contains the common elements of A and B and only those, that is, the elements that are contained **both in A and in B**.

Useful Remark. If, say, $B \subseteq A$, then $A \cap B = B$ (why?).

Special case: for every set A we have that $A \cap \emptyset = \emptyset$.

Recall also notation that we used earlier:

the union of A and B, denoted by $A \cup B$, is the set that contains the elements of A as well as the elements of B.

In other words, $x \in A \cup B$ **if and only if** $x \in A$ or $x \in B$.

Remark. If, say, $B \subseteq A$, then $A \cup B = A$ (why?).

Terminology

Given a field $\mathbb{F}$, and a vector space V over $\mathbb{F}$, we call any subspace W of V which can be spanned by exactly one non-zero vector of V a *line*.

In other words, the lines in V are sets containing all the scalar multiples of some non-zero vector $\bar{x}$ of V and only those elements; that is, sets of the form

$$\{\lambda \cdot \bar{x} : \lambda \in \mathbb{F}\}$$

where $\bar{x} \neq \bar{0}$.

Note that the terminology is inspired by settings such as $\mathbb{R}^2$ and $\mathbb{R}^3$, where the common visualisation of such sets is to draw them precisely as **straight lines passing through the origin** *(with the word 'line' having a geometric meaning now).*

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 15

Tuesday September 29

Substructures of Vector Spaces: Subspaces

Definition. Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Given a (non-empty) subset W of V, we can consider the restrictions of the operations of vector addition and of scalar multiplication in V to elements of W as follows:

$$(\bar{x}, \bar{y}) \in W \times W \ \mapsto \ \bar{x} + \bar{y} \in V$$
$$(r, \bar{x}) \in \mathbb{F} \times W \ \mapsto \ r \cdot \bar{x} \in V$$

- We say that W is *closed under the vector addition in V* if, for all $\bar{x}, \bar{y} \in W$, we have that $\bar{x} + \bar{y} \in W$ as well. Similarly, we say that W is *closed under the scalar multiplication in V* if, for all $\bar{x} \in W$ and for all $r \in \mathbb{F}$, we have that $r \cdot \bar{x} \in W$ as well.

- If W is closed under both the vector addition and the scalar multiplication in V, then we say that W is a subspace of V if W together with the restricted operations is a vector space over $\mathbb{F}$ itself.

Reminder: Sufficient to check

Let V be a vector space over $\mathbb{F}$, and let U be a subset of V.

U is a subspace of V
if and only if

1. U is closed under vector addition,
2. U is closed under scalar multiplication
3. U contains $\bar{0}_V$.

Note also that we could replace the 3rd requirement above by the requirement

"U is non-empty"

and obtain another triple of *necessary and sufficient* conditions for U to be a subspace (why would it again be sufficient to check conditions 1, 2 and the new 3rd one? because if we can find some vector $\bar{x} \in U$, we will also have $\bar{0}_V = 0_{\mathbb{F}} \cdot \bar{x} \in U$, as long as it also holds that U is closed under scalar multiplication).

Reminder: Subspaces coincide with linear spans

Theorem

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Consider a subset W of V.

Then W is a subspace of V **if and only if** W is equal to the linear span $\mathrm{span}(S)$ of some subset S of V.

Intersection of subspaces

Proposition 1

Let V be a vector space over a field $\mathbb{F}$, and let W_1, W_2 be two subspaces of V.

Then $W_1 \cap W_2$ is a subspace of V too.

Proof of Proposition 1

We need to show that $W_1 \cap W_2$

1. is closed under vector addition;
2. is closed under scalar multiplication;
3. contains $\bar{0}_V$.

– We have already seen that $\bar{0}_V \in W_1 \cap W_2$ (since $\bar{0}_V$ is an element of both W_1 and W_2).

– We check that $W_1 \cap W_2$ is closed under vector addition: that is, we need to show that, for every $\bar{x}, \bar{y} \in W_1 \cap W_2$, we have $\bar{x} + \bar{y} \in W_1 \cap W_2$ too.

Consider $\bar{x}, \bar{y} \in W_1 \cap W_2$.

- Then $\bar{x}, \bar{y} \in W_1$. Since W_1 is a subspace of V, we know that it is closed under vector addition. Therefore, $\bar{x} + \bar{y} \in W_1$.
- Similarly, $\bar{x}, \bar{y} \in W_2$. Since W_2 is a subspace of V, it is closed under vector addition. Therefore, $\bar{x} + \bar{y} \in W_2$.

Combining the two, we see that $\bar{x} + \bar{y} \in W_1 \cap W_2$ *(and since $\bar{x}, \bar{y} \in W_1 \cap W_2$ were arbitrary, this gives us that $W_1 \cap W_2$ is closed under vector addition).*

Proof of Proposition 1 (cont.)

– Finally, we show that $W_1 \cap W_2$ is closed under scalar multiplication: that is, we need to show that, for every $\bar{z} \in W_1 \cap W_2$ and every $r \in \mathbb{F}$, we have $r \cdot \bar{z} \in W_1 \cap W_2$ too.

Consider $\bar{z} \in W_1 \cap W_2$ and $r \in \mathbb{F}$.

- Then $\bar{z} \in W_1$. Since W_1 is a subspace of V, we know that it is closed under scalar multiplication. Therefore, $r \cdot \bar{z} \in W_1$.
- Similarly, $\bar{z} \in W_2$. Since W_2 is a subspace of V, it is closed under scalar multiplication. Therefore, $r \cdot \bar{z} \in W_2$.

Combining the two, we see that $r \cdot \bar{z} \in W_1 \cap W_2$ *(and since $\bar{z} \in W_1 \cap W_2$, $r \in \mathbb{F}$ were arbitrary, this gives us that $W_1 \cap W_2$ is closed under scalar multiplication).*

Having checked that conditions 1, 2 and 3 hold for $W_1 \cap W_2$, we conclude that $W_1 \cap W_2$ is a subspace of V.

On the other hand...

Let V be a vector space over a field $\mathbb{F}$, and let U_1, U_2 be two subspaces of V.

The union $U_1 \cup U_2$ of U_1 and U_2 will **not necessarily** be a subspace of V.

One example that shows this: let $V = \mathbb{R}^4$ (viewed as a vector space over $\mathbb{R}$), and let

$$U_1 = \operatorname{span}(\bar{e}_2) = \left\{ \begin{pmatrix} 0 \\ \lambda \\ 0 \\ 0 \end{pmatrix} : \lambda \in \mathbb{R} \right\}, \qquad U_2 = \operatorname{span}(\bar{e}_3) = \left\{ \begin{pmatrix} 0 \\ 0 \\ \mu \\ 0 \end{pmatrix} : \mu \in \mathbb{R} \right\}.$$

Question 1. What is $U_1 \cup U_2$ here?

It is the set $\left\{ \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ 0 \end{pmatrix} \in \mathbb{R}^4 : \text{at most one of } x_2, x_3 \text{ is non-zero} \right\}$.

Question 2. Why is $U_1 \cup U_2$ here not a subspace of $\mathbb{R}^4$?

Because it is not closed under vector addition: e.g. $\bar{e}_2, \bar{e}_3$ are both in $U_1 \cup U_2$, but $\bar{e}_2 + \bar{e}_3 \notin U_1 \cup U_2$.

Useful Question to ask: Given two subspaces W_1, W_2 of a vector space V, what is the smallest subspace of V out of those that contain the union $W_1 \cup W_2$ of W_1 and W_2 (or equivalently, that contain both W_1 and W_2)? Is there such a smallest subspace? Yes, the subspace $\operatorname{span}(W_1 \cup W_2)$.

Past exam problem

Consider the following sets of vectors from $\mathbb{Z}_3^4$:

$$S_1 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}, \qquad S_2 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$

$$S_3 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Exactly two of them have the same linear span: find which two, and also justify why the span of the remaining set is different (the scalars are taken from $\mathbb{Z}_3$).

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 16

Wednesday September 30

Past exam problem

Consider the following sets of vectors from $\mathbb{Z}_3^4$:

$$S_1 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}, \qquad S_2 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$

$$S_3 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Exactly two of them have the same linear span: find which two, and also justify why the span of the remaining set is different (the scalars are taken from $\mathbb{Z}_3$).

Formulating an approach

Assume that we have guessed that S_1 and S_2 have the same linear span. That is, assume that we want to show

$$\mathrm{span}(S_1) = \mathrm{span}(S_2)\,.$$

This is equivalent (why?) to showing two inclusions:

$$\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2) \qquad \text{and} \qquad \mathrm{span}(S_2) \subseteq \mathrm{span}(S_1)\,.$$

Observe also that, since we have $S_1 \subseteq \mathrm{span}(S_1)$, the first inclusion (once we show it) will imply

$$S_1 \subseteq \mathrm{span}(S_2)\,.$$

Similarly, the second inclusion will imply

$$S_2 \subseteq \mathrm{span}(S_1)\,.$$

In other words, the latter inclusions are (seemingly) weaker inclusions that we could first try to show as a sanity check (clearly if one of the weaker inclusions does not hold, the corresponding stronger inclusion will not hold either; but what if the weaker inclusions do hold? did we 'waste' time trying to show these instead of the stronger ones right away?).

Formulating an approach (cont.)

Important Claim

To show that $\mathrm{span}(S_1) = \mathrm{span}(S_2)$, it suffices to show that

$$S_1 \subseteq \mathrm{span}(S_2) \qquad \text{and} \qquad S_2 \subseteq \mathrm{span}(S_1)\,.$$

In other words,

$$\mathrm{span}(S_1) = \mathrm{span}(S_2)$$

is equivalent to

$$\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2) \quad \text{and} \quad \mathrm{span}(S_2) \subseteq \mathrm{span}(S_1),$$

which in turn is equivalent to

$$S_1 \subseteq \mathrm{span}(S_2) \quad \text{and} \quad S_2 \subseteq \mathrm{span}(S_1).$$

With this approach in mind

- How do we check that $S_1 \subseteq \mathrm{span}(S_2)$? We need to check that every vector in S_1 can be written as a linear combination of the vectors in S_2 (recall that we can reduce this to determining whether certain systems of linear equations have solutions; how?).
- Analogously we can check that $S_2 \subseteq \mathrm{span}(S_1)$ (in fact, for this particular example / problem this inclusion can be verified very quickly, because we have $S_2 \subseteq S_1$).
- Assuming that our guess is correct, and we do have $\mathrm{span}(S_1) = \mathrm{span}(S_2)$, how do we show afterwards that $\mathrm{span}(S_3) \neq \mathrm{span}(S_1)$?
  Again, we rely on the claim, and we note that the converse of what we want to show,

  $$\mathrm{span}(S_3) = \mathrm{span}(S_1)\,,$$

  would be equivalent to

  $$S_3 \subseteq \mathrm{span}(S_1) \qquad \text{and} \qquad S_1 \subseteq \mathrm{span}(S_3)\,.$$

  Thus, to disprove $\mathrm{span}(S_3) = \mathrm{span}(S_1)$, it would suffice to show that one of the latter inclusions fails, or in other words, that
  - one of the vectors in S_3 cannot be written as a linear combination of the vectors in S_1,
  - or one of the vectors in S_1 cannot be written as a linear combination of the vectors in S_3.

Justifying the claim

Why is $\mathrm{span}(S_1) = \mathrm{span}(S_2)$ equivalent to

$$S_1 \subseteq \mathrm{span}(S_2) \qquad \text{and} \qquad S_2 \subseteq \mathrm{span}(S_1)\,?$$

In other words, why, for instance, does the "weaker" inclusion $S_1 \subseteq \mathrm{span}(S_2)$ imply the "stronger" $\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2)$?

We'll justify this next time.

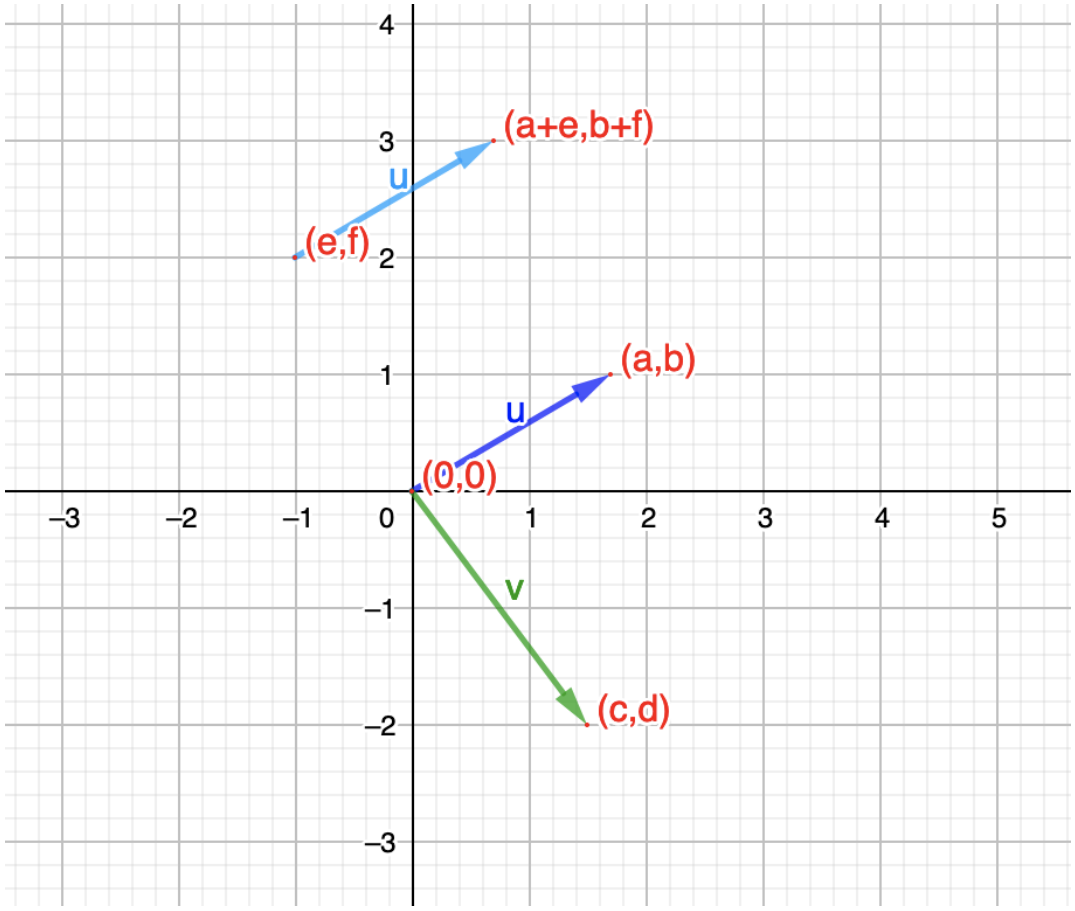
Next Main Topic:

Using Linear Algebra tools to do Geometry in $\mathbb{R}^2$ and $\mathbb{R}^3$

Visualisations

Recall that we identify $\mathbb{R}$ with a line, the 'real line'.

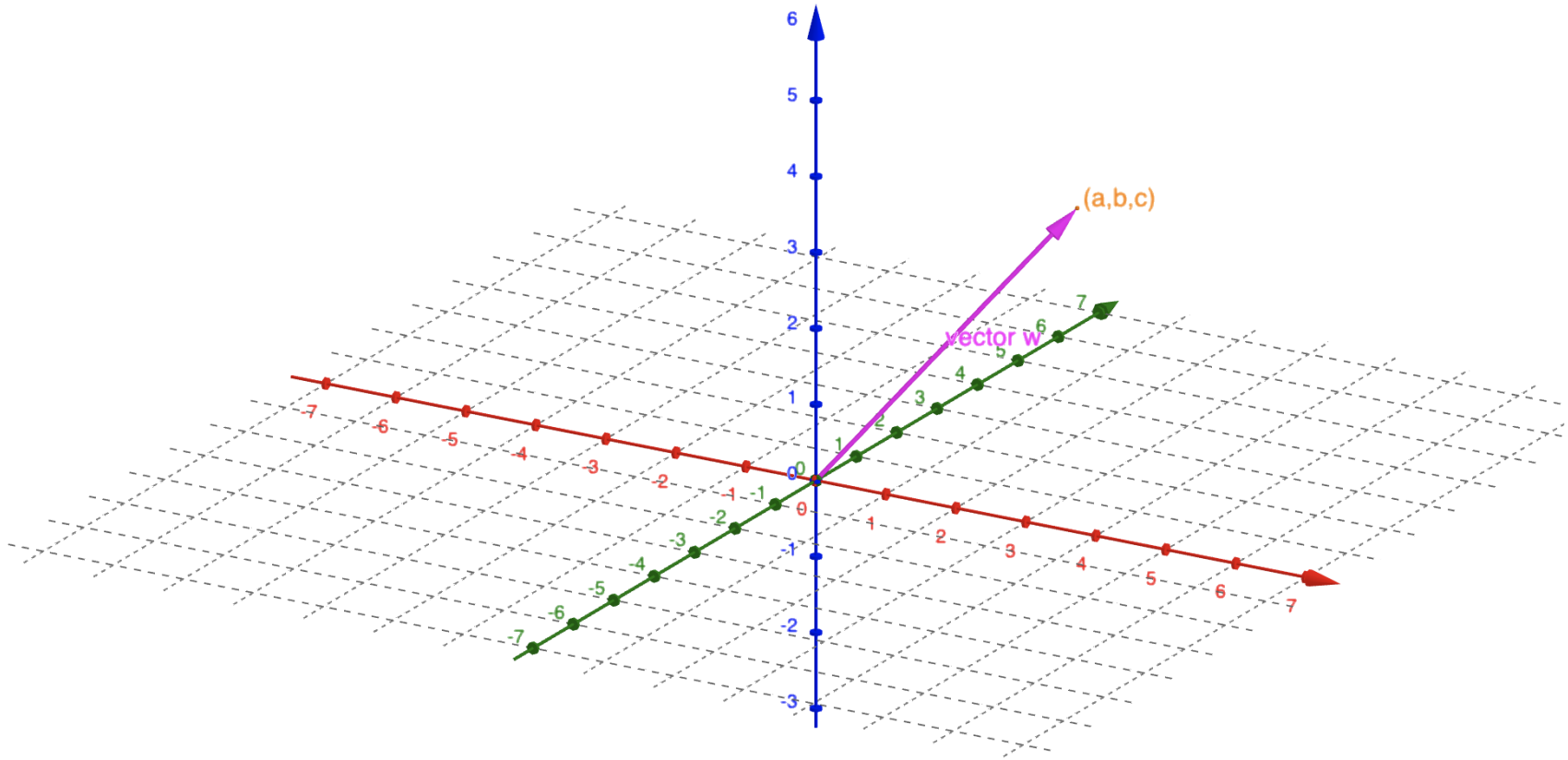
Based on this, we can visualise $\mathbb{R}^2$ as follows:



Pictorially we 'represent' a vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ in $\mathbb{R}^2$ by an arrow with its tail at the origin (that is, the point of intersection of the two axes), and its tip/head at the point (x_1, x_2), **or by any translate of this arrow**. Thus a vector is determined by its direction and its length, but not by its position.

Visualisations (cont.)

Similarly we visualise $\mathbb{R}^3$ as follows:



Here the vector w is the vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ in $\mathbb{R}^3$.

Some definitions / concepts with geometric meaning or use

Definition 1: Dot Product in $\mathbb{R}^n$

Definition: *Dot Product, or Scalar Product, or Inner Product*

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix}$ be two vectors in $\mathbb{R}^n$.

Then the *dot product* of $\bar{x}$ and $\bar{y}$ is denoted by $\langle \bar{x}, \bar{y} \rangle$ and is given by

$$\langle \bar{x}, \bar{y} \rangle := x_1 y_1 + x_2 y_2 + \cdots + x_{n-1} y_{n-1} + x_n y_n = \sum_{i=1}^{n} x_i y_i \, .$$

Very Important Observation. The dot product $\langle \bar{x}, \bar{y} \rangle$ of the vectors $\bar{x}$ and $\bar{y}$ is a **number** in $\mathbb{R}$.

Fundamental Properties of the Dot Product

Theorem

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$, $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix}$ and $\bar{z} = \begin{pmatrix} z_1 \\ z_2 \\ \ddots \\ z_{n-1} \\ z_n \end{pmatrix}$ be vectors in $\mathbb{R}^n$, and let $\lambda \in \mathbb{R}$.

Then

(i) $\langle \bar{x}, \bar{y} \rangle = \langle \bar{y}, \bar{x} \rangle$ *(symmetry of the dot product)*

(ii) $\langle \bar{x}, \bar{y} + \bar{z} \rangle = \langle \bar{x}, \bar{y} \rangle + \langle \bar{x}, \bar{z} \rangle$ *(Problem in HW2)*

(iii) $\langle \lambda \cdot \bar{x}, \bar{y} \rangle = \lambda \langle \bar{x}, \bar{y} \rangle = \langle \bar{x}, \lambda \cdot \bar{y} \rangle$.

(iv) $\langle \bar{x}, \bar{x} \rangle \geqslant 0$. Moreover, $\langle \bar{x}, \bar{x} \rangle = 0$ if and only if $\bar{x} = \bar{0}$. *(Positive-definiteness of the dot product; Problem in HW2)*

Properties (ii) and (iii) combined give the *linearity (or rather, bilinearity) of the dot product* (more about this later in the term).

Definition 2: Length or Norm of a vector in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} \in \mathbb{R}^n$. The *length* (or *norm*) of $\bar{x}$ is denoted by $\|\bar{x}\|$ and is given by

$$\|\bar{x}\| := \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2} \, .$$

Important Remark. Because the dot product is positive-definite, we have that $\|\bar{x}\| = 0$ if and only if $\bar{x}$ is the zero vector.

This allows us to also give the following definition:

Definition 3: Angle between two vectors in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}$ be two **non-zero** vectors in $\mathbb{R}^n$.

The *angle* between $\bar{x}$ and $\bar{y}$ is the *unique* real number θ such that

- $\theta \in [0, \pi]$
- and $\cos(\theta) = \dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$.

We say that $\bar{x}$ and $\bar{y}$ are *orthogonal* if the angle between them is equal to $\frac{\pi}{2}$; equivalently, if $\langle \bar{x}, \bar{y} \rangle = 0$.

Convention. The zero vector is orthogonal to every vector in $\mathbb{R}^n$.

Question. Does this definition make sense? *(see next lecture)*

Comments on the definitions

- Given that one position of the vector $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$ is the arrow with tail at the origin (that is, the n-tuple $(0, 0, \ldots, 0, 0)$) and tip at the point $(x_1, x_2, \ldots, x_{n-1}, x_n)$, we see that $\|\bar{x}\|$ coincides with the 'length of this arrow', that is, the Euclidean distance of the point $(x_1, x_2, \ldots, x_{n-1}, x_n)$ from the origin.

- If we consider two non-zero vectors $\bar{x}, \bar{y}$ in $\mathbb{R}^n$, and we visualise them as having the same initial point P_0 (that is, we consider such positions of them), then **the angle between $\bar{x}$ and $\bar{y}$ is the same as the angle of the two directed line segments we get**, whose one endpoint is the point P_0 and the other endpoint is either one of the terminal points of the vectors $\bar{x}, \bar{y}$.

  Thus we have come up with a formula for the angle between $\bar{x}$ and $\bar{y}$, which we can now find without having to draw the vectors.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 17

Friday October 2

Finishing a discussion from last time...

Past Exam Problem. Consider the following sets of vectors from $\mathbb{Z}_3^4$:

$$S_1 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}, \qquad S_2 = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$

$$S_3 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Exactly two of them have the same linear span: find which two, and also justify why the span of the remaining set is different (the scalars are taken from $\mathbb{Z}_3$).

Formulating an approach

Assume that we have guessed that S_1 and S_2 have the same linear span. That is, assume that we want to show

$$\mathrm{span}(S_1) = \mathrm{span}(S_2)\,.$$

This is equivalent (why?) to showing two inclusions:

$$\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2) \qquad \text{and} \qquad \mathrm{span}(S_2) \subseteq \mathrm{span}(S_1)\,.$$

Observe also that, since we have $S_1 \subseteq \mathrm{span}(S_1)$, the first inclusion (once we show it) will imply

$$S_1 \subseteq \mathrm{span}(S_2)\,.$$

Similarly, the second inclusion will imply

$$S_2 \subseteq \mathrm{span}(S_1)\,.$$

In other words, the latter inclusions are (seemingly) weaker inclusions that we could first try to show as a sanity check (clearly if one of the weaker inclusions does not hold, the corresponding stronger inclusion will not hold either; but what if the weaker inclusions do hold? did we 'waste' time trying to show these instead of the stronger ones right away?).

Important Claim

To show that $\mathrm{span}(S_1) = \mathrm{span}(S_2)$, it suffices to show that

$$S_1 \subseteq \mathrm{span}(S_2) \qquad \text{and} \qquad S_2 \subseteq \mathrm{span}(S_1)\,.$$

With this approach in mind

- How do we check that $S_1 \subseteq \mathrm{span}(S_2)$? We need to check that every vector in S_1 can be written as a linear combination of the vectors in S_2 (recall that we can reduce this to determining whether certain systems of linear equations have solutions; how?).
- Analogously we can check that $S_2 \subseteq \mathrm{span}(S_1)$ (in fact, for this particular example / problem this inclusion can be verified very quickly, because we have $S_2 \subseteq S_1$).
- Assuming that our guess is correct, and we do have $\mathrm{span}(S_1) = \mathrm{span}(S_2)$, how do we show afterwards that $\mathrm{span}(S_3) \neq \mathrm{span}(S_1)$?
  Again, we rely on the claim, and we note that the converse of what we want to show,

  $$\mathrm{span}(S_3) = \mathrm{span}(S_1)\,,$$

  would be equivalent to

  $$S_3 \subseteq \mathrm{span}(S_1) \qquad \text{and} \qquad S_1 \subseteq \mathrm{span}(S_3)\,.$$

  Thus, to disprove $\mathrm{span}(S_3) = \mathrm{span}(S_1)$, it would suffice to show that one of the latter inclusions fails, or in other words, that
  - one of the vectors in S_3 cannot be written as a linear combination of the vectors in S_1,
  - or one of the vectors in S_1 cannot be written as a linear combination of the vectors in S_3.

Justifying the claim

Looking into what's going on in detail:

write $S_1 = \{\bar{x}_1, \bar{x}_2, \bar{x}_3\}$ and $S_2 = \{\bar{y}_1, \bar{y}_2\}$ (and for the time being ignore what $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and $\bar{y}_1, \bar{y}_2$ exactly are).

Why does $S_1 \subseteq \mathrm{span}(S_2)$ imply $\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2)$?

Assume that we have already found scalars $\lambda_1, \mu_1, \lambda_2, \mu_2, \lambda_3, \mu_3 \in \mathbb{Z}_3$ such that

$$\bar{x}_1 = \lambda_1 \cdot \bar{y}_1 + \mu_1 \cdot \bar{y}_2, \qquad \bar{x}_2 = \lambda_2 \cdot \bar{y}_1 + \mu_2 \cdot \bar{y}_2,$$
$$\text{and} \qquad \bar{x}_3 = \lambda_3 \cdot \bar{y}_1 + \mu_3 \cdot \bar{y}_2.$$

Consider now an arbitrary element $\bar{z} \in \mathrm{span}(S_1)$. Then $\bar{z}$ can be written in the form

$$\bar{z} = \alpha \cdot \bar{x}_1 + \beta \cdot \bar{x}_2 + \gamma \cdot \bar{x}_3$$

for some $\alpha, \beta, \gamma \in \mathbb{Z}_3$.

Justifying the claim (cont.)

But then

$$\begin{aligned}\bar{z} &= \alpha \cdot \bar{x}_1 + \beta \cdot \bar{x}_2 + \gamma \cdot \bar{x}_3 \\ &= \alpha \cdot (\lambda_1 \cdot \bar{y}_1 + \mu_1 \cdot \bar{y}_2) + \beta \cdot (\lambda_2 \cdot \bar{y}_1 + \mu_2 \cdot \bar{y}_2) \\ &\qquad + \gamma \cdot (\lambda_3 \cdot \bar{y}_1 + \mu_3 \cdot \bar{y}_2)\end{aligned}$$

$$= (\alpha\lambda_1 + \beta\lambda_2 + \gamma\lambda_3) \cdot \bar{y}_1 + (\alpha\mu_1 + \beta\mu_2 + \gamma\mu_3) \cdot \bar{y}_2$$

where we use both distributive laws, the associativity of scalar multiplication, and the associativity and commutativity of vector addition in $\mathbb{Z}_3^4$,

and thus $\bar{z} \in \mathrm{span}(S_2)$.

Since $\bar{z}$ was an arbitrary element of $\mathrm{span}(S_1)$, we can conclude that $\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2)$.

Using more advanced language that we have already introduced

Why does $S_1 \subseteq \mathrm{span}(S_2)$ imply $\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2)$?

- We have seen that $W_2 = \mathrm{span}(S_2)$ is a subspace of $\mathbb{Z}_3^4$.

We have also seen that in this case we have

$$W_2 = \mathrm{span}(W_2).$$

In other words, $\mathrm{span}(S_2) = \mathrm{span}\big(\mathrm{span}(S_2)\big)$.

- Going back to the inclusion $S_1 \subseteq \mathrm{span}(S_2) = W_2$, we note that it implies that

$$\mathrm{span}(S_1) \subseteq \mathrm{span}(W_2)$$

(this is because the linear span of S_1 contains all finite sums of scalar multiples of elements of S_1; but all elements of S_1 are also elements of W_2, so all finite sums of scalar multiples of these elements are contained in $\mathrm{span}(W_2)$ too).

Combining the above, we see that

$$S_1 \subseteq \mathrm{span}(S_2) \text{ implies } \mathrm{span}(S_1) \subseteq \mathrm{span}\big(\mathrm{span}(S_2)\big) = \mathrm{span}(S_2).$$

Using Linear Algebra tools to do Geometry in $\mathbb{R}^2$ and $\mathbb{R}^3$

Definition 1: Dot Product in $\mathbb{R}^n$

Definition: *Dot Product, or Scalar Product, or Inner Product*

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}$ be two vectors in $\mathbb{R}^n$.

Then the *dot product* of $\bar{x}$ and $\bar{y}$ is denoted by $\langle \bar{x}, \bar{y} \rangle$ and is given by

$$\begin{aligned} \langle \bar{x}, \bar{y} \rangle &:= \sum_{i=1}^{n} x_i y_i \\ &= x_1 y_1 + x_2 y_2 + \cdots + x_{n-1} y_{n-1} + x_n y_n. \end{aligned}$$

Very Important Observation. The dot product $\langle \bar{x}, \bar{y} \rangle$ of the vectors $\bar{x}$ and $\bar{y}$ is a **number** in $\mathbb{R}$.

Fundamental Properties of the Dot Product

Theorem

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$, $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix}$ and $\bar{z} = \begin{pmatrix} z_1 \\ z_2 \\ \ddots \\ z_{n-1} \\ z_n \end{pmatrix}$ be vectors in $\mathbb{R}^n$, and let $\lambda \in \mathbb{R}$.

Then

(i) $\langle \bar{x}, \bar{y} \rangle = \langle \bar{y}, \bar{x} \rangle$ *(symmetry of the dot product)*

(ii) $\langle \bar{x}, \bar{y} + \bar{z} \rangle = \langle \bar{x}, \bar{y} \rangle + \langle \bar{x}, \bar{z} \rangle$ *(Problem in HW2)*

(iii) $\langle \lambda \cdot \bar{x}, \bar{y} \rangle = \lambda \langle \bar{x}, \bar{y} \rangle = \langle \bar{x}, \lambda \cdot \bar{y} \rangle$.

(iv) $\langle \bar{x}, \bar{x} \rangle \geqslant 0$. Moreover, $\langle \bar{x}, \bar{x} \rangle = 0$ if and only if $\bar{x} = \bar{0}$. *(Positive-definiteness of the dot product; Problem in HW2)*

Properties (ii) and (iii) combined give the *linearity (or rather, bilinearity) of the dot product* (more about this later in the term).

Verification of Properties (i) and (iii)

(i) We have

$$\langle \bar{x}, \bar{y} \rangle = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i \qquad \text{(by the commutativity of multiplication in } \mathbb{R}\text{)}$$
$$= \langle \bar{y}, \bar{x} \rangle .$$

(iii) We have

$$\langle \lambda \cdot \bar{x}, \bar{y} \rangle = \sum_{i=1}^{n} (\lambda x_i) \cdot y_i = \sum_{i=1}^{n} \lambda \cdot (x_i y_i) \qquad \text{(by the associativity of multiplication in } \mathbb{R}\text{)}$$
$$= \lambda \cdot \sum_{i=1}^{n} x_i y_i \qquad \text{(distributive law in } \mathbb{R}\text{)}$$
$$= \lambda \langle \bar{x}, \bar{y} \rangle .$$

Moreover, the equality $\langle \bar{x}, \lambda \cdot \bar{y} \rangle = \lambda \langle \bar{x}, \bar{y} \rangle$ can be shown analogously, or more simply follows from this combined with Property (i).

Definition 2: Length or Norm of a vector in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} \in \mathbb{R}^n$. The *length* (or *norm*) of $\bar{x}$ is denoted by $\|\bar{x}\|$ and is given by

$$\|\bar{x}\| := \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2} \, .$$

Important Remark. Because the dot product is positive-definite, we have that $\|\bar{x}\| = 0$ if and only if $\bar{x}$ is the zero vector.

This allows us to also give the following definition:

Definition 3: Angle between two vectors in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix}$ be two **non-zero** vectors in $\mathbb{R}^n$.

The *angle* between $\bar{x}$ and $\bar{y}$ is the *unique* real number θ such that

- $\theta \in [0, \pi]$
- and $\cos(\theta) = \dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$.

We say that $\bar{x}$ and $\bar{y}$ are *orthogonal* if the angle between them is equal to $\frac{\pi}{2}$; equivalently, if $\langle \bar{x}, \bar{y} \rangle = 0$.

Convention. The zero vector is orthogonal to every vector in $\mathbb{R}^n$.

Question. Does this definition make sense?

The Cauchy-Schwarz inequality

For the definition of the angle we just gave to make sense, we need to be able to view the expression

$$\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$$

as the cosine of an angle.

But this would happen if and only if $\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$ belonged to the range of the cosine function, that is,

$$\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} \in [-1, 1].$$

This is indeed true, and follows by the so-called **Cauchy-Schwarz inequality**: for every two vectors $\bar{z}, \bar{w}$ in $\mathbb{R}^n$, we have

$$|\langle \bar{z}, \bar{w} \rangle| \leqslant \|\bar{z}\| \cdot \|\bar{w}\|,$$

or equivalently

$$-\|\bar{z}\| \cdot \|\bar{w}\| \leqslant \langle \bar{z}, \bar{w} \rangle \leqslant \|\bar{z}\| \cdot \|\bar{w}\|.$$

The Cauchy-Schwarz inequality

For every two vectors $\bar{z}, \bar{w}$ in $\mathbb{R}^n$, we have

$$|\langle \bar{z}, \bar{w} \rangle| \leqslant \|\bar{z}\| \cdot \|\bar{w}\|,$$

or equivalently

$$-\|\bar{z}\| \cdot \|\bar{w}\| \leqslant \langle \bar{z}, \bar{w} \rangle \leqslant \|\bar{z}\| \cdot \|\bar{w}\|.$$

Moreover,

- $\langle \bar{z}, \bar{w} \rangle = \|\bar{z}\| \cdot \|\bar{w}\|$ **if and only if** $\bar{z}$ and $\bar{w}$ are *parallel* and have the same direction *(that is, one vector is a positive (or non-negative) scalar multiple of the other)*.
- $\langle \bar{z}, \bar{w} \rangle = -\|\bar{z}\| \cdot \|\bar{w}\|$ **if and only if** $\bar{z}$ and $\bar{w}$ are *parallel* and have opposite directions *(that is, one vector is a negative scalar multiple of the other)*.

Parallel vectors

In other words, given two non-zero vectors $\bar{x}, \bar{y}$ in $\mathbb{R}^n$,

- we have that $\dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} = 1$,

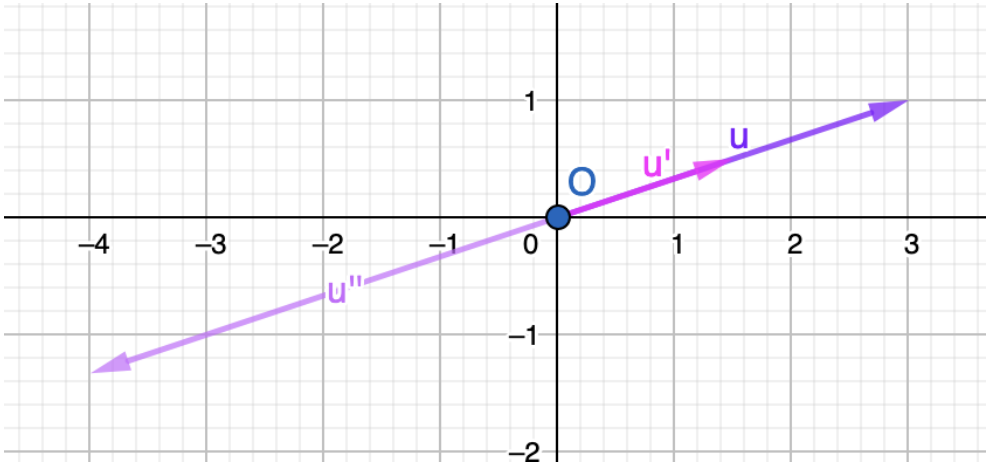
  or equivalently the angle between $\bar{x}$ and $\bar{y}$ is equal to 0,

  if and only if $\bar{x}$ and $\bar{y}$ are parallel and have the same direction;

- we have that $\dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} = -1$,

  or equivalently the angle between $\bar{x}$ and $\bar{y}$ is equal to π,

  if and only if $\bar{x}$ and $\bar{y}$ are parallel and have opposite directions.



For instance, the angle between vectors u and u' is 0,
while the angle between vectors u and u'' is π.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 18

Monday October 5

Using Linear Algebra tools to do Geometry in $\mathbb{R}^2$ and $\mathbb{R}^3$

Definitions from Last Two Lectures

Definition 2: Length or Norm of a vector in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} \in \mathbb{R}^n$. The *length* (or *norm*) of $\bar{x}$ is denoted by $\|\bar{x}\|$ and is given by

$$\|\bar{x}\| := \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2} \, .$$

Important Remark. Because the dot product is positive-definite, we have that $\|\bar{x}\| = 0$ if and only if $\bar{x}$ is the zero vector.

This allows us to also give the following definition:

Definition 3: Angle between two vectors in $\mathbb{R}^n$

Definition

Let $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$ and $\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ddots \\ y_{n-1} \\ y_n \end{pmatrix}$ be two **non-zero** vectors in $\mathbb{R}^n$.

The *angle* between $\bar{x}$ and $\bar{y}$ is the *unique* real number θ such that

- $\theta \in [0, \pi]$
- and $\cos(\theta) = \dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$.

We say that $\bar{x}$ and $\bar{y}$ are *orthogonal* if the angle between them is equal to $\frac{\pi}{2}$; equivalently, if $\langle \bar{x}, \bar{y} \rangle = 0$.

Convention. The zero vector is orthogonal to every vector in $\mathbb{R}^n$.

Question. Does this definition make sense?

The Cauchy-Schwarz inequality

For the definition of the angle we just gave to make sense, we need to be able to view the expression

$$\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$$

as the cosine of an angle.

But this would happen if and only if $\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|}$ belonged to the range of the cosine function, that is,

$$\frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} \in [-1, 1].$$

This is indeed true, and follows by the so-called **Cauchy-Schwarz inequality**: for every two vectors $\bar{z}, \bar{w}$ in $\mathbb{R}^n$, we have

$$|\langle \bar{z}, \bar{w} \rangle| \leqslant \|\bar{z}\| \cdot \|\bar{w}\|,$$

or equivalently

$$-\|\bar{z}\| \cdot \|\bar{w}\| \leqslant \langle \bar{z}, \bar{w} \rangle \leqslant \|\bar{z}\| \cdot \|\bar{w}\|.$$

The Cauchy-Schwarz inequality

For every two vectors $\bar{z}, \bar{w}$ in $\mathbb{R}^n$, we have

$$|\langle \bar{z},\, \bar{w} \rangle| \leqslant \|\bar{z}\| \cdot \|\bar{w}\|,$$

or equivalently

$$-\|\bar{z}\| \cdot \|\bar{w}\| \leqslant \langle \bar{z},\, \bar{w} \rangle \leqslant \|\bar{z}\| \cdot \|\bar{w}\|.$$

Moreover,

- $\langle \bar{z},\, \bar{w} \rangle = \|\bar{z}\| \cdot \|\bar{w}\|$ **if and only if** $\bar{z}$ and $\bar{w}$ are *parallel* and have the same direction *(that is, one vector is a positive (or non-negative) scalar multiple of the other)*.
- $\langle \bar{z},\, \bar{w} \rangle = -\|\bar{z}\| \cdot \|\bar{w}\|$ **if and only if** $\bar{z}$ and $\bar{w}$ are *parallel* and have opposite directions *(that is, one vector is a negative scalar multiple of the other)*.

Parallel vectors

In other words, given two non-zero vectors $\bar{x}, \bar{y}$ in $\mathbb{R}^n$,

- we have that $\dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} = 1$,

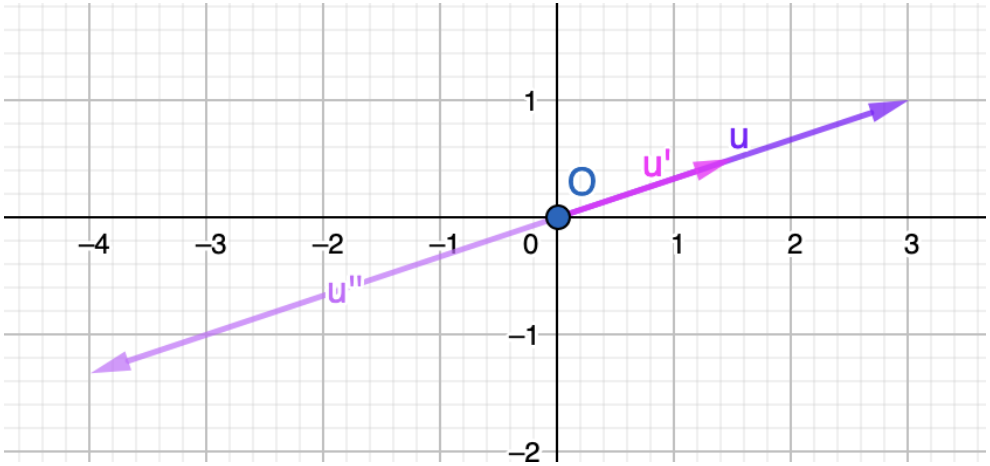
  or equivalently the angle between $\bar{x}$ and $\bar{y}$ is equal to 0,

  if and only if $\bar{x}$ and $\bar{y}$ are parallel and have the same direction;

- we have that $\dfrac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \cdot \|\bar{y}\|} = -1$,

  or equivalently the angle between $\bar{x}$ and $\bar{y}$ is equal to π,

  if and only if $\bar{x}$ and $\bar{y}$ are parallel and have opposite directions.



For instance, the angle between vectors u and u' is 0,
while the angle between vectors u and u'' is π.

Some more comments on the definitions

- Given that one position of the vector $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$ is the arrow with tail at the origin (that is, the n-tuple $(0, 0, \ldots, 0, 0)$) and tip at the point $(x_1, x_2, \ldots, x_{n-1}, x_n)$, we see that $\|\bar{x}\|$ coincides with the 'length of this arrow', that is, the Euclidean distance of the point $(x_1, x_2, \ldots, x_{n-1}, x_n)$ from the origin.

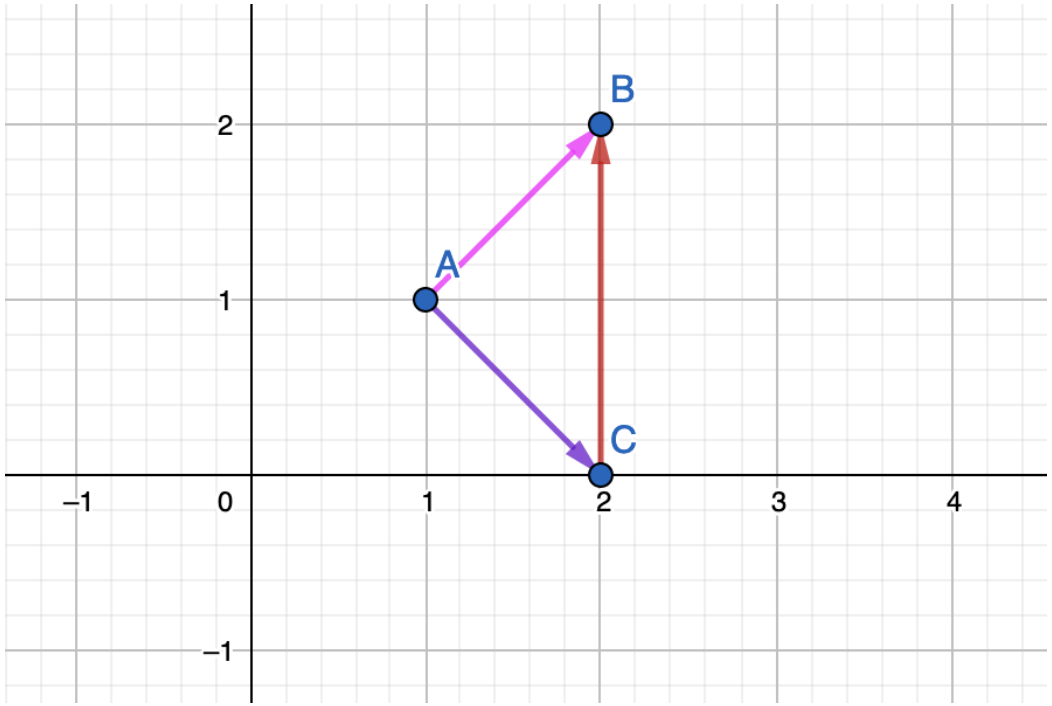
- If we consider two non-zero vectors $\bar{x}, \bar{y}$ in $\mathbb{R}^n$, and we visualise them as having the same initial point P_0 (that is, we consider such positions of them), then **the angle between $\bar{x}$ and $\bar{y}$ is the same as the angle of the two directed line segments we get**, whose one endpoint is the point P_0 and the other endpoint is either one of the terminal points of the vectors $\bar{x}, \bar{y}$.

  Thus we have come up with a formula for the angle between $\bar{x}$ and $\bar{y}$, which we can now find without having to draw the vectors.

Practice Problems

Example 1. Consider the points $A = (1, 1)$, $B = (2, 2)$ and $C = (2, 0)$ in $\mathbb{R}^2$. Show that they determine a (non-degenerate) triangle, and find the lengths of the edges of this triangle, as well as the angles between any two edges.

We could try drawing these points on the plane to verify that they determine a triangle (and also measure the edge lengths and angles in our drawing):



However, now we can also apply the formulas we gave above (and use only the coordinates of the points).

Answer to Example 1

Let us write $\overrightarrow{AB}$ for the vector with initial point A and terminal point B (note that this would be one of the representations of this vector).

Similarly, let us write $\overrightarrow{AC}$ for the vector with initial point A and terminal point C, and $\overrightarrow{CB}$ for the vector with initial point C and terminal point B.

Then

$$\overrightarrow{AB} = \begin{pmatrix} 2-1 \\ 2-1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \overrightarrow{AC} = \begin{pmatrix} 2-1 \\ 0-1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$
$$\text{and} \quad \overrightarrow{CB} = \begin{pmatrix} 2-2 \\ 2-0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

The lengths of these vectors are

$$\|\overrightarrow{AB}\| = \sqrt{1^2+1^2} = \sqrt{2}, \quad \|\overrightarrow{AC}\| = \sqrt{1^2+(-1)^2} = \sqrt{2},$$
$$\text{and} \quad \|\overrightarrow{CB}\| = \sqrt{0^2+2^2} = 2.$$

Answer to Example 1 (cont.)

We can then check: $\langle \overrightarrow{AB}, \overrightarrow{AC} \rangle = 1 \cdot 1 + 1 \cdot (-1) = 0$, thus the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are orthogonal.

This also implies that the three points are **not** *collinear* (that is, they are not all found on the same line), and hence they form a triangle.

The lengths that we found above are the lengths of the edges of this triangle, while one of its angles, the angle $B\widehat{A}C$, is a right angle.

Moreover

$$\frac{\langle \overrightarrow{BA}, \overrightarrow{BC} \rangle}{\|\overrightarrow{BA}\| \cdot \|\overrightarrow{BC}\|} = \frac{\langle -\overrightarrow{AB}, -\overrightarrow{CB} \rangle}{\| -\overrightarrow{AB}\| \cdot \| -\overrightarrow{CB}\|}$$

$$= \frac{\langle \overrightarrow{AB}, \overrightarrow{CB} \rangle}{\|\overrightarrow{AB}\| \cdot \|\overrightarrow{CB}\|} = \frac{2}{\sqrt{2} \cdot 2} = \frac{1}{\sqrt{2}} \in (-1, 1),$$

thus the vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$ are **not** parallel, and $A\widehat{B}C = \arccos\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}$.

Similarly, $\frac{\langle \overrightarrow{CA}, \overrightarrow{CB} \rangle}{\|\overrightarrow{CA}\| \cdot \|\overrightarrow{CB}\|} = \frac{\langle -\overrightarrow{AC}, \overrightarrow{CB} \rangle}{\| -\overrightarrow{AC}\| \cdot \|\overrightarrow{CB}\|} = \frac{-(-2)}{\sqrt{2} \cdot 2} = \frac{1}{\sqrt{2}}$,

thus $A\widehat{C}B = \arccos\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}$, as we expected.

Practice Problems (cont.)

Example 2. Find the distance between the points $P(2, -1, 0)$ and $Q(4, 9, 3)$ in $\mathbb{R}^3$.

Solution. This distance coincides with the length of any vector that has initial point one of P or Q and terminal point the other point. For instance, it is equal to the length of the vector

$$\overrightarrow{PQ} = \begin{pmatrix} 4-2 \\ 9-(-1) \\ 3-0 \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \\ 3 \end{pmatrix}.$$

We thus get that the distance between P and Q is given by

$$\|\overrightarrow{PQ}\| = \sqrt{2^2 + 10^2 + 3^2} = \sqrt{113}.$$

Equations of lines in $\mathbb{R}^2$

Normal form

General form of a linear equation representing a line in $\mathbb{R}^2$:

$$ax + by + c = 0, \quad \text{where } a, b, c \in \mathbb{R}, \ (a, b) \neq (0, 0).$$

The set of all points $(x, y) \in \mathbb{R}^2$ satisfying this (linear) equation forms a line.

Moreover, the slope of this line is $-\frac{a}{b}$ (if $b \neq 0$, otherwise the slope is not defined).

Terminology. The above form is sometimes called the *normal form* of the line. This is because, if we initially assume that $c = 0$, then the equation can be equivalently rewritten as

$$\left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix} \right\rangle = 0,$$

and is satisfied precisely by those points (x, y) such that the corresponding vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is orthogonal (or *normal*) to the vector $\begin{pmatrix} a \\ b \end{pmatrix}$.

In the more general case, where $c \neq 0$, the set of all points satisfying the equation is merely a translate of a set of the latter type.

Equivalent ways of representing a line in $\mathbb{R}^2$

Slope-intercept form: Whenever the slope of a line in $\mathbb{R}^2$ is defined (that is, whenever the line is not parallel to the y-axis), we can represent it by an equation of the form

$$y = kx + q_0.$$

Here k is the slope of the line, while q_0 is its y-intercept (that is, **the second coordinate of the unique point** at which the line intersects the y-axis; note that the requirement that exactly one such point exists is the reason why we don't consider lines that are parallel to the y-axis here).

'Disadvantage' of this form: Not every line in $\mathbb{R}^2$ can be described in this way.

Equivalent ways of representing a line in $\mathbb{R}^2$

Two-point form: Recall that given two **different** points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ in $\mathbb{R}^2$, there is **exactly one line** ℓ_0 passing through both of them.

We note that the line ℓ_0 should have the same direction as the vector $\overrightarrow{P_1P_2} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}$.

Also, for any other point $P(x, y)$ on ℓ_0, we should have that again the direction of ℓ_0 is the same as that of the vector $\overrightarrow{P_1P}$, or in other words the vector has the same direction as $\overrightarrow{P_1P_2}$ (or opposite direction).

This means exactly that we can find $r \in \mathbb{R}$ such that

$$\overrightarrow{P_1P} = r \cdot \overrightarrow{P_1P_2} \qquad \Leftrightarrow \qquad \frac{x - x_1}{x_2 - x_1} = r = \frac{y - y_1}{y_2 - y_1}$$

(where we assume that $x_1 \neq x_2$ and $y_1 \neq y_2$ for the second group of equalities to make sense).

In the most general case (that is, when we don't assume anything about x_1, x_2, y_1, y_2 besides $(x_1, y_1) \neq (x_2, y_2)$), we can represent ℓ_0 by the equation

$$(x_2 - x_1)(y - y_1) = (y_2 - y_1)(x - x_1).$$

Important Remarks

1. By algebraically manipulating any of the different forms representing a line in $\mathbb{R}^2$ that we mentioned above, we can always obtain a normal form for the line.

2. The observation that, given two **different** points P_1 and P_2 in $\mathbb{R}^n$, there is **exactly one line** passing through both of them, remains valid even in higher-dimensional Euclidean spaces, that is, when $n > 2$. Again we can describe this line by observing that, for every other point P on it, we must have

$$\overrightarrow{P_1P} = r \cdot \overrightarrow{P_1P_2}$$

for some $r \in \mathbb{R}$.

Vector equation of a line in $\mathbb{R}^2$

Note that we used the last observation to get a two-point form for a line in $\mathbb{R}^2$. We can now also use it to give a vector equation for the line; in fact, observe that we have already used vector notation to some extent.

If ℓ_0 is a line in $\mathbb{R}^2$, and P_1, P_2 are two different points contained in it, then every other point $P(x, y)$ contained in ℓ_0 must satisfy

$$\overrightarrow{P_1P} = r \cdot \overrightarrow{P_1P_2}$$

for some $r \in \mathbb{R}$.

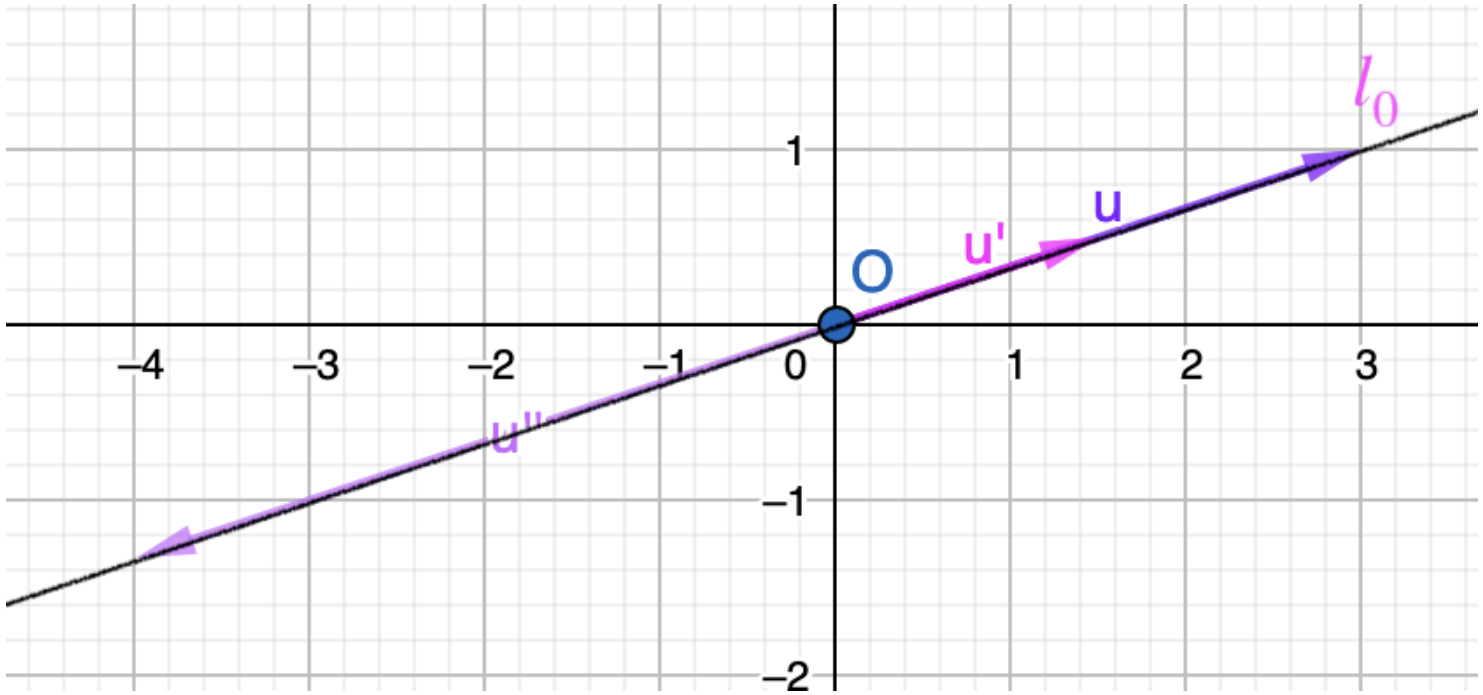
We also note that we can write $\overrightarrow{P_1P} = \overrightarrow{OP} - \overrightarrow{OP_1}$, where O stands for the origin in $\mathbb{R}^2$.

Thus, the arbitrary point $P(x, y)$ on ℓ_0, which we identify with the vector $\overrightarrow{OP} = \binom{x}{y}$ which has initial point the origin and terminal point the point P, will be contained in the set of solutions to the vector equation

$$\binom{x}{y} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1P_2}, \qquad r \in \mathbb{R}.$$

The vector $\overrightarrow{P_1P_2}$ here gives the direction of the line, while the vector $\overrightarrow{OP_1}$ gives one point contained in the line.

Simplest example: the line passes through the origin



A vector equation for the line ℓ_0 here is the following:

$$\begin{pmatrix} x \\ y \end{pmatrix} = r \cdot u = r \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \qquad r \in \mathbb{R}.$$

Practice Problems

Example 3. (i) Consider the line ℓ_1 in $\mathbb{R}^2$ given by the equation

$$\ell_1 : \ x - 2y = 0.$$

Find a vector equation representing ℓ_1.

(ii) Consider the line ℓ_2 in $\mathbb{R}^2$ with vector equation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \mu \in \mathbb{R}.$$

Find a linear equation in normal form representing ℓ_2.

Answer to Example 3

(i) We need to find two points on ℓ_1.

- By setting $x = 0$ and solving for y in the linear equation representing ℓ_1, we get that the origin $O(0, 0)$ belongs to ℓ_1.
- By setting $x = 2$ and solving for y in the equation representing ℓ_1, we get that the point $P_2(2, 1)$ belongs to ℓ_1 as well.

We conclude that a vector equation for ℓ_1 is the equation

$$\binom{x}{y} = r \cdot \overrightarrow{OP_2} = r \cdot \binom{2}{1}, \qquad r \in \mathbb{R}.$$

(ii) We use the given vector equation to find two points on ℓ_2, and then based on these we write the two-point form for ℓ_2 which we can quickly transform to the normal form.

- By setting $\mu = 0$, we see that one point belonging to ℓ_2 is the point $Q_1(x_1, y_1)$ with $x_1 = 1$, $y_1 = 1$.
- By setting $\mu = 1$, we find that a second point belonging to ℓ_2 is the point $Q_2(x_2, y_2)$ with $x_2 = 2$, $y_2 = 0$.

We conclude that an equation in two-point form for ℓ_2 is the equation

$$y - 1 = (2 - 1)(y - 1) = (0 - 1)(x - 1) = -x + 1.$$

Therefore $x + y - 2 = 0$ is an equation in normal form representing ℓ_2.

Vector equations of lines in $\mathbb{R}^3$

Vector equation of a line in $\mathbb{R}^3$

We will see that in $\mathbb{R}^3$ one linear equation can never represent a line.

Instead, we can describe a line by a system of two linear equations in the unknowns x, y, z,

and any such system *(except in some special cases, which we will soon understand)* represents a line (that is, the set of all points (x, y, z) whose coordinates solve the linear system forms a line).

E.g. the system

$$\left\{\begin{array}{rcrcrcr} x & + & y & + & z & = & 0 \\ x & + & 2y & + & 3z & = & -1 \end{array}\right\}$$

of linear equations defines a specific line in $\mathbb{R}^3$.

Question. How can we find a vector equation representing this line? What should such a vector equation look like?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 19

Tuesday October 6

From Last Time: Vector equation of a line in $\mathbb{R}^2$

If ℓ_0 is a line in $\mathbb{R}^2$, and P_1, P_2 are two different points contained in it, then every other point $P(x, y)$ contained in ℓ_0 must satisfy

$$\overrightarrow{P_1P} = r \cdot \overrightarrow{P_1P_2}$$

for some $r \in \mathbb{R}$.

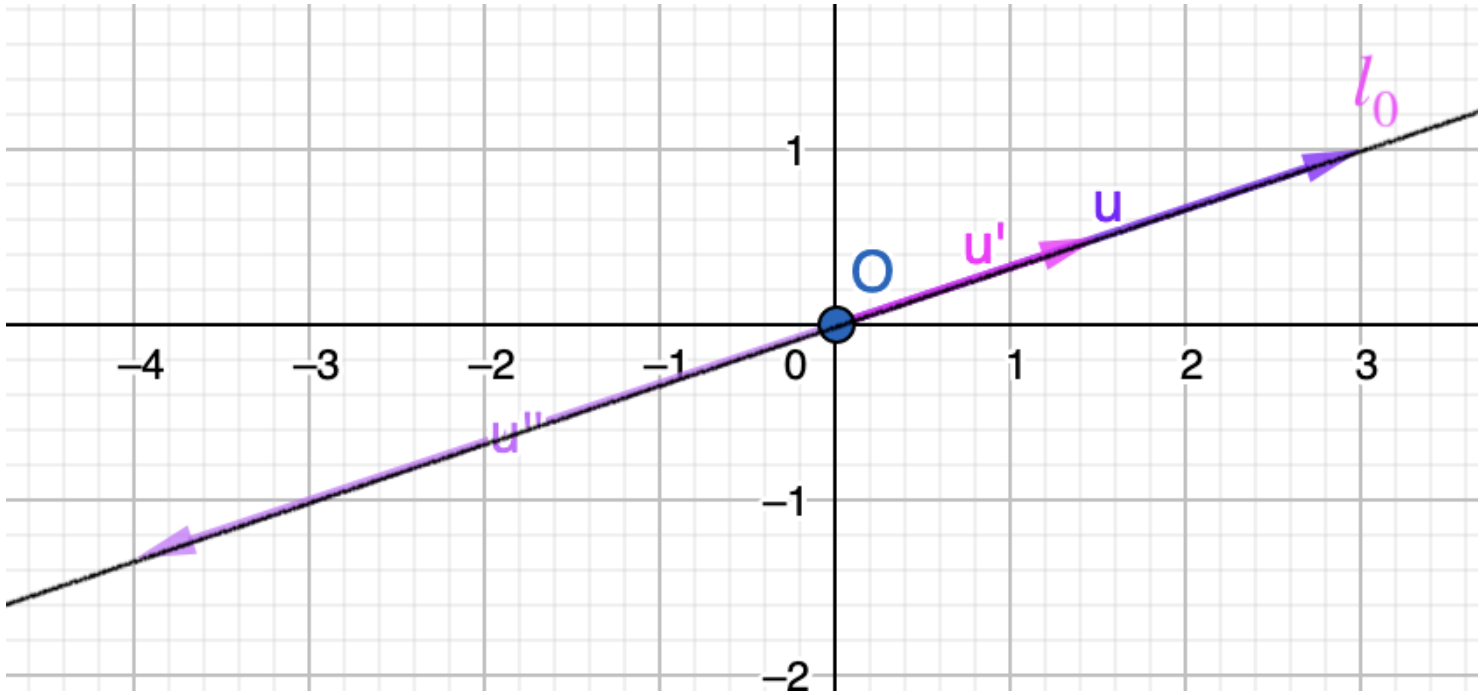
We also note that we can write $\overrightarrow{P_1P} = \overrightarrow{OP} - \overrightarrow{OP_1}$, where O stands for the origin in $\mathbb{R}^2$.

Thus, the arbitrary point $P(x, y)$ on ℓ_0, which we identify with the vector $\overrightarrow{OP} = \binom{x}{y}$ which has initial point the origin and terminal point the point P, will be contained in the set of solutions to the vector equation

$$\binom{x}{y} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1P_2}, \qquad r \in \mathbb{R}.$$

The vector $\overrightarrow{P_1P_2}$ here gives the direction of the line, while the vector $\overrightarrow{OP_1}$ gives one point contained in the line.

Simplest example: the line passes through the origin



A vector equation for the line ℓ_0 here is the following:

$$\begin{pmatrix} x \\ y \end{pmatrix} = r \cdot u = r \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \qquad r \in \mathbb{R}.$$

Practice Problem

Example 3 from last time. (i) Consider the line ℓ_1 in $\mathbb{R}^2$ given by the equation

$$\ell_1 : \ x - 2y = 0.$$

Find a vector equation representing ℓ_1.

(ii) Consider the line ℓ_2 in $\mathbb{R}^2$ with vector equation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \mu \in \mathbb{R}.$$

Find a linear equation in normal form representing ℓ_2.

Recall: Answer to Example 3

(i) We need to find two points on ℓ_1.

- By setting $x = 0$ and solving for y in the linear equation representing ℓ_1, we get that the origin $O(0, 0)$ belongs to ℓ_1.
- By setting $x = 2$ and solving for y in the equation representing ℓ_1, we get that the point $P_2(2, 1)$ belongs to ℓ_1 as well.

We conclude that a vector equation for ℓ_1 is the equation

$$\binom{x}{y} = r \cdot \overrightarrow{OP_2} = r \cdot \binom{2}{1}, \qquad r \in \mathbb{R}.$$

(ii) We use the given vector equation to find two points on ℓ_2, and then based on these we write the two-point form for ℓ_2 which we can quickly transform to the normal form.

- By setting $\mu = 0$, we see that one point belonging to ℓ_2 is the point $Q_1(x_1, y_1)$ with $x_1 = 1$, $y_1 = 1$.
- By setting $\mu = 1$, we find that a second point belonging to ℓ_2 is the point $Q_2(x_2, y_2)$ with $x_2 = 2$, $y_2 = 0$.

We conclude that an equation in two-point form for ℓ_2 is the equation

$$y - 1 = (2 - 1)(y - 1) = (0 - 1)(x - 1) = -x + 1.$$

Therefore $x + y - 2 = 0$ is an equation in normal form representing ℓ_2.

Vector equations of lines in $\mathbb{R}^3$

Vector equation of a line in $\mathbb{R}^3$

We will see that in $\mathbb{R}^3$ one linear equation can never represent a line.

Instead, we can describe a line by a system of two linear equations in the unknowns x, y, z,

and any such system *(except in some special cases, which we will soon understand)* represents a line (that is, the set of all points (x, y, z) whose coordinates solve the linear system forms a line).

E.g. the system

$$\left\{\begin{array}{rcrcrcr} x & + & y & + & z & = & 0 \\ x & + & 2y & + & 3z & = & -1 \end{array}\right\}$$

of linear equations defines a specific line in $\mathbb{R}^3$.

Question. How can we find a vector equation representing this line? What should such a vector equation look like?

Vector equation of a line in $\mathbb{R}^3$ (cont.)

E.g. the system

$$\left\{\begin{array}{rcrcrcr} x & + & y & + & z & = & 0 \\ x & + & 2y & + & 3z & = & -1 \end{array}\right\}$$

of linear equations defines a specific line in $\mathbb{R}^3$.

Question. How can we find a vector equation representing this line? What should such a vector equation look like?

Again, we rely on the key fact/observation that, if we know two different points on the line, then we can completely determine the line. In fact, if P_1, P_2 are points in $\mathbb{R}^3$ which are contained in the line represented by the above linear system (in other words, such that their coordinates satisfy the linear system), then a vector equation for the line will be

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{OP_1} + \lambda \cdot \overrightarrow{P_1P_2}, \qquad \lambda \in \mathbb{R}$$

(note that now all vectors appearing in this equation are 3-dimensional).

Finding a vector equation for the line in this specific example

We need to find two points in $\mathbb{R}^3$ whose coordinates satisfy the given linear system.

We can first try to 'manipulate' the system in order to get an equivalent system of (slightly) simpler form:

$$\left\{\begin{array}{rcrcrcr} x & + & y & + & z & = & 0 \\ x & + & 2y & + & 3z & = & -1 \end{array}\right\} \quad \overset{E_2 - E_1 \to E_2'}{\Longleftrightarrow}$$

$$\left\{\begin{array}{rcrcrcr} x & + & y & + & z & = & 0 \\ & & 3y & + & 4z & = & -1 \end{array}\right\}.$$

We now see that

- if we set $z = 0$, we obtain that the point $P_1(1/3, -1/3, 0)$ is contained in this line;
- if we set $z = -1/4$, we obtain that the point $P_2(1/4, 0, -1/4)$ is also contained in the line.

We thus find that a vector equation for the line is the equation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1/3 \\ -1/3 \\ 0 \end{pmatrix} + \lambda \cdot \begin{pmatrix} -1/12 \\ 1/3 \\ -1/4 \end{pmatrix}, \qquad \lambda \in \mathbb{R}.$$

Equations of planes in $\mathbb{R}^3$

Normal form

General form of a linear equation in unknowns x, y, z (with non-trivial solution set):

$$ax + by + cz + d = 0, \quad \text{where } a, b, c, d \in \mathbb{R}, \ (a, b, c) \neq (0, 0, 0).$$

Now, in $\mathbb{R}^3$, this linear equation represents a **plane**! That is, the set of all points $(x, y, z) \in \mathbb{R}^3$ satisfying this (linear) equation forms a plane of $\mathbb{R}^3$.

Question. Why is this called a *normal form* of the plane? Because, if we initially assume that $d = 0$, then the equation can be equivalently rewritten as

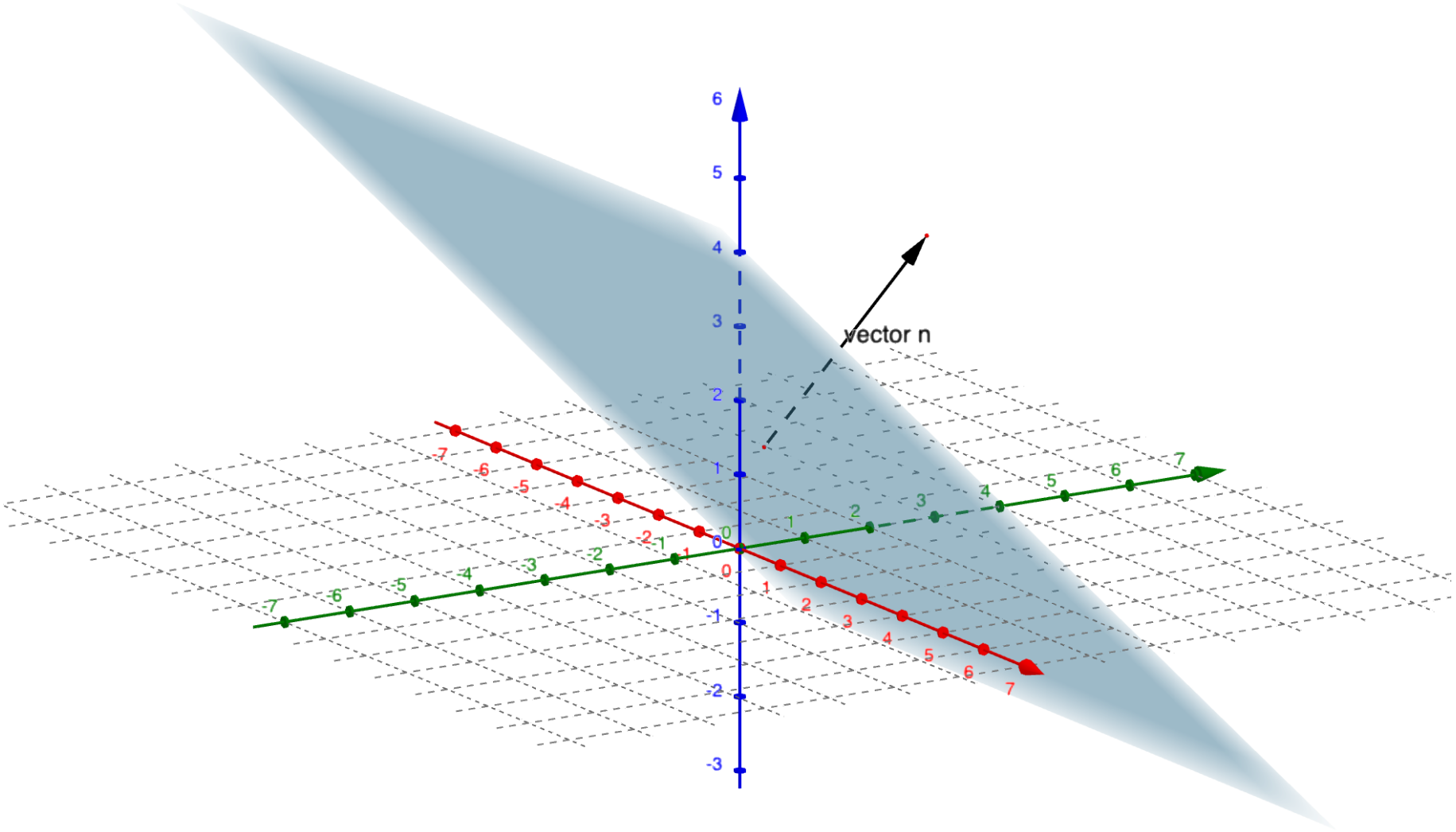
$$\left\langle \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\rangle = 0,$$

and is satisfied precisely by those points (x, y, z) such that the corresponding vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is orthogonal (or *normal*) to the vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

In fact, we also call the vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ a *normal vector* to the plane represented by the equation $ax + by + cz = 0$.

In the more general case, where $d \neq 0$, the set of all points satisfying the equation is merely a translate of a set of the latter type (and the vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is still called a normal vector to the plane).

Example



The vector in the picture is a normal vector to the given plane; note that n is orthogonal to any 'direction' contained in the plane (that is, any vector with initial and terminal points both belonging to the plane).

Vector equation of a plane in $\mathbb{R}^3$

Note that we can draw vectors / 'arrows' with several different directions within a given plane $\mathcal{P}$ in $\mathbb{R}^3$.

This suggests that we shouldn't be able to determine the plane simply by knowing one point contained in the plane and one 'direction' in the plane.

On the other hand, it turns out that, if we know **two different** directions in the plane, along with one point P_0 of it, then we can completely describe the plane $\mathcal{P}$.

Observe that this is equivalent to the fact which we (probably already) know from Euclidean geometry, that we can completely determine a plane in $\mathbb{R}^3$ if we know three different points contained in the plane **which are not collinear**.

Conclusion. Assuming that P_0, P_1, P_2 are three different points contained in $\mathcal{P}$, **which are not collinear**, one vector equation for $\mathcal{P}$ is the following:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{OP_0} + \lambda \cdot \overrightarrow{P_0P_1} + \mu \cdot \overrightarrow{P_0P_2}, \qquad \lambda, \mu \in \mathbb{R}.$$

Practice Problem

Example 4. Consider the plane $\mathcal{P}_1$ in $\mathbb{R}^3$ represented by the linear equation

$$x - y + 2z + 1 = 0\,.$$

Find a vector equation for $\mathcal{P}_1$.

Note. This is an equation of the form

$$ax + by + cz + d = 0 \qquad \text{with } d \neq 0,$$

so this plane will **not** contain the origin.

Answer to Example 4

First we need to find three non-collinear points contained in the plane $\mathcal{P}_1$.

- If we set $x = y = 0$ and solve for z in the equation representing $\mathcal{P}_1$, we obtain that the point $P_0(0, 0, -0.5)$ is contained in $\mathcal{P}_1$.
- If we set $x = y = 1$, we obtain that the point $P_1(1, 1, -0.5)$ is also contained in $\mathcal{P}_1$.
- Now, to find a third point P_2 in the plane $\mathcal{P}_1$ in such a way that the vectors $\overrightarrow{P_0P_1}$ and $\overrightarrow{P_0P_2}$ won't be parallel, we first note the following:

  $\overrightarrow{P_0P_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, so every vector parallel to $\overrightarrow{P_0P_1}$ (in other words, every scalar multiple of $\overrightarrow{P_0P_1}$) will have the property that **its first two components are equal.**

  On the other hand, the first two components of the vector $\overrightarrow{P_0P_2}$ will coincide with the first two coordinates of the point P_2; **thus if we pick $x = 0$, $y = 1$, and then solve for z in the equation representing $\mathcal{P}_1$, we get that the point $P_2(0, 1, 0)$ is contained in $\mathcal{P}_1$, and also that the points P_0, P_1, P_2 are not collinear.**

We conclude that a vector equation for the plane $\mathcal{P}_1$ is the equation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -0.5 \end{pmatrix} + \lambda \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \mu \cdot \begin{pmatrix} 0 \\ 1 \\ 0.5 \end{pmatrix}, \qquad \lambda, \mu \in \mathbb{R}.$$

Intersection of two planes of $\mathbb{R}^3$

Important Fact

Let $\mathcal{P}_1 : a_1x + b_1y + c_1z + d_1 = 0$ and $\mathcal{P}_2 : a_2x + b_2y + c_2z + d_2 = 0$ be two planes in $\mathbb{R}^3$. There are exactly three possibilities regarding the 'position' of these planes relative to each other (and in each instance exactly one of the following holds):

(i) the planes coincide;

(ii) the planes are parallel and have no common points;

(iii) the planes intersect in a line.

Note that one of (i) or (ii) will hold **if and only if** the normal vector $\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}$ to the plane $\mathcal{P}_1$ is parallel to the normal vector $\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}$ to the plane $\mathcal{P}_2$.

Otherwise the planes intersect in a line of $\mathbb{R}^3$, which we can then represent by the linear system

$$\left\{ \begin{array}{ccccccc} a_1x & + & b_1y & + & c_1z & = & -d_1 \\ a_2x & + & b_2y & + & c_2z & = & -d_2 \end{array} \right\}.$$

Next Main Topic:

A systematic study of Systems of Linear Equations with coefficients from a field $\mathbb{F}$

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 20

Wednesday October 7

Next Main Topic:

A systematic study of Systems of Linear Equations with coefficients from a field $\mathbb{F}$

Examples we have seen

Example 1. Consider the following linear system in 3 unknowns with coefficients from $\mathbb{R}$:

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & & & 2x_3 & = & 9 \end{array}\right\}.$$

Recall that we found its solution set via back substitution: it's the set $\{(-8.5,\, 1,\, 4.5)\}$ (in other words, we must have $x_1 = -8.5$, $x_2 = 1$, $x_2 = 4.5$).

Examples we have seen

Example 2. Consider the following system in 3 unknowns with coefficients from $\mathbb{R}$:

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ -x_1 & + & 3x_2 & - & 3x_3 & = & -2 \\ 2x_1 & & & + & 4x_3 & = & 1 \end{array}\right\} \quad \begin{array}{c} E_2+\frac{1}{3}E_1\to E_2' \\ E_3-\frac{2}{3}E_1\to E_3' \\ \Longleftrightarrow \end{array}$$

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & 4x_2 & - & \frac{2}{3}x_3 & = & 1 \end{array}\right\} \quad \begin{array}{c} E_3-4E_2\to E_3' \\ \Longleftrightarrow \end{array}$$

$$\left\{\begin{array}{rcrcrcr} 3x_1 & - & 6x_2 & + & 7x_3 & = & 0 \\ & & x_2 & - & \frac{2}{3}x_3 & = & -2 \\ & & & & 2x_3 & = & 9 \end{array}\right\}.$$

Examples we have seen

Example 3. Consider the following system in 4 unknowns with coefficients from $\mathbb{Z}_5$:

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{3E_1 \to E_1'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{E_2+E_1 \to E_2'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & 2x_2 & & & + & 2x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{3E_2 \to E_2'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & & & + & x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \end{array}\right\} \xrightarrow{E_3 - E_2 \to E_3'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & & & + & x_4 & = & 3 \\ & & & & 4x_3 & - & 2x_4 & = & 3 \end{array}\right\}.$$

Terminology

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$ (including the constant terms b_i):

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

A solution to $\mathcal{LS}_1$ is an n-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ with components from $\mathbb{F}$ such that setting $x_1 = \lambda_1, x_2 = \lambda_2, \ldots, x_n = \lambda_n$ turns all equations of the system into true (mathematical) statements.

There are 3 possibilities for the set of solutions of the system:

- it contains **no** solutions. *(in this case, the system is called inconsistent)*
- it contains **exactly one** solution.
- it contains more than one solution, and in fact at least as many as the number of elements of $\mathbb{F}$. *(so e.g. if $\mathbb{F} = \mathbb{R}$, there are infinitely many solutions in this case)*

In the latter two cases, the system is called consistent.

Terminology (cont.)

Moreover, we call an equation of the system *non-trivial* if at least one of its coefficients (including the constant term) is non-zero.

- The system is called *exactly determined* if the number n of unknowns is equal to the number of non-trivial equations.
- The system is called *overdetermined* if the number n of unknowns is strictly smaller than the number of non-trivial equations.
- The system is called *underdetermined* if the number n of unknowns is strictly larger than the number of non-trivial equations.
  Clearly, the system will be underdetermined if the number n of unknowns is strictly larger than the number m of equations.

Recall: Another Viewpoint

We have said that the system

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

is consistent **if and only if** the vector

$$\begin{pmatrix} b_1 \\ b_2 \\ \ddots \\ b_{m-1} \\ b_m \end{pmatrix}$$

is in the linear span of the vectors

$$\begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \ddots \\ a_{m-1,1} \\ a_{m,1} \end{pmatrix}, \quad \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \ddots \\ a_{m-1,2} \\ a_{m,2} \end{pmatrix}, \ldots\ldots\ldots, \quad \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \ddots \\ a_{m-1,n} \\ a_{m,n} \end{pmatrix}.$$

Equivalent systems

Terminology

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ and $\mathcal{LS}_2$ be two systems of linear equations with coefficients from $\mathbb{F}$ in the same set of unknowns (say $x_1, x_2, \ldots, x_n$).

The systems are called *equivalent* if they have the same set of solutions.

From the examples we have seen so far, it might be already clear that, for certain systems of linear equations it's considerably easier (and even less 'costly' computationally) than it is for others to determine their set of solutions (for instance, when we can use back substitution or forward substitution).

This suggests the following approach:

Question / Idea: Given a system of linear equations $\mathcal{LS}_1$, is it guaranteed that we can find another system $\mathcal{LS}_2$ that is 'easier' to solve and which is equivalent to the first system? (In which case, by finding the set of solutions to $\mathcal{LS}_2$, we will have solved $\mathcal{LS}_1$ as well.)

Method of *Gaussian elimination / row reduction*

The following operations / ways of manipulating a linear system always give an equivalent system:

1. Multiplying (both sides of) a linear equation of the system **by a non-zero constant**.
2. Adding a multiple of one equation of the system, say $Eq(i)$, to another equation, say $Eq(j)$, and replacing $Eq(j)$ by the new equation we get *(note that, here, $Eq(i)$ is kept where it was, unchanged).*
3. Swapping two equations (that is, writing $Eq(i)$ where we had $Eq(j)$ before, and writing $Eq(j)$ where we had $Eq(i)$).

Question. Why does an application of either of these operations lead to an equivalent system?

Revisiting a previous example (slightly changed)

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xleftrightarrow{E_1 \leftrightarrow E_2}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xleftrightarrow{3E_1 \to E_1'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xleftrightarrow{E_3+E_1 \to E_3'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ & & 2x_2 & & & + & 2x_4 & = & 1 \end{array}\right\} \xleftrightarrow{E_3-2E_2 \to E_3'}$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ & & & & 2x_3 & + & 4x_4 & = & 4 \end{array}\right\}.$$

Staircase Systems of Linear Equations

Definition

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

The system is called *an (upper) staircase* system if the following hold:

1. if there are any trivial equations, they are below any non-trivial equation;
2. the first non-zero coefficient of every non-trivial equation is found (strictly) to the right of the first non-zero coefficient of any previous equation.

Staircase Systems of Linear Equations (cont.)

Terminology

- Given a staircase system $\mathcal{LS}_0$, we call *pivot* any coefficient (including perhaps a constant term coefficient) which is the **first non-zero** coefficient in an equation.

  Note. Clearly there is only one pivot in each non-trivial equation of the system. Also by the second condition we stated above, there is **at most** one pivot in each column.

- We call a column a *pivot column* if it contains a coefficient that is a pivot.

- Finally, any variable / unknown found in a pivot column is called a *pivot variable*, whereas any variable that is **not** in a pivot column is called a *free variable*.

  Note. It the system is underdetermined, in particular if $m < n$, then we necessarily have free variables (why?).

The method of Gaussian elimination works:

it allows us to determine the solution set of every system of linear equations (with coefficients from a field $\mathbb{F}$)

Analysing the method

To reach the conclusion we just stated, we could try proving two things:

- That for every system $\mathcal{LS}_1$ of linear equations in unknowns $x_1, x_2, \ldots, x_n$ with coefficients from a field $\mathbb{F}$, we can find an **equivalent** staircase system $\mathcal{LS}_0$ of linear equations in the same unknowns (via consecutive applications of the method of Gaussian elimination);

- That we can (fairly efficiently) determine the solution set of every staircase linear system $\mathcal{LS}_0$ with coefficients from $\mathbb{F}$.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 21

Friday October 9

A systematic study of Systems of Linear Equations

Recall: Equivalent systems

Terminology

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ and $\mathcal{LS}_2$ be two systems of linear equations with coefficients from $\mathbb{F}$ in the same set of unknowns (say $x_1, x_2, \ldots, x_n$).

The systems are called *equivalent* if they have the same set of solutions.

Question / Idea: Given a system of linear equations $\mathcal{LS}_1$, is it guaranteed that we can find another system $\mathcal{LS}_2$ that is 'easier' to solve and which is equivalent to the first system? (In which case, by finding the set of solutions to $\mathcal{LS}_2$, we will have solved $\mathcal{LS}_1$ as well.)

Method of *Gaussian elimination / row reduction*

The following operations / ways of manipulating a linear system always give an equivalent system:

1. Multiplying (both sides of) a linear equation of the system **by a non-zero constant**.
2. Adding a multiple of one equation of the system, say $Eq(i)$, to another equation, say $Eq(j)$, and replacing $Eq(j)$ by the new equation we get *(note that, here, $Eq(i)$ is kept where it was, unchanged).*
3. Swapping two equations (that is, writing $Eq(i)$ where we had $Eq(j)$ before, and writing $Eq(j)$ where we had $Eq(i)$).

Question. Why does an application of either of these operations lead to an equivalent system?

Recall: Staircase Systems of Linear Equations

Definition

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

The system is called *an (upper) staircase* system if the following hold:

1. if there are any trivial equations, they are below any non-trivial equation;
2. the first non-zero coefficient of every non-trivial equation is found (strictly) to the right of the first non-zero coefficient of any previous equation.

Recall: Staircase Systems of Linear Equations (cont.)

Terminology

- Given a staircase system $\mathcal{LS}_0$, we call *pivot* any coefficient (including perhaps a constant term coefficient) which is the **first non-zero** coefficient in an equation.

  Note. Clearly there is only one pivot in each non-trivial equation of the system. Also by the second condition we stated above, there is **at most** one pivot in each column.

- We call a column a *pivot column* if it contains a coefficient that is a pivot.

- Finally, any variable / unknown found in a pivot column is called a *pivot variable*, whereas any variable that is **not** in a pivot column is called a *free variable*.

  Note. It the system is underdetermined, in particular if $m < n$, then we necessarily have free variables (why?).

One immediate consequence of the definition of 'staircase system'

If $\mathcal{LS}_0$ is a staircase system of m linear equations, then

- the first non-zero coefficient a_{2,j_2} in the 2nd equation has to be to the right of the first non-zero coefficient in the 1st equation, and thus we must have $j_2 \geqslant 2$;
- the first non-zero coefficient a_{3,j_3} in the 3rd equation has to be to the right of the first non-zero coefficient in the 2nd equation, and thus we must have $j_3 > j_2 \geqslant 2 \Rightarrow j_3 \geqslant 3$;
- continuing like this, we see that, for $2 \leqslant k \leqslant m$, the first non-zero coefficient a_{k,j_k} in the k-th equation (if any such zero coefficients exist) has to be to the right of the first non-zero coefficient in the $(k-1)$-th equation, and thus inductively we obtain that $j_k > j_{k-1} > \cdots > j_3 > j_2 \geqslant 2 \Rightarrow j_k \geqslant k$.

In other words,

Conclusion

For a staircase system of m linear equations, we necessarily have that $a_{k,j} = 0$ whenever $j < k$ (in other words, coefficients below the 'main diagonal' must be zero).

Note that this is not enough for the linear system to be staircase: e.g. the following is **not** a staircase system, but satisfies the above conclusion:

$$\left\{ \begin{array}{rcrcrcl} 2x_1 & - & 8x_2 & + & x_3 & + & x_4 = 0 \\ & & & & 7x_3 & - & x_4 = 1 \\ & & & & -x_3 & + & 6x_4 = 0.8 \end{array} \right\}.$$

The method of Gaussian elimination works:

it allows us to determine the solution set of every system of linear equations (with coefficients from a field $\mathbb{F}$)

Analysing the method

To reach the conclusion we just stated, we could try proving two things:

- That for every system $\mathcal{LS}_1$ of linear equations in unknowns $x_1, x_2, \ldots, x_n$ with coefficients from a field $\mathbb{F}$, we can find an **equivalent** staircase system $\mathcal{LS}_0$ of linear equations in the same unknowns (via consecutive applications of the method of Gaussian elimination);

- That we can (fairly efficiently) determine the solution set of every staircase linear system $\mathcal{LS}_0$ with coefficients from $\mathbb{F}$.

Reminder: Solving one linear equation $ax = b$ in one unknown

There are exactly three cases to consider:

Case 1: $a = b = 0$. Then, **no matter what element of $\mathbb{F}$ we set x equal to**, the equality will hold true. **In other words, we have as many solutions as the elements of $\mathbb{F}$.**

Case 2: $a = 0$, $b \neq 0$. Then, no matter what element of $\mathbb{F}$ we set x equal to, the LHS will equal 0, while the RHS will be non-zero $\rightsquigarrow$ **absurd!** thus we have no solutions

Case 3: $a \neq 0$. Then, no matter what the exact value of a is, and no matter what b is, we can find a **unique solution**. How? Note that $a \neq 0$ implies that a^{-1} exists. But then, we can multiply both sides of the equation by a^{-1}, and also use the associativity of the multiplication, to obtain that

$$x = (a^{-1}a)x = a^{-1} \cdot (ax) = a^{-1} \cdot b.$$

Note that this tells us that the only value that would make the equality true is the element $a^{-1} \cdot b$ (and also that this value is a solution, since $a \cdot (a^{-1}b) = b$ as we wanted).

Analogously...

Let $\mathcal{LS}_1$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from a field $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

There are exactly three possibilities for the set of solutions to $\mathcal{LS}_1$:

- it contains **no** solutions. *(in this case, the system is called inconsistent)*
- it contains **exactly one** solution.
- it contains more than one solution, and in fact at least as many as the number of elements of $\mathbb{F}$. *(so e.g. if $\mathbb{F} = \mathbb{R}$, there are infinitely many solutions in this case, but if e.g. $\mathbb{F} = \mathbb{Z}_p$ for some prime p, we will have finitely many, but definitely more than one)*

In the latter two cases, the system is called consistent.

In the case of a staircase system

Theorem 1

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a staircase system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$ (and at least one non-trivial equation).

(I) If there is a pivot in the column of constant terms, then the system is inconsistent.

(II) Otherwise, if there is no pivot in the column of constant terms, then the system is consistent. Moreover,

(a) if there are n pivots, or in other words (in this case) all variables are pivot variables, there is a **unique** solution to $\mathcal{LS}_0$.

(b) if there are fewer than n pivots, then we will have **free variables**. In this case, if the number of free variables is k (with $k \geqslant 1$ and $k < n$ (why?)), then the set of solutions

- is infinite if $\mathbb{F}$ is an infinite set;
- contains $|\mathbb{F}|^k$ different solutions if $\mathbb{F}$ is finite, where $|\mathbb{F}|$ denotes the cardinality of $\mathbb{F}$.

See next lecture for the proof of the theorem.

Past Exam Problem

Consider the following systems of linear equations. Determine whether they are staircase systems, and for each one of them that is, determine the size of its solution set.

$$\left\{\begin{array}{rcrcrcrcrcl} x_1 & + & 0x_2 & + & 2x_3 & + & 0x_4 & + & 9x_5 & = & 7 \\ 0x_1 & - & 7x_2 & + & 8x_3 & - & 4x_4 & + & 0x_5 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 4x_4 & + & x_5 & = & 5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & - & 3x_5 & = & 8 \end{array}\right\} \text{ (coefficients from } \mathbb{Z}_{11}),$$

$$\left\{\begin{array}{rcrcrcrcl} x_1 & + & 0x_2 & + & 2x_3 & + & 4x_4 & = & 3 \\ 0x_1 & + & 3x_2 & + & 0x_3 & - & 4x_4 & = & 0 \\ 0x_1 & + & 0x_2 & + & 4x_3 & + & x_4 & = & 1 \\ 0x_1 & + & 0x_2 & + & 0x_3 & - & 3x_4 & = & 2 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & = & 0 \end{array}\right\} \text{ (coefficients from } \mathbb{Z}_5),$$

$$\left\{\begin{array}{rcrcrcrcrcrcl} 2x_1 & - & 3.5x_2 & + & 17x_3 & + & 0x_4 & + & 9x_5 & + & x_6 & = & 2 \\ 0x_1 & + & 2x_2 & + & 0x_3 & - & 4x_4 & + & 10x_5 & + & 21x_6 & = & 5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 3x_5 & + & 0x_6 & = & -1 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & -2x_4 & + & x_5 & + & 0x_6 & = & 2.5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 8x_5 & + & 17x_6 & = & 6 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 0x_5 & + & x_6 & = & 3 \end{array}\right\}$$

(coefficients from $\mathbb{Q}$).

Suggested solution

Regarding the first system: this has no trivial equations, so the first condition in the definition of a staircase system is satisfied.

We now find the first non-zero coefficient of each of its equations; they are highlighted below:

$$\left\{\begin{array}{rcrcrcrcrcl} 1x_1 & + & 0x_2 & + & 2x_3 & + & 0x_4 & + & 9x_5 & = & 7 \\ 0x_1 & + & -7x_2 & + & 8x_3 & - & 4x_4 & + & 0x_5 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 4x_4 & + & x_5 & = & 5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & -3x_5 & = & 8 \end{array}\right\}.$$

We can verify that each such coefficient is to the right of the previous first non-zero coefficients of an equation of the system, and thus the second condition in the definition of a staircase system is satisfied as well.

The above observations combined show that this system is a staircase system.

Also, none of its pivots, namely the coefficients we highlighted, is in the column of constant terms, therefore the system is consistent.

Note finally that the system has 1 free variable, the variable x_3 (since none of the coefficients of x_3 is a pivot of the system). Thus its solution set will contain $|\mathbb{Z}_{11}| = 11$ different solutions.

Suggested solution (cont.)

Regarding the second system: it has only one trivial equation, which appears below any other equation of the system. Thus the first condition in the definition of a staircase system is satisfied.

We now find the first non-zero coefficient of each of its non-trivial equations; they are highlighted below:

$$\left\{\begin{array}{rcrcrcrcl} 1x_1 & + & 0x_2 & + & 2x_3 & + & 4x_4 & = & 3 \\ 0x_1 & + & 3x_2 & + & 0x_3 & - & 4x_4 & = & 0 \\ 0x_1 & + & 0x_2 & + & 4x_3 & + & x_4 & = & 1 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & -3x_4 & = & 2 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & = & 0 \end{array}\right\}.$$

We can verify that each such coefficient is to the right of the previous first non-zero coefficients of a non-trivial equation of the system, and thus the second condition in the definition of a staircase system is satisfied as well.

The above observations combined show that this system is a staircase system too.

Moreover, we see here as well that none of the pivots is in the column of constant terms, therefore the second system is consistent.

At the same time, all the variables here are pivot variables, therefore the second system has a unique solution.

Suggested solution (cont.)

Regarding the last system: it has no trivial equations, so the first condition in the definition of a staircase system is satisfied.

We find again the first non-zero coefficient of each of the equations of this system; they are highlighted below:

$$\left\{\begin{array}{rcrcrcrcrcrcr} 2x_1 & - & 3.5x_2 & + & 17x_3 & + & 0x_4 & + & 9x_5 & + & x_6 & = & 2 \\ 0x_1 & + & 2x_2 & + & 0x_3 & - & 4x_4 & + & 10x_5 & + & 21x_6 & = & 5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 3x_5 & + & 0x_6 & = & -1 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & -2x_4 & + & x_5 & + & 0x_6 & = & 2.5 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 8x_5 & + & 17x_6 & = & 6 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & + & 0x_5 & + & 1x_6 & = & 3 \end{array}\right\}.$$

Here we see that the first non-zero coefficients of the 3rd equation, the 4th equation and the 5th equation of the system fail to satisfy the second condition in the definition of a staircase system, so this system is **not** staircase.

Question. Could you use Gaussian elimination in order to determine the size of the solution set of the third system of the problem? *(In other words, could you use Gaussian elimination in order to find a staircase system that is equivalent to this third system, and then apply Theorem 1 to the staircase system you will have found?)*

How would you use the method in this specific case?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 22

Tuesday October 13

1st Midterm, Problem 3 (ii)

Consider the following two subspaces of $\mathbb{R}^4$: $\mathrm{span}(\bar{x}, \bar{y})$ and $\mathrm{span}(\bar{e}_2, \bar{e}_4)$, where

$$\bar{x} = \begin{pmatrix} 5 \\ -3 \\ 4 \\ 0.8 \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} 3 \\ 7 \\ 2.4 \\ 1 \end{pmatrix}, \quad \bar{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \bar{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Do these subspaces have any common **non-zero** vectors? Justify your answer.

(Here $\mathbb{R}^4$ should be viewed as a vector space over $\mathbb{R}$, with standard vector addition and scalar multiplication, as we defined them in class.)

1st Midterm, Problem 4

Recall that in HW1, Problem 5 you had to show that $\mathbb{Z}_3^2$ with operations

$$\begin{aligned} \text{addition} \qquad & (a,b)+(c,d) := (a+c, b+d) \\ \text{multiplication} \qquad & (a,b)\cdot(c,d) := (ac-bd, ad+bc) \end{aligned}$$

is a field. Let us write $\mathbb{K}$ for this field.

Consider now the vector space $\mathbb{K}^2$ over $\mathbb{K}$ with coordinate-wise vector addition and scalar multiplication (you don't need to verify that this is a vector space over $\mathbb{K}$).

Determine whether the vector

$$\bar{u} = \begin{pmatrix} (1,2) \\ (2,1) \end{pmatrix}$$

can be written as a linear combination of the vectors

$$\bar{w}_1 = \begin{pmatrix} (2,1) \\ (2,2) \end{pmatrix} \quad \text{and} \quad \bar{w}_2 = \begin{pmatrix} (2,0) \\ (1,1) \end{pmatrix}$$

of the form

$$\lambda \cdot \bar{w}_1 \; + \; \bar{w}_2$$

(that is, a linear combination where the second scalar is equal to the multiplicative identity in $\mathbb{K}$). Justify your answer.

[*Clarification.* Here we write $\mathbb{Z}_3 = \{0,1,2\}$, that is, we omit brackets.]

Recall: Staircase Systems of Linear Equations

Definition

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

The system is called *an (upper) staircase* system if the following hold:

1. if there are any trivial equations, they are below any non-trivial equation;
2. the first non-zero coefficient of every non-trivial equation is found (strictly) to the right of the first non-zero coefficient of any previous equation.

Recall: Staircase Systems of Linear Equations (cont.)

Terminology

- Given a staircase system $\mathcal{LS}_0$, we call *pivot* any coefficient (including perhaps a constant term coefficient) which is the **first non-zero** coefficient in an equation.

  Note. Clearly there is only one pivot in each non-trivial equation of the system. Also by the second condition we stated above, there is **at most** one pivot in each column.

- We call a column a *pivot column* if it contains a coefficient that is a pivot.

- Finally, any variable / unknown found in a pivot column is called a *pivot variable*, whereas any variable that is **not** in a pivot column is called a *free variable*.

  Note. It the system is underdetermined, in particular if $m < n$, then we necessarily have free variables (why?).

Reminder from last time

Theorem 1

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a staircase system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$ (and at least one non-trivial equation).

(I) If there is a pivot in the column of constant terms, then the system is inconsistent.

(II) Otherwise, if there is no pivot in the column of constant terms, then the system is consistent. Moreover,

(a) if there are n pivots, or in other words (in this case) all variables are pivot variables, there is a **unique** solution to $\mathcal{LS}_0$.

(b) if there are fewer than n pivots, then we will have **free variables**. In this case, if the number of free variables is k (with $k \geqslant 1$ and $k < n$ (why?)), then the set of solutions

- is infinite if $\mathbb{F}$ is an infinite set;
- contains $|\mathbb{F}|^k$ different solutions if $\mathbb{F}$ is finite, where $|\mathbb{F}|$ denotes the cardinality of $\mathbb{F}$.

Justification of Theorem 1

(I) If there is a pivot in the column of constant terms:

then the system must have an equation of the form

$$0x_1 + 0x_2 + \cdots + 0x_{n-1} + 0x_n = b_{i_0}$$

where $b_{i_0} \neq 0$ (note that b_{i_0} is the pivot found in this equation, and it is the pivot in the column of constant terms which we assume exists in this case). But this equation has no solutions, and thus the whole linear system cannot have solutions either.

Justification of Theorem 1 (cont.)

(II a) If there is **no** pivot in the column of constant terms, and we have n pivots (thus, all variables are pivot variables):

then the system must have the following form

$$\left\{\begin{array}{ccccccccccc}
a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n-1}x_{n-1} & + & a_{1,n}x_n & = & b_1 \\
 & & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n-1}x_{n-1} & + & a_{2,n}x_n & = & b_2 \\
 & & & & \ddots\ddots & \ddots & \ddots\ddots & & \ddots & & \ddots \\
 & & & & & & a_{n-1,n-1}x_{n-1} & + & a_{n-1,n}x_n & = & b_{n-1} \\
 & & & & & & & & a_{n,n}x_n & = & b_n
\end{array}\right\}$$

with perhaps a few more trivial equations at the bottom where it is guaranteed that the coefficients

$$a_{1,1}, \quad a_{2,2}, \ \ldots\ldots, \quad a_{n-1,n-1}, \quad a_{n,n}$$

are **non-zero** elements of $\mathbb{F}$ *(these will be the pivots of the system).*

Justification of Theorem 1 (cont.)

(II a) (cont.) If there is **no** pivot in the column of constant terms, and we have n pivots (thus, all variables are pivot variables):

But then we can check that the system has a unique solution which we can find via back substitution. In more detail,

- we can solve for x_n in the last equation we wrote above, and we will find a unique value for x_n that would satisfy this last equation (this also implies that any solution to the entire system must also set x_n equal to this value, because otherwise the last equation would not be satisfied).
- Once we solve for x_n, we can replace it in all the previous equations too by the value we found for it. In particular, this will turn the $(n-1)$-th equation into a linear equation in one unknown now, the unknown x_{n-1}, which is multiplied by a non-zero coefficient. Thus, similarly to before, we can solve for x_{n-1}, and we will find a unique value for it.
- Continuing like this, we see that, based on the values we will have already found for, say, the last $n-j$ variables, there is a **unique** value for the variable x_j that would satisfy the j-th equation (here it is understood that, in this j-th equation, we have already replaced the last $n-j$ unknowns by the values we found for them).

Justification of Theorem 1 (cont.)

(II b) If there is **no** pivot in the column of constant terms, and we have **fewer than *n*** pivots (so there are free variables, say k free variables):

we can then suppose that the free variables are

$$x_{j_1},\ x_{j_2},\ \ldots\ldots,\ x_{j_{k-1}},\ x_{j_k}$$

with $j_1 < j_2 < \cdots < j_{k-1} < j_k$, while the pivot variables are

$$x_{i_1},\ x_{i_2},\ \ldots\ldots,\ x_{i_{r-1}},\ x_{i_r}$$

where $r = n - k$ and $i_1 < i_2 < \cdots < i_{r-1} < i_r$.

But then, because of the commutativity and the associativity of addition in $\mathbb{F}$, we could rewrite the entire system we have as follows:

$$\left\{\begin{array}{ccccc}
a_{1,i_1}x_{i_1} + a_{1,i_2}x_{i_2} + \cdots + a_{1,i_r}x_{i_r} & + & a_{1,j_1}x_{j_1} + a_{1,j_2}x_{j_2} + \cdots + a_{1,j_k}x_{j_k} & = & b_1 \\
a_{2,i_1}x_{i_1} + a_{2,i_2}x_{i_2} + \cdots + a_{2,i_r}x_{i_r} & + & a_{2,j_1}x_{j_1} + a_{2,j_2}x_{j_2} + \cdots + a_{2,j_k}x_{j_k} & = & b_2 \\
\ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots & \ddots & \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots & \ddots & \ddots \\
a_{r,i_1}x_{i_1} + a_{r,i_2}x_{i_2} + \cdots + a_{r,i_r}x_{i_r} & + & a_{r,j_1}x_{j_1} + a_{r,j_2}x_{j_2} + \cdots + a_{r,j_k}x_{j_k} & = & b_r
\end{array}\right\}$$

(and perhaps there are a few more trivial equations at the bottom)

Justification of Theorem 1 (cont.)

(II b) (cont.) If there is **no** pivot in the column of constant terms, and we have **fewer than n** pivots (so there are free variables, say k free variables):

But then, for every choice of values for the free variables, the system will take the following form:

$$\left\{\begin{array}{rcccccccc}
a_{1,i_1}x_{i_1} & + a_{1,i_2}x_{i_2} & + \cdots\cdots & + a_{1,i_{l-1}}x_{i_{l-1}} & + a_{1,i_l}x_{i_l} & + & c_1 & = & b_1 \\
 & a_{2,i_2}x_{i_2} & + \cdots\cdots & + a_{2,i_{l-1}}x_{i_{l-1}} & + a_{2,i_l}x_{i_l} & + & c_2 & = & b_2 \\
 & & \ddots\ddots & \ddots\ddots & \ddots & & \ddots & & \ddots \\
 & & & a_{l-1,i_{l-1}}x_{i_{l-1}} & + a_{l-1,i_l}x_{i_l} & + & c_{l-1} & = & b_{l-1} \\
 & & & & a_{l,i_l}x_{i_l} & + & c_l & = & b_l
\end{array}\right\}$$

(and perhaps there are a few more trivial equations at the bottom). Note that here we have already replaced the variables $x_{j_1}, x_{j_2}, \ldots, x_{j_{k-1}}$ and x_{j_k} by the values we chose for them, and hence expressions of the form

$$a_{t,j_1}x_{j_1} + a_{t,j_2}x_{j_2} + \cdots\cdots + a_{t,j_k}x_{j_k},$$

which are found in the second part of the LHS of each of the r equations above, are now replaced by a single constant c_t from the field $\mathbb{F}$.

Observe now that such a system is like the one we discussed in Case (IIa) above, and thus it has a unique solution for the unknowns/pivot variables $x_{i_1}, x_{i_2}, \ldots, x_{i_{r-1}}$ and x_{i_r}.

Justification of Theorem 1 (cont.)

(II b) (cont.) If there is **no** pivot in the column of constant terms, and we have **fewer than *n*** pivots (so there are free variables, say k free variables):

In other words, we have reached the following conclusion: for every choice of values for the k free variables, the entire system has a unique solution.

Thus, the number of different solutions to the system equals the number of different choices we have to assign values to the k free variables. This leads to the conclusion stated in the theorem, that is, that there are $|\mathbb{F}|^k$ different solutions to $\mathcal{LS}_0$, and completes the proof in case (IIb) too *(note that $|\mathbb{F}|^k$ is infinite when $\mathbb{F}$ is an infinite set).*

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 23

Wednesday October 14

How the method of Gaussian elimination works

In other words, given a system $\mathcal{LS}_1$ of linear equations in unknowns $x_1, x_2, \ldots, x_n$ with coefficients from a field $\mathbb{F}$, how can we use the method of Gaussian elimination to find the solution set to this system?

We have two main steps to go through:

Step 1: We need to show that we can find a staircase system $\mathcal{LS}_0$ of linear equations in the same unknowns which is **equivalent** to $\mathcal{LS}_1$ **(and this is done via consecutive applications of the method of Gaussian elimination).**

Step 2: We need to verify that we can (fairly efficiently) determine the solution set of any staircase linear system $\mathcal{LS}_0$ with coefficients from $\mathbb{F}$.

Reminder (regarding Step 2 of the process)

Theorem 1 in the last two lectures

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a staircase system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$.

(I) If there is a pivot in the column of constant terms, then the system is inconsistent.

(II) Otherwise, if there is no pivot in the column of constant terms, then the system is consistent. Moreover,

(a) if there are n pivots, or in other words (in this case) all variables are pivot variables, there is a **unique** solution to $\mathcal{LS}_0$.

(b) if there are fewer than n pivots, then we will have **free variables**. In this case, if the number of free variables is k (with $k \geqslant 1$ and $k < n$ (why?)), then the set of solutions

- is infinite if $\mathbb{F}$ is an infinite set;
- contains $|\mathbb{F}|^k$ different solutions if $\mathbb{F}$ is finite, where $|\mathbb{F}|$ denotes the cardinality of $\mathbb{F}$.

What about Step 1? How do we justify it?

First, we introduce a slight change in notation,
and at the same time a new concept:

More convenient notation?

$$\left\{\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}$$

We can choose to stop writing the variables in each equation, assuming we already know what the unknowns of the linear system are, and assuming we have been consistent in arranging all the terms involving, say, the variable x_i to appear in the same column, the i-th column, as we have done above.

It should be clear then that **we could simply store the coefficients** of the unknowns in each equation (still arranged in the same way), as well as the vector of the constant terms, given that this information alone allows us to get back the original system at any point:

$$\begin{array}{ccccc} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{array} \quad \text{and} \quad \begin{pmatrix} b_1 \\ b_2 \\ \ddots \\ b_{m-1} \\ b_m \end{pmatrix}$$

Further notation

Often we also include all the coefficients of the unknowns together with the constant terms in the equations in one arrangement as follows:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{pmatrix}.$$

Moreover, for further clarity sometimes, we draw a vertical line segment before the column of the constant terms:

$$\left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{array}\right).$$

These 'arrangements' turn out to be very important in Linear Algebra.

Matrices

Definition

Let $\mathbb{F}$ be a field. A *matrix* with entries from $\mathbb{F}$, and with m rows and n columns, is an array of elements from $\mathbb{F}$ arranged as follows:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

(again note that here, for all indices i, j with $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, $a_{i,j}$ is an element of $\mathbb{F}$).

The set of all these arrays is denoted by $\mathbb{F}^{m\times n}$, and we refer to its elements as $m \times n$ matrices with entries from $\mathbb{F}$.

We'll soon discuss 'nice' properties that the set $\mathbb{F}^{m\times n}$ has.

Operations with matrices

Based on what we have just discussed,

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{k-1} \\ x_k \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$$

is another way to denote the 'variable' vector

$$\begin{pmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots\cdots + a_{1,n}x_n \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots\cdots + a_{2,n}x_n \\ \ddots\quad\ddots\quad\ddots\quad\ddots \\ a_{m-1,1}x_1 + a_{m-1,2}x_2 + \cdots\cdots + a_{m-1,n}x_n \\ a_{m,1}x_1 + a_{2,2}x_2 + \cdots\cdots + a_{m,n}x_n \end{pmatrix},$$

in other words, a more 'compact' notation for it.

Completely analogously we can now give the following definition:

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Given a matrix $A \in \mathbb{F}^{m \times n}$ and a vector $\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$, we set

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} = ???$$

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Given a matrix $A \in \mathbb{F}^{m \times n}$ and a vector $\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$, we set

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} =$$

$$\begin{pmatrix} a_{1,1}c_1 + a_{1,2}c_2 + \cdots\cdots + a_{1,n}c_n \\ a_{2,1}c_1 + a_{2,2}c_2 + \cdots\cdots + a_{2,n}c_n \\ \ddots\quad\ddots\quad\ddots\quad\ddots \\ a_{m-1,1}c_1 + a_{m-1,2}c_2 + \cdots\cdots + a_{m-1,n}c_n \\ a_{m,1}c_1 + a_{2,2}c_2 + \cdots\cdots + a_{m,n}c_n \end{pmatrix} \in \mathbb{F}^m.$$

An 'auxiliary', but quite useful, notion

Let $\mathbb{F}$ be a field (not necessarily the field $\mathbb{R}$ of real numbers).

Definition: *Dot Product*

Let $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-1} \\ d_n \end{pmatrix}$ be two vectors in $\mathbb{F}^n$.

Then we can still define the *dot product* $\langle \bar{c}, \bar{d} \rangle$ of $\bar{c}$ and $\bar{d}$ as follows:

$$\langle \bar{c}, \bar{d} \rangle \ := \ \sum_{i=1}^{n} c_i d_i \ = \ c_1 d_1 + c_2 d_2 + \cdots + c_{n-1} d_{n-1} + c_n d_n.$$

Important Observation

Unless $\mathbb{F}$ is the field $\mathbb{R}$ of real numbers, or the field $\mathbb{Q}$ of rational numbers (or even the field $\mathbb{C}$ or complex numbers as we will see later, in which case we will need to introduce a slightly different definition), the notion of 'dot product' is not as useful anymore because it doesn't have all the properties we discussed in previous lectures.

In fact,

— it still satisfies the **symmetry** and the **linearity (in the 1st argument)** (that is, properties (i), (ii) and (iii) of the Theorem in Lecture 17),

— but it is no longer **positive-definite**.

We can still make use of the notation though when working with matrices!

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Having introduced the dot-product notation, we can give again the definition of how we multiply a vector $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$ by a matrix $A \in \mathbb{F}^{m \times n}$:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} =$$

$$\begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \in \mathbb{F}^m.$$

Multiplication of two matrices?

What if we had two vectors $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}$ from $\mathbb{F}^n$, and we wanted to multiply both of them by the matrix $A \in \mathbb{F}^{m \times n}$ above? Suppose that we also decided first to write one vector next to the other:

$$\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \quad \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}.$$

Multiplication of two matrices?

What if we had two vectors $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}$ from $\mathbb{F}^n$, and we wanted to multiply both of them by the matrix $A \in \mathbb{F}^{m \times n}$ above? Suppose that we also decided first to write one vector next to the other:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \left[\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \quad \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix} \right]$$

$$= \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{d} \rangle \\ \langle \mathrm{Row}_2(A), \bar{d} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{d} \rangle \\ \langle \mathrm{Row}_m(A), \bar{d} \rangle \end{pmatrix}.$$

Then we could also write the outputs of the operations $A \cdot \bar{c}$ and $A \cdot \bar{d}$ one next to the other as above.

Multiplication of two matrices?

Observe now that the 'arrangements'

$$\left[\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}\right] \quad \text{and} \quad \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots \ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{d} \rangle \\ \langle \mathrm{Row}_2(A), \bar{d} \rangle \\ \ddots \ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{d} \rangle \\ \langle \mathrm{Row}_m(A), \bar{d} \rangle \end{pmatrix}$$

look a lot like the new objects we are discussing: in particular, a matrix with n rows and 2 columns, and a matrix with m rows and 2 columns respectively.

This motivates the following definition:

Multiplication of two matrices

Definition

Consider two matrices with entries from a field $\mathbb{F}$, say, a matrix $A \in \mathbb{F}^{m\times n}$ and a matrix $B \in \mathbb{F}^{k\times l}$. First of all, we introduce some

Notation. Often, to clarify how we will denote the entries of A or of B, we may also write $A = \big(a_{i,j}\big)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$, and analogously $B = \big(b_{r,s}\big)_{\substack{1\leqslant r\leqslant k\\ 1\leqslant s\leqslant l}}$.

If we have $n = k$, that is, if the number of columns of A is equal to the number of rows of B, then we can define $A \cdot B$ to be the matrix $C = \big(c_{i,s}\big)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant s\leqslant l}} \in \mathbb{F}^{m\times l}$ whose entries are given by the following rule:

$$c_{i,s} := \big\langle \mathrm{Row}_i(A), \mathrm{Col}_s(B)\big\rangle = \sum_{j=1}^{n} a_{i,j}\, b_{j,s}\,.$$

In other words, $A \cdot B :=$

$$\begin{pmatrix} \big\langle \mathrm{Row}_1(A), \mathrm{Col}_1(B)\big\rangle & \big\langle \mathrm{Row}_1(A), \mathrm{Col}_2(B)\big\rangle & \cdots\cdots & \big\langle \mathrm{Row}_1(A), \mathrm{Col}_l(B)\big\rangle \\ \big\langle \mathrm{Row}_2(A), \mathrm{Col}_1(B)\big\rangle & \big\langle \mathrm{Row}_2(A), \mathrm{Col}_2(B)\big\rangle & \cdots\cdots & \big\langle \mathrm{Row}_2(A), \mathrm{Col}_l(B)\big\rangle \\ \ddots & \ddots & \ddots\ \ddots & \ddots \\ \big\langle \mathrm{Row}_m(A), \mathrm{Col}_1(B)\big\rangle & \big\langle \mathrm{Row}_m(A), \mathrm{Col}_2(B)\big\rangle & \cdots\cdots & \big\langle \mathrm{Row}_m(A), \mathrm{Col}_l(B)\big\rangle \end{pmatrix}$$

An example

Consider two matrices with real entries:

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 0 & 0 \\ 4 & 1 & 5 \end{pmatrix} \in \mathbb{R}^{3\times 3} \qquad \text{and} \qquad B = \begin{pmatrix} 2 & 0 \\ 1 & 3 \\ -1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Question 1. Is the product $A \cdot B$ defined?

Answer. Yes, because the number of columns of A equals the number of rows of B. In fact, we will have $A \cdot B \in \mathbb{R}^{3\times 2}$ and

$$A \cdot B = \begin{pmatrix} \langle \mathrm{Row}_1(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_1(A), \mathrm{Col}_2(B) \rangle \\ \langle \mathrm{Row}_2(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_2(A), \mathrm{Col}_2(B) \rangle \\ \langle \mathrm{Row}_3(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_3(A), \mathrm{Col}_2(B) \rangle \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 6 & 0 \\ 4 & 8 \end{pmatrix}.$$

Question 2. Is the product $B \cdot A$ defined?

Answer. No, because the number of columns of B (which is 2) is different from the number of rows of A (which is 3).

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 24

Friday October 16

Reminder: Matrices

Definition

Let $\mathbb{F}$ be a field. A *matrix* with entries from $\mathbb{F}$, and with m rows and n columns, is an array of elements from $\mathbb{F}$ arranged as follows:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

(again note that here, for all indices i, j with $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, $a_{i,j}$ is an element of $\mathbb{F}$).

The set of all these arrays is denoted by $\mathbb{F}^{m\times n}$, and we refer to its elements as $m \times n$ matrices with entries from $\mathbb{F}$.

We'll soon discuss 'nice' properties that the set $\mathbb{F}^{m\times n}$ has.

Operations with matrices

Recall the convention from last time, that

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{k-1} \\ x_k \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix}$$

is another way to denote the 'variable' vector

$$\begin{pmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots\cdots + a_{1,n}x_n \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots\cdots + a_{2,n}x_n \\ \ddots \quad \ddots \quad \ddots \quad \ddots \\ a_{m-1,1}x_1 + a_{m-1,2}x_2 + \cdots\cdots + a_{m-1,n}x_n \\ a_{m,1}x_1 + a_{2,2}x_2 + \cdots\cdots + a_{m,n}x_n \end{pmatrix},$$

in other words, a more 'compact' notation for it.

Completely analogously we can now give the following definition:

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Given a matrix $A \in \mathbb{F}^{m \times n}$ and a vector $\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$, we set

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} = ???$$

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Given a matrix $A \in \mathbb{F}^{m \times n}$ and a vector $\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$, we set

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} =$$

$$\begin{pmatrix} a_{1,1}c_1 + a_{1,2}c_2 + \cdots\cdots + a_{1,n}c_n \\ a_{2,1}c_1 + a_{2,2}c_2 + \cdots\cdots + a_{2,n}c_n \\ \ddots\quad\ddots\quad\ddots\quad\ddots \\ a_{m-1,1}c_1 + a_{m-1,2}c_2 + \cdots\cdots + a_{m-1,n}c_n \\ a_{m,1}c_1 + a_{2,2}c_2 + \cdots\cdots + a_{m,n}c_n \end{pmatrix} \in \mathbb{F}^m.$$

An 'auxiliary', but quite useful, notion

Let $\mathbb{F}$ be a field (not necessarily the field $\mathbb{R}$ of real numbers).

Definition: *Dot Product*

Let $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-1} \\ d_n \end{pmatrix}$ be two vectors in $\mathbb{F}^n$.

Then we can still define the *dot product* $\langle \bar{c}, \bar{d} \rangle$ of $\bar{c}$ and $\bar{d}$ as follows:

$$\langle \bar{c}, \bar{d} \rangle \ := \ \sum_{i=1}^{n} c_i d_i \ = \ c_1 d_1 + c_2 d_2 + \cdots + c_{n-1} d_{n-1} + c_n d_n.$$

Important Observation

Unless $\mathbb{F}$ is the field $\mathbb{R}$ of real numbers, or the field $\mathbb{Q}$ of rational numbers (or even the field $\mathbb{C}$ or complex numbers as we will see later, in which case we will need to introduce a slightly different definition), the notion of 'dot product' is not as useful anymore because it doesn't have all the properties we discussed in previous lectures.

In fact,

— it still satisfies the **symmetry** and the **linearity (in the 1st argument)** (that is, properties (i), (ii) and (iii) of the Theorem in Lecture 17),

— but it is no longer **positive-definite**.

We can still make use of the notation though when working with matrices!

Multiplication of an n-dimensional vector by an $m \times n$ matrix

Having introduced the dot-product notation, we can give again the definition of how we multiply a vector $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \in \mathbb{F}^n$ by a matrix $A \in \mathbb{F}^{m \times n}$:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} =$$

$$\begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \in \mathbb{F}^m.$$

Multiplication of two matrices?

What if we had two vectors $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}$ from $\mathbb{F}^n$, and we wanted to multiply both of them by the matrix $A \in \mathbb{F}^{m \times n}$ above? Suppose that we also decided first to write one vector next to the other:

$$\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \quad \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}.$$

Multiplication of two matrices?

What if we had two vectors $\bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix}$ and $\bar{d} = \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}$ from $\mathbb{F}^n$, and we wanted to multiply both of them by the matrix $A \in \mathbb{F}^{m \times n}$ above? Suppose that we also decided first to write one vector next to the other:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix} \cdot \left[\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \quad \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix} \right]$$

$$= \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{d} \rangle \\ \langle \mathrm{Row}_2(A), \bar{d} \rangle \\ \ddots\ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{d} \rangle \\ \langle \mathrm{Row}_m(A), \bar{d} \rangle \end{pmatrix}.$$

Then we could also write the outputs of the operations $A \cdot \bar{c}$ and $A \cdot \bar{d}$ one next to the other as above.

Multiplication of two matrices?

Observe now that the 'arrangements'

$$\left[\begin{pmatrix} c_1 \\ c_2 \\ \ddots \\ c_{k-1} \\ c_k \\ \ddots \\ c_{n-1} \\ c_n \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \ddots \\ d_{k-1} \\ d_k \\ \ddots \\ d_{n-1} \\ d_n \end{pmatrix}\right] \quad \text{and} \quad \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{c} \rangle \\ \langle \mathrm{Row}_2(A), \bar{c} \rangle \\ \ddots \ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{c} \rangle \\ \langle \mathrm{Row}_m(A), \bar{c} \rangle \end{pmatrix} \begin{pmatrix} \langle \mathrm{Row}_1(A), \bar{d} \rangle \\ \langle \mathrm{Row}_2(A), \bar{d} \rangle \\ \ddots \ddots \\ \langle \mathrm{Row}_{m-1}(A), \bar{d} \rangle \\ \langle \mathrm{Row}_m(A), \bar{d} \rangle \end{pmatrix}$$

look a lot like the new objects we are discussing: in particular, a matrix with n rows and 2 columns, and a matrix with m rows and 2 columns respectively.

This motivates the following definition:

Multiplication of two matrices

Definition

Consider two matrices with entries from a field $\mathbb{F}$, say, a matrix $A \in \mathbb{F}^{m\times n}$ and a matrix $B \in \mathbb{F}^{k\times l}$. First of all, we introduce some

Notation. Often, to clarify how we will denote the entries of A or of B, we may also write $A = \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$, and analogously $B = \left(b_{r,s}\right)_{\substack{1\leqslant r\leqslant k\\ 1\leqslant s\leqslant l}}$.

If we have $n = k$, that is, if the number of columns of A is equal to the number of rows of B, then we can define $A \cdot B$ to be the matrix $C = \left(c_{i,s}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant s\leqslant l}} \in \mathbb{F}^{m\times l}$ whose entries are given by the following rule:

$$c_{i,s} := \left\langle \mathrm{Row}_i(A), \mathrm{Col}_s(B)\right\rangle = \sum_{j=1}^{n} a_{i,j}\, b_{j,s}\,.$$

In other words, $A \cdot B := \left(A\cdot \mathrm{Col}_1(B) \,\middle|\, A\cdot \mathrm{Col}_2(B) \,\middle|\, \cdots\cdots \,\middle|\, A\cdot \mathrm{Col}_l(B)\right) =$

$$\begin{pmatrix} \left\langle \mathrm{Row}_1(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_1(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_1(A), \mathrm{Col}_l(B)\right\rangle \\ \left\langle \mathrm{Row}_2(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_2(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_2(A), \mathrm{Col}_l(B)\right\rangle \\ \ddots & \ddots & \ddots\ddots & \ddots \\ \left\langle \mathrm{Row}_m(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_m(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_m(A), \mathrm{Col}_l(B)\right\rangle \end{pmatrix}$$

Example from last time

Consider two matrices with real entries:

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 0 & 0 \\ 4 & 1 & 5 \end{pmatrix} \in \mathbb{R}^{3\times 3} \qquad \text{and} \qquad B = \begin{pmatrix} 2 & 0 \\ 1 & 3 \\ -1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Question 1. Is the product $A \cdot B$ defined?

Answer. Yes, because the number of columns of A equals the number of rows of B. In fact, we will have $A \cdot B \in \mathbb{R}^{3\times 2}$ and

$$A \cdot B = \begin{pmatrix} \langle \mathrm{Row}_1(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_1(A), \mathrm{Col}_2(B) \rangle \\ \langle \mathrm{Row}_2(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_2(A), \mathrm{Col}_2(B) \rangle \\ \langle \mathrm{Row}_3(A), \mathrm{Col}_1(B) \rangle & \langle \mathrm{Row}_3(A), \mathrm{Col}_2(B) \rangle \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 6 & 0 \\ 4 & 8 \end{pmatrix}.$$

Question 2. Is the product $B \cdot A$ defined?

Answer. No, because the number of columns of B (which is 2) is different from the number of rows of A (which is 3).

More operations with matrices?

Addition of two matrices

Definition

Consider two matrices with entries from a field $\mathbb{F}$, say, a matrix $A = \left(a_{i,j}\right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \in \mathbb{F}^{m \times n}$ and a matrix $B = \left(b_{r,s}\right)_{\substack{1 \leqslant r \leqslant k \\ 1 \leqslant s \leqslant l}} \in \mathbb{F}^{k \times l}$.

If we have that $m = k$ and $n = l$, then we can define the matrix $A + B$ by doing *entry-wise addition*:

$$A + B := \begin{pmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \cdots\cdots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \cdots\cdots & a_{2,n} + b_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots \\ a_{m-1,1} + b_{m-1,1} & a_{m-1,2} + b_{m-1,2} & \cdots\cdots & a_{m-1,n} + b_{m-1,n} \\ a_{m,1} + b_{m,1} & a_{m,2} + b_{m,2} & \cdots\cdots & a_{m,n} + b_{m,n} \end{pmatrix}.$$

Observe that $A + B \in \mathbb{F}^{m \times n}$ as well.

Multiplication of a matrix by a scalar

Definition

Let $A = \left(a_{i,j}\right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$ be a matrix with entries from a field $\mathbb{F}$, and let $\lambda \in \mathbb{F}$.

We define

$$\lambda \cdot A := \begin{pmatrix} \lambda a_{1,1} & \lambda a_{1,2} & \cdots\cdots & \lambda a_{1,n-1} & \lambda a_{1,n} \\ \lambda a_{2,1} & \lambda a_{2,2} & \cdots\cdots & \lambda a_{2,n-1} & \lambda a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ \lambda a_{m-1,1} & \lambda a_{m-1,2} & \cdots\cdots & \lambda a_{m-1,n-1} & \lambda a_{m-1,n} \\ \lambda a_{m,1} & \lambda a_{m,2} & \cdots\cdots & \lambda a_{m,n-1} & \lambda a_{m,n} \end{pmatrix},$$

that is, scalar multiplication is defined *entry-wise*.

Observe that $\lambda \cdot A \in \mathbb{F}^{m \times n}$ as well.

Important Remark: when are two matrices equal?

Let $\mathbb{F}$ be a field, and let A, B be two matrices with entries from $\mathbb{F}$, $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{k \times l}$.

We have that $\boldsymbol{A = B}$ **if and only if**

- $\boldsymbol{m = k}$ and $\boldsymbol{n = l}$
- and for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$ the (i, j)-th entry of A **equals** the (i, j)-th entry of B.

Example. If A, B, C are the following matrices with entries from $\mathbb{Z}_5$:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 4 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \\ -1 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 4 \end{pmatrix},$$

then $A = B$ but $A \neq C \neq B$.

Properties of the operations we defined?

We can verify the following

Theorem

The set $\mathbb{F}^{m\times n}$ together with this addition and this scalar multiplication is a vector space over $\mathbb{F}$.

What about other properties

(involving the multiplication of matrices too)?

Past Homework Problem

Consider the following matrices:

$$A = \begin{pmatrix} 1 & 1 & -4 \\ 11 & -5 & 7 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{pmatrix} -1 & -2 \\ 10 & 9 \\ 8 & -6 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \quad \bar{c} = \begin{pmatrix} 1 & -5 \end{pmatrix} \in \mathbb{R}^{1\times 2},$$

$$D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad E = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 1 & 0 & 3 \end{pmatrix} \in \mathbb{Z}_5^{3\times 3}, \quad \bar{u} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{Z}_5^{3\times 1}.$$

Find all products of any two (different) matrices from above that make sense, as well as the following expressions, again if they are defined: $AB + I_2$, $A(I_3 + B)$, $AB\bar{c}$, $I_3 + (E^2 - E) \cdot (I_3 - E)^{-1}$ (is it possible to simplify any of the expressions you are asked to find, in order to do fewer and easier computations? can you simplify regardless of what the given matrices are?).

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 25

Monday October 19

Reminder: Matrices

Definition

Let $\mathbb{F}$ be a field. A *matrix* with entries from $\mathbb{F}$, and with m rows and n columns, is an array of elements from $\mathbb{F}$ arranged as follows:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

(again note that here, for all indices i, j with $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, $a_{i,j}$ is an element of $\mathbb{F}$).

The set of all these arrays is denoted by $\mathbb{F}^{m\times n}$, and we refer to its elements as $m \times n$ matrices with entries from $\mathbb{F}$.

Reminder: Equality of Matrices

Let $\mathbb{F}$ be a field, and let A, B be two matrices with entries from $\mathbb{F}$, $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{k \times l}$.

We have that $\boldsymbol{A = B}$ **if and only if**

- $\boldsymbol{m = k}$ and $\boldsymbol{n = l}$
- and for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$ the (i, j)-th entry of A **equals** the (i, j)-th entry of B.

Example. If A, B, C are the following matrices with entries from $\mathbb{Z}_5$:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 4 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \\ -1 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 4 \end{pmatrix},$$

then $A = B$ but $A \neq C \neq B$.

Multiplication of two matrices

Definition

Consider two matrices with entries from a field $\mathbb{F}$, say, a matrix $A \in \mathbb{F}^{m\times n}$ and a matrix $B \in \mathbb{F}^{k\times l}$. First of all, we introduce some

Notation. Often, to clarify how we will denote the entries of A or of B, we may also write $A = \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$, and analogously $B = \left(b_{r,s}\right)_{\substack{1\leqslant r\leqslant k\\ 1\leqslant s\leqslant l}}$.

If we have $n = k$, that is, if the number of columns of A is equal to the number of rows of B, then we can define $A \cdot B$ to be the matrix $C = \left(c_{i,s}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant s\leqslant l}} \in \mathbb{F}^{m\times l}$ whose entries are given by the following rule:

$$c_{i,s} := \left\langle \mathrm{Row}_i(A), \mathrm{Col}_s(B)\right\rangle = \sum_{j=1}^{n} a_{i,j}\, b_{j,s}\,.$$

In other words, $A \cdot B := \left(A\cdot \mathrm{Col}_1(B) \;\middle|\; A\cdot \mathrm{Col}_2(B) \;\middle|\; \cdots\cdots \;\middle|\; A\cdot \mathrm{Col}_l(B) \right) =$

$$\begin{pmatrix} \left\langle \mathrm{Row}_1(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_1(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_1(A), \mathrm{Col}_l(B)\right\rangle \\ \left\langle \mathrm{Row}_2(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_2(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_2(A), \mathrm{Col}_l(B)\right\rangle \\ \ddots & \ddots & \ddots\;\ddots & \ddots \\ \left\langle \mathrm{Row}_m(A), \mathrm{Col}_1(B)\right\rangle & \left\langle \mathrm{Row}_m(A), \mathrm{Col}_2(B)\right\rangle & \cdots\cdots & \left\langle \mathrm{Row}_m(A), \mathrm{Col}_l(B)\right\rangle \end{pmatrix}$$

Addition of two matrices

Definition

Consider two matrices with entries from a field $\mathbb{F}$, say, a matrix $A = \left(a_{i,j}\right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \in \mathbb{F}^{m \times n}$ and a matrix $B = \left(b_{r,s}\right)_{\substack{1 \leqslant r \leqslant k \\ 1 \leqslant s \leqslant l}} \in \mathbb{F}^{k \times l}$.

If we have that $m = k$ and $n = l$, then we can define the matrix $A + B$ by doing *entry-wise addition*:

$$A + B := \begin{pmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \cdots\cdots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \cdots\cdots & a_{2,n} + b_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots \\ a_{m-1,1} + b_{m-1,1} & a_{m-1,2} + b_{m-1,2} & \cdots\cdots & a_{m-1,n} + b_{m-1,n} \\ a_{m,1} + b_{m,1} & a_{m,2} + b_{m,2} & \cdots\cdots & a_{m,n} + b_{m,n} \end{pmatrix}.$$

Observe that $A + B \in \mathbb{F}^{m \times n}$ as well.

Multiplication of a matrix by a scalar

Definition

Let $A = \left(a_{i,j}\right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$ be a matrix with entries from a field $\mathbb{F}$, and let $\lambda \in \mathbb{F}$.

We define

$$\lambda \cdot A := \begin{pmatrix} \lambda a_{1,1} & \lambda a_{1,2} & \cdots\cdots & \lambda a_{1,n-1} & \lambda a_{1,n} \\ \lambda a_{2,1} & \lambda a_{2,2} & \cdots\cdots & \lambda a_{2,n-1} & \lambda a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ \lambda a_{m-1,1} & \lambda a_{m-1,2} & \cdots\cdots & \lambda a_{m-1,n-1} & \lambda a_{m-1,n} \\ \lambda a_{m,1} & \lambda a_{m,2} & \cdots\cdots & \lambda a_{m,n-1} & \lambda a_{m,n} \end{pmatrix},$$

that is, scalar multiplication is defined *entry-wise*.

Observe that $\lambda \cdot A \in \mathbb{F}^{m \times n}$ as well.

ATTENTION to the definitions

Let $A \in \mathbb{F}^{m \times n}$, $B \in \mathbb{F}^{k \times l}$ and $\bar{c} \in \mathbb{F}^{s} \equiv \mathbb{F}^{s \times 1}$.

- $A \cdot \bar{c}$ makes sense only when $n = s$ (the number of columns of the matrix A equals the number of components of the vector $\bar{c}$). Also, in such a case $A \cdot \bar{c} \in \mathbb{F}^{m}$.

- $A \cdot B$ makes sense only when $n = k$ (the number of columns of the matrix we write on the left equals the number of rows of the matrix we write on the right). Also, in such a case, $A \cdot B \in \mathbb{F}^{m \times l}$.

- $A + B$ makes sense only when $m = k$ and $n = l$. Also, in such a case, $A + B \in \mathbb{F}^{m \times n}$.

Important properties these operations satisfy

Reminder: A structure on $\mathbb{F}^{m\times n}$

Theorem

The set $\mathbb{F}^{m\times n}$ together with this addition and this scalar multiplication is a vector space over $\mathbb{F}$.

The theorem implies (and relies on checking) the following eight properties:

(i) For all $A, B \in \mathbb{F}^{m\times n}$, $A + B = B + A$.

(ii) For all $A, B, C \in \mathbb{F}^{m\times n}$, $(A + B) + C = A + (B + C)$.

(iii) If we write $\overline{O}$ for the matrix in $\mathbb{F}^{m\times n}$ all of whose entries are equal to $0_{\mathbb{F}}$, then $\overline{O}$ is the neutral element of addition in $\mathbb{F}^{m\times n}$: for all $A \in \mathbb{F}^{m\times n}$, we have that $\overline{O} + A = A = A + \overline{O}$.

(iv) Given an arbitrary matrix $A = \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} \in \mathbb{F}^{m\times n}$, the matrix $\left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$, which is also in $\mathbb{F}^{m\times n}$, is the additive inverse of A: we have that

$$\left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} + \left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} = \left(\left[a_{i,j} + (-a_{i,j})\right]\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} = \overline{O}$$

$$= \left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} + \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}.$$

Note that the additive inverse of A is also given by $(-1_{\mathbb{F}}) \cdot A$.

A structure on $\mathbb{F}^{m\times n}$ (cont.)

Theorem

The set $\mathbb{F}^{m\times n}$ together with this addition and this scalar multiplication is a vector space over $\mathbb{F}$.

Properties hinted at in the statement of the theorem, cont.:

(v) For all $A \in \mathbb{F}^{m\times n}$, $1_{\mathbb{F}} \cdot A = A$.

(vi) For all $A \in \mathbb{F}^{m\times n}$ and all scalars $\lambda, \mu \in \mathbb{F}$, we have that

$$\lambda \cdot \big(\mu \cdot A\big) = (\lambda\mu) \cdot A\,.$$

and the two Distributive Laws of a vector space structure:

(vii) For all $A, B \in \mathbb{F}^{m\times n}$ and all $\lambda \in \mathbb{F}$, we have that

$$\lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B\,.$$

(viii) For all $A \in \mathbb{F}^{m\times n}$ and all scalars $\lambda, \mu \in \mathbb{F}$, we have that

$$(\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A\,.$$

A few more important properties

- **Distributive Law** *(matrix multiplication distributes over matrix addition)* For all matrices $A, B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times l}$ and $E \in \mathbb{F}^{k\times m}$, we have that

$$(A+B)C = AC + BC, \qquad E(A+B) = EA + EB.$$

(you have to verify this yourselves in HW3, Problem 3)

- **Associativity** For all matrices $A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times k}$ and $C \in \mathbb{F}^{k\times l}$, we have that

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

(left as an exercise for now, we'll return to this again)

- *Special case of associativity* For all matrices $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$ and all scalars $\lambda \in \mathbb{F}$, we have that

$$\lambda \cdot (A \cdot B) = (\lambda \cdot A) \cdot B = A \cdot (\lambda \cdot B).$$

What about the order in which we write two matrices we want to multiply? Does it matter?

In general, matrix multiplication is not commutative.

That is, there exist (plenty of examples of) matrices A, B such that $A \cdot B \neq B \cdot A$.

First of all, when would both products $A \cdot B$ and $B \cdot A$ be defined?

Answer. Whenever we have $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times m}$ (check this yourselves).

In such a case, as we already have seen, $A \cdot B \in \mathbb{F}^{m \times m}$ and $B \cdot A \in \mathbb{F}^{n \times n}$. Thus,

- if $m \neq n$, we definitely have $A \cdot B \neq B \cdot A$. *(recall when we have equality of matrices)*
- if $m = n$, it may still happen that $A \cdot B \neq B \cdot A$. *(you are asked to find such examples in HW3, Problem 2)*

In general, multiplication in $\mathbb{F}^{n\times n}$ is not commutative

Terminology

A matrix $A \in \mathbb{F}^{n\times n}$ (that is, a matrix which has as many rows as columns) is called a *square* matrix.

When $A, B \in \mathbb{F}^{n\times n}$, it may happen that $A \cdot B = B \cdot A$: e.g. for the matrices

$$A = \begin{pmatrix} -3 & 1 & 1 \\ 0 & -5 & 3 \\ 0 & 1 & -5 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 3 \\ 0 & 1 & 0 \end{pmatrix}$$

in $\mathbb{R}^{3\times 3}$, we have that

$$AB = \begin{pmatrix} -6 & -2 & 0 \\ 0 & 3 & -15 \\ 0 & -5 & 3 \end{pmatrix} = BA\,.$$

Terminology

Given matrices $A, B \in \mathbb{F}^{n\times n}$, if we have that $A \cdot B = B \cdot A$, we say that the matrices A, B *commute*.

On the other hand, whenever $n \geqslant 2$, we can always find examples of square matrices $C, E \in \mathbb{F}^{n\times n}$ such that $C \cdot E \neq E \cdot C$.

Important examples/types of matrices

Terminology

A square matrix $D = \left(d_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$d_{i,j} = 0 \quad \text{whenever} \quad i \neq j$$

is called a *diagonal* matrix.

Remark. In general, the (i,i)-th entries of a square matrix A are called its *diagonal* entries, while the remaining entries are the non-diagonal entries. Thus, a diagonal matrix is a square matrix which has **all non-diagonal entries equal to zero**.

Terminology

- A square matrix $U = \left(u_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$u_{i,j} = 0 \quad \text{whenever} \quad i > j$$

is called an *upper triangular* matrix. In other words, an upper triangular matrix is a square matrix which has all entries **below the diagonal** equal to 0.

- Similarly, a square matrix $L = \left(l_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$l_{i,j} = 0 \quad \text{whenever} \quad i < j$$

is called a *lower triangular* matrix. In other words, a lower triangular matrix is a square matrix which has all entries **above the diagonal** equal to 0.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 26

Tuesday October 20

Reminder: Basic Operations on Matrices and Important Properties they satisfy

Reminder: A structure on $\mathbb{F}^{m\times n}$

Theorem

The set $\mathbb{F}^{m\times n}$ together with entry-wise matrix addition and entry-wise scalar multiplication is a vector space over $\mathbb{F}$.

The theorem implies (and relies on checking) the following eight properties:

(i) For all $A, B \in \mathbb{F}^{m\times n}$, $A + B = B + A$.

(ii) For all $A, B, C \in \mathbb{F}^{m\times n}$, $(A + B) + C = A + (B + C)$.

(iii) If we write $\overline{O}$ for the matrix in $\mathbb{F}^{m\times n}$ all of whose entries are equal to $0_{\mathbb{F}}$, then $\overline{O}$ is the neutral element of addition in $\mathbb{F}^{m\times n}$: for all $A \in \mathbb{F}^{m\times n}$, we have that $\overline{O} + A = A = A + \overline{O}$.

(iv) Given an arbitrary matrix $A = \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} \in \mathbb{F}^{m\times n}$, the matrix $\left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$, which is also in $\mathbb{F}^{m\times n}$, is the additive inverse of A: we have that

$$\left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} + \left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} = \left(\left[a_{i,j} + (-a_{i,j})\right]\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} = \overline{O}$$

$$= \left(-a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} + \left(a_{i,j}\right)_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}.$$

Note that the additive inverse of A is also given by $(-1_{\mathbb{F}}) \cdot A$.

A structure on $\mathbb{F}^{m\times n}$ (cont.)

Theorem

The set $\mathbb{F}^{m\times n}$ together with entry-wise matrix addition and entry-wise scalar multiplication is a vector space over $\mathbb{F}$.

Properties hinted at in the statement of the theorem, cont.:

(v) For all $A \in \mathbb{F}^{m\times n}$, $1_{\mathbb{F}} \cdot A = A$.

(vi) For all $A \in \mathbb{F}^{m\times n}$ and all scalars $\lambda, \mu \in \mathbb{F}$, we have that

$$\lambda \cdot (\mu \cdot A) = (\lambda\mu) \cdot A\,.$$

and the two Distributive Laws of a vector space structure:

(vii) For all $A, B \in \mathbb{F}^{m\times n}$ and all $\lambda \in \mathbb{F}$, we have that

$$\lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B\,.$$

(viii) For all $A \in \mathbb{F}^{m\times n}$ and all scalars $\lambda, \mu \in \mathbb{F}$, we have that

$$(\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A\,.$$

A few more important properties

- **Distributive Law** *(matrix multiplication distributes over matrix addition)* For all matrices $A, B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times l}$ and $E \in \mathbb{F}^{k\times m}$, we have that

$$(A+B)C = AC + BC, \qquad E(A+B) = EA + EB.$$

(you have to verify this yourselves in HW3, Problem 3)

- **Associativity** For all matrices $A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times k}$ and $C \in \mathbb{F}^{k\times l}$, we have that

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

(left as an exercise for now, we'll return to this again)

- *Special case of associativity* For all matrices $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$ and all scalars $\lambda \in \mathbb{F}$, we have that

$$\lambda \cdot (A \cdot B) = (\lambda \cdot A) \cdot B = A \cdot (\lambda \cdot B).$$

What about the order in which we write two matrices we want to multiply? Does it matter?

In general, matrix multiplication is not commutative.

That is, there exist (plenty of examples of) matrices A, B such that $A \cdot B \neq B \cdot A$.

First of all, when would both products $A \cdot B$ and $B \cdot A$ be defined?

Answer. Whenever we have $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times m}$ (check this yourselves).

In such a case, as we already have seen, $A \cdot B \in \mathbb{F}^{m \times m}$ and $B \cdot A \in \mathbb{F}^{n \times n}$. Thus,

- if $m \neq n$, we definitely have $A \cdot B \neq B \cdot A$. *(recall when we have equality of matrices)*
- if $m = n$, it may still happen that $A \cdot B \neq B \cdot A$. *(you are asked to find such examples in HW3, Problem 2)*

In general, multiplication in $\mathbb{F}^{n\times n}$ is not commutative

Terminology

A matrix $A \in \mathbb{F}^{n\times n}$ (that is, a matrix which has as many rows as columns) is called a *square* matrix.

When $A, B \in \mathbb{F}^{n\times n}$, it may happen that $A \cdot B = B \cdot A$: e.g. for the matrices

$$A = \begin{pmatrix} -3 & 1 & 1 \\ 0 & -5 & 3 \\ 0 & 1 & -5 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 3 \\ 0 & 1 & 0 \end{pmatrix}$$

in $\mathbb{R}^{3\times 3}$, we have that

$$AB = \begin{pmatrix} -6 & -2 & 0 \\ 0 & 3 & -15 \\ 0 & -5 & 3 \end{pmatrix} = BA\,.$$

Terminology

Given matrices $A, B \in \mathbb{F}^{n\times n}$, if we have that $A \cdot B = B \cdot A$, we say that the matrices A, B *commute*.

On the other hand, whenever $n \geqslant 2$, we can always find examples of square matrices $C, E \in \mathbb{F}^{n\times n}$ such that $C \cdot E \neq E \cdot C$.

Important examples/types of matrices

Terminology

A square matrix $D = \left(d_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$d_{i,j} = 0 \quad \text{whenever} \quad i \neq j$$

is called a *diagonal* matrix.

Remark. In general, the (i,i)-th entries of a square matrix A are called its *diagonal* entries, while the remaining entries are the non-diagonal entries. Thus, a diagonal matrix is a square matrix which has **all non-diagonal entries equal to zero**.

Terminology

- A square matrix $U = \left(u_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$u_{i,j} = 0 \quad \text{whenever} \quad i > j$$

is called an *upper triangular* matrix. In other words, an upper triangular matrix is a square matrix which has all entries **below the diagonal** equal to 0.

- Similarly, a square matrix $L = \left(l_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ which satisfies

$$l_{i,j} = 0 \quad \text{whenever} \quad i < j$$

is called a *lower triangular* matrix. In other words, a lower triangular matrix is a square matrix which has all entries **above the diagonal** equal to 0.

Important examples/types of matrices (cont.)

Useful Observation

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$.

A is diagonal **if and only if**
A is both upper triangular and lower triangular.

For each $n \geqslant 1$, consider the following diagonal matrix in $\mathbb{F}^{n\times n}$:

$$I_n := \begin{pmatrix} 1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 1 & \cdots\cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = \begin{pmatrix} | & | & | & & | \\ \bar{e}_1 & \bar{e}_2 & \bar{e}_3 & \cdots & \bar{e}_n \\ | & | & | & & | \end{pmatrix}.$$

Then I_n acts

- as a *left multiplicative identity* for matrices $B \in \mathbb{F}^{n\times k}$,
- and as a *right multiplicative identity* for matrices $C \in \mathbb{F}^{m\times n}$.

Indeed, for all matrices $B \in \mathbb{F}^{n\times k}$ and $C \in \mathbb{F}^{m\times n}$, we have that

$$I_n \cdot B = B \qquad \text{and} \qquad C \cdot I_n = C.$$

Important examples/types of matrices (cont.)

Useful Observation

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$.

A is diagonal **if and only if**
A is both upper triangular and lower triangular.

For each $n \geqslant 1$, consider the following diagonal matrix in $\mathbb{F}^{n\times n}$:

$$I_n := \begin{pmatrix} 1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 1 & \cdots\cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = \begin{pmatrix} | & | & | & & | \\ \bar{e}_1 & \bar{e}_2 & \bar{e}_3 & \cdots & \bar{e}_n \\ | & | & | & & | \end{pmatrix}.$$

Then I_n acts

- as a *left multiplicative identity* for matrices $B \in \mathbb{F}^{n\times k}$,
- and as a *right multiplicative identity* for matrices $C \in \mathbb{F}^{m\times n}$.

Indeed, for all matrices $B \in \mathbb{F}^{n\times k}$ and $C \in \mathbb{F}^{m\times n}$, we have that

$$I_n \cdot B = B \qquad \text{and} \qquad C \cdot I_n = C.$$

In particular, I_n acts as a multiplicative identity within $\mathbb{F}^{n\times n}$.

Recall: Definition of the notion of 'commutative ring'

Definition. A *commutative ring* $\mathcal{R}$ is a set of elements together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x, y) \in \mathcal{R} \times \mathcal{R} \ \mapsto \ x + y \in \mathcal{R} \\ \textit{multiplication} \qquad & (x, y) \in \mathcal{R} \times \mathcal{R} \ \mapsto \ x \cdot y \in \mathcal{R} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathcal{R}$, $\quad x + y = y + x$ *(commutativity)*

(ii) for all $x, y, z \in \mathcal{R}$, $\quad (x + y) + z = x + (y + z)$ *(associativity)*

(iii) there exists an element 0 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 0 + x = x + 0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathcal{R}$, there exists an element $w = w_x$ such that

$$x + w = w + x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

Definition of the notion of 'commutative ring' (cont.)

(i') for all $x, y \in \mathcal{R}$, $\quad x \cdot y = y \cdot x$ *(commutativity)*

(ii') for all $x, y, z \in \mathcal{R}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathcal{R}$ such that

$$\text{for all } x \in \mathcal{R}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

and finally

(viii) for all $x, y, z \in \mathcal{R}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

The last property, called the Distributive Law, relates the operations of addition and multiplication in $\mathcal{R}$.

Definition of the notion of **'ring'**

Definition. A (not necessarily commutative) *ring* $\mathcal{S}$ is a set of elements together with two operations/functions of the following form:

$$\begin{aligned} \textit{addition} \qquad & (x,y) \in \mathcal{S}\times\mathcal{S} \;\mapsto\; x+y \in \mathcal{S} \\ \textit{multiplication} \qquad & (x,y) \in \mathcal{S}\times\mathcal{S} \;\mapsto\; x\cdot y \in \mathcal{S} \end{aligned}$$

which satisfy the following properties:

(i) for all $x, y \in \mathcal{S}$, $\quad x+y = y+x$ *(commutativity)*

(ii) for all $x, y, z \in \mathcal{S}$, $\quad (x+y)+z = x+(y+z)$ *(associativity)*

(iii) there exists an element 0 in $\mathcal{S}$ such that

$$\text{for all } x \in \mathcal{S}, \quad 0+x = x+0 = x$$

(neutral element of addition, or additive identity)

(iv) for every $x \in \mathcal{S}$, there exists an element $w = w_x$ such that

$$x+w = w+x = 0$$

(negative element, or additive inverse) usually denoted by $-x$

(to be continued...)

Definition of the notion of **'ring'** (cont.)

(ii') for all $x, y, z \in \mathcal{S}$, $\quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$ *(associativity)*

(iii') there exists an element 1 in $\mathcal{S}$ such that

$$\text{for all } x \in \mathcal{S}, \quad 1 \cdot x = x \cdot 1 = x$$

(neutral element of multiplication, or multiplicative identity)

and finally

(vii) for all $x, y, z \in \mathcal{S}$,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$\text{and} \quad x \cdot (y + z) = x \cdot y + x \cdot z$$

Note that now it is absolutely necessary that we require both the **right distributive property** and the **left distributive property**.

Another structure on $\mathbb{F}^{n\times n}$

Theorem

The set $\mathbb{F}^{n\times n}$ together with entry-wise matrix addition and the matrix multiplication we have defined is a (non-commutative) ring.

Question. Are there any (multiplicatively) invertible elements in this structure?

Terminology

A square matrix $A \in \mathbb{F}^{n\times n}$ is called *invertible* if there is a matrix $B \in \mathbb{F}^{n\times n}$ such that

$$A \cdot B = I_n = B \cdot A\,.$$

Moreover, if such a matrix B exists, then it is unique, and we call it the *inverse* of A and denote it by A^{-1}.

Is there some nice ‘method’ by which we can tell if a matrix $A \in \mathbb{F}^{n \times n}$ is invertible?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 27

Wednesday October 21

Recall: Structure on $\mathbb{F}^{n\times n}$

Theorem

The set $\mathbb{F}^{n\times n}$ together with entry-wise matrix addition and the matrix multiplication we have defined is a (non-commutative) ring.

Question. Are there any (multiplicatively) invertible elements in this structure?

Terminology

A square matrix $A \in \mathbb{F}^{n\times n}$ is called *invertible* if there is a matrix $B \in \mathbb{F}^{n\times n}$ such that

$$A \cdot B = I_n = B \cdot A\,.$$

Moreover, if such a matrix B exists, then it is unique, and we call it the *inverse* of A. We denote it by A^{-1}.

Examples of invertible and non-invertible matrices

- The matrix

$$A_1 = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{3\times 3}$$

is invertible because, if

$$B_1 = \begin{pmatrix} 0.5 & -1.5 & -1.5 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{3\times 3},$$

then we have $A_1 B_1 = I_3 = B_1 A_1$.

- Similarly the matrix

$$A_2 = \begin{pmatrix} 0 & 2 & 4 & 4 \\ 2 & 3 & 1 & 1 \\ 4 & 3 & 2 & 4 \\ 0 & 2 & 4 & 0 \end{pmatrix} \in \mathbb{Z}_5^{4\times 4}$$

is invertible *(we will soon see how we could check this).*

Examples of invertible and non-invertible matrices (cont.)

- On the other hand, the matrix

$$A_3 = \begin{pmatrix} 1 & -i & 3 \\ 0 & 0 & 0 \\ 2+i & 14 & 2i \end{pmatrix} \in \mathbb{C}^{3\times 3}$$

is not invertible (why?).

Exercise 1

Let $\mathbb{F}$ be a field, and let E be a matrix in $\mathbb{F}^{n\times n}$ which has at least one zero row. Then E cannot be invertible.

Justification. Suppose that the i_0-th row of the matrix E is zero. Moreover, assume towards a contradiction that we could find a matrix $Q \in \mathbb{F}^{n\times n}$ such that

$$E \cdot Q = I_n = Q \cdot E.$$

This would imply that the (i_0, i_0)-th entry of the matrix $E \cdot Q$ is equal to 1. At the same time, we have that this entry equals

$$\langle \mathrm{Row}_{i_0}(E),\, \mathrm{Col}_{i_0}(Q) \rangle = \langle \bar{0},\, \mathrm{Col}_{i_0}(Q) \rangle = 0.$$

Thus we reach a contradiction: the (i_0, i_0)-th entry of the matrix $E \cdot Q$ must be both equal to 1 and equal to 0. We conclude that the matrix E cannot have an inverse, or, in other words, that it is not invertible.

Examples of invertible and non-invertible matrices (cont.)

- On the other hand, the matrix

$$A_3 = \begin{pmatrix} 1 & -i & 3 \\ 0 & 0 & 0 \\ 2+i & 14 & 2i \end{pmatrix} \in \mathbb{C}^{3\times 3}$$

  is not invertible (why?).

Exercise 1

Let $\mathbb{F}$ be a field, and let E be a matrix in $\mathbb{F}^{n\times n}$ which has at least one zero row. Then E cannot be invertible.

One more similar exercise:

Exercise 2

Let $\mathbb{F}$ be a field, and let Q be a matrix in $\mathbb{F}^{n\times n}$ which has at least one zero **column**. Then Q cannot be invertible.

Examples of invertible and non-invertible matrices (cont.)

- On the other hand, the matrix

$$A_3 = \begin{pmatrix} 1 & -i & 3 \\ 0 & 0 & 0 \\ 2+i & 14 & 2i \end{pmatrix} \in \mathbb{C}^{3\times 3}$$

is not invertible.

We just saw that one immediate justification here is that A_3 has a zero row. However, this is a rather 'weak' criterion, because we can also find non-invertible matrices which have no zero rows or columns (we can even find non-invertible matrices all of whose entries are non-zero).

- The matrix

$$A_4 = \begin{pmatrix} 4 & 5 & 1 \\ 4 & 1 & 3 \\ 1 & 6 & 4 \end{pmatrix} \in \mathbb{Z}_7^{3\times 3}$$

is not invertible *(we will soon see how to verify this).*

Past Homework Problem

Consider the following matrices:

$$A = \begin{pmatrix} 1 & 1 & -4 \\ 11 & -5 & 7 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{pmatrix} -1 & -2 \\ 10 & 9 \\ 8 & -6 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \quad \bar{c} = \begin{pmatrix} 1 & -5 \end{pmatrix} \in \mathbb{R}^{1\times 2},$$

$$D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad E = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 1 & 0 & 3 \end{pmatrix} \in \mathbb{Z}_5^{3\times 3}, \quad \bar{u} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{Z}_5^{3\times 1}.$$

Find all products of any two (different) matrices from above that make sense, as well as the following expressions, again if they are defined: $AB + I_2$, $A(I_3 + B)$, $AB\bar{c}$, $I_3 + (E^2 - E) \cdot (I_3 - E)^{-1}$ (is it possible to simplify any of the expressions you are asked to find, in order to do fewer and easier computations? can you simplify regardless of what the given matrices are?).

Solution

The products $A\bar{c}$, $B\bar{c}$, BD, $\bar{c}B$, $\bar{c}D$, DA, $D\bar{c}$ and $\bar{u}E$ are not defined because the dimensions don't match. Similarly, the products AE, EA, BE, EB, $\bar{c}E$, $E\bar{c}$, DE, ED, $A\bar{u}$, $\bar{u}A$, $B\bar{u}$, $\bar{u}B$, $\bar{c}\bar{u}$, $\bar{u}\bar{c}$, $D\bar{u}$, $\bar{u}D$ are not defined simply because the entries of the matrices we are trying to multiply in each case come from different fields none of which is a subfield of the other.

Thus the only products that are defined are the following:

$$AB = \begin{pmatrix} -23 & 31 \\ -5 & -109 \end{pmatrix}, \quad BA = \begin{pmatrix} -23 & 9 & -10 \\ 109 & -35 & 23 \\ -58 & 38 & -74 \end{pmatrix},$$

$$\bar{c}A = \begin{pmatrix} -54 & 26 & -39 \end{pmatrix}, \quad AD = \begin{pmatrix} -3 & 2 & -32 \\ -33 & -10 & 56 \end{pmatrix},$$

$$DB = \begin{pmatrix} 3 & 6 \\ 20 & 18 \\ 64 & -48 \end{pmatrix}, \quad \text{and} \quad E\bar{u} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}.$$

Moreover, $A(I_3 + B)$ is not defined because $I_3 + B$ is not defined given that B does not have the same number of columns as I_3. Similarly, $AB\bar{c}$ is not defined because AB has 2 columns while $\bar{c}$ has 1 row.

On the other hand,

$$AB + I_2 = \begin{pmatrix} -22 & 31 \\ -5 & -108 \end{pmatrix}.$$

Solution (cont.)

Finally, if we can show that $I_3 - E$ is invertible, then the last expression will make sense, and we will be able to write

$$\begin{aligned} I_3 + (E^2 - E) \cdot (I_3 - E)^{-1} &= I_3 + E(E - I_3) \cdot (I_3 - E)^{-1} \\ &= I_3 + E \cdot \big((-1) \cdot (I_3 - E)\big) \cdot (I_3 - E)^{-1} \\ &= I_3 - E \end{aligned}$$

where we used associativity and the distributive law for matrix multiplication, as well as properties of multiplication of a matrix by a scalar which we have stated in Lectures 24 and 25.

Thus it remains to check whether $I_3 - E$ is invertible. **We will see an efficient method to verify this in the coming lectures.**

The transpose of a matrix

Definition

Let $\mathbb{F}$ be a field, and let $A = \left(a_{i,j}\right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \in \mathbb{F}^{m \times n}$ be a matrix with entries from $\mathbb{F}$ which has m rows and n columns.

The *transpose* of A is defined to be a matrix $C = \left(c_{j,i}\right)_{\substack{1 \leqslant j \leqslant n \\ 1 \leqslant i \leqslant m}} \in \mathbb{F}^{n \times m}$ which satisfies

$$c_{j,i} := a_{i,j}$$

for all $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$.

In other words, the j-th row of the transpose of A coincides with the j-th column of A, while the i-th column of the transpose of A coincides with the i-th row of A.

We denote the transpose of A by A^T.

Definition

A square matrix $A = \left(a_{i,j}\right)_{1 \leqslant i,j \leqslant n} \in \mathbb{F}^{n \times n}$ is called *symmetric* if it satisfies $A = A^T$. In other words, if we have $a_{i,j} = a_{j,i}$ for all $i, j \in \{1, 2, \ldots, n\}$.

Find the transposes of the following matrices

$$A = \begin{pmatrix} 1 & 1 & -4 \\ 11 & -5 & 7 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{pmatrix} -1 & -2 \\ 10 & 9 \\ 8 & -6 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \quad \bar{c} = \begin{pmatrix} 1 & -5 \end{pmatrix} \in \mathbb{R}^{1\times 2},$$

$$D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad E = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 1 & 0 & 3 \end{pmatrix} \in \mathbb{Z}_5^{3\times 3}, \quad \bar{u} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{Z}_5^{3\times 1}.$$

We have that

$$A^T = \begin{pmatrix} 1 & 11 \\ 1 & -5 \\ -4 & 7 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \qquad E^T = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 2 & 3 \end{pmatrix} \in \mathbb{Z}_5^{3\times 3}$$

$$\text{and} \quad \bar{c}^T = \begin{pmatrix} 1 \\ -5 \end{pmatrix} \in \mathbb{R}^{2\times 1}$$

The rest left as exercise.

Back to relating matrices and systems of linear equations

Suppose we have a system $\mathcal{LS}_1$ of m linear equations in n unknowns with coefficients from a field $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Then the matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

is called the *coefficient matrix* of $\mathcal{LS}_1$, while the matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ \ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{pmatrix}$$

is called the *augmented matrix* of $\mathcal{LS}_1$.

Suppose we have a system $\mathcal{LS}_1$ of m linear equations in n unknowns with coefficients from a field $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Then the matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

is called the *coefficient matrix* of $\mathcal{LS}_1$, while the matrix

$$\left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{array}\right)$$

is called the *augmented matrix* of $\mathcal{LS}_1$.

Gaussian elimination for matrices?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 28

Friday October 23

Recall: Structure on $\mathbb{F}^{n\times n}$

Theorem

The set $\mathbb{F}^{n\times n}$ together with entry-wise matrix addition and the matrix multiplication we have defined is a (non-commutative) ring.

Question. Are there any (multiplicatively) invertible elements in this structure?

Terminology

A square matrix $A \in \mathbb{F}^{n\times n}$ is called *invertible* if there is a matrix $B \in \mathbb{F}^{n\times n}$ such that

$$A \cdot B = I_n = B \cdot A\,.$$

Moreover, if such a matrix B exists, then it is unique, and we call it the *inverse* of A. We denote it by A^{-1}.

Why is working with invertible matrices useful?

Recall how we analysed solving one linear equation $ax = b$ in one unknown:

There are exactly three cases to consider:

Case 1: $a = b = 0$. Then, **no matter what element of $\mathbb{F}$ we set x equal to**, the equality will hold true. **In other words, we have as many solutions as the elements of $\mathbb{F}$.**

Case 2: $a = 0$, $b \neq 0$. Then, no matter what element of $\mathbb{F}$ we set x equal to, the LHS will equal 0, while the RHS will be non-zero $\rightsquigarrow$ **absurd!** thus we have no solutions

Case 3: $a \neq 0$. Then, no matter what the exact value of a is, and no matter what b is, we can find a **unique solution**. How? Note that $a \neq 0$ implies that a^{-1} exists. But then, we can multiply both sides of the equation by a^{-1}, and also use the associativity of the multiplication, to obtain that

$$x = (a^{-1}a)x = a^{-1} \cdot (ax) = a^{-1} \cdot b.$$

Note that this tells us that the only value that would make the equality true is the element $a^{-1} \cdot b$ (and also that this value is a solution, since $a \cdot (a^{-1}b) = b$ as we wanted).

Why is working with invertible matrices useful?

Assume now that we have a system of n linear equations in n unknowns, and suppose that by using matrix notation we can write it as follows:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n-1} & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \ddots \\ b_{n-1} \\ b_n \end{pmatrix}$$

or more simply $A\bar{x} = \bar{b}$.

If the matrix A were invertible, and A^{-1} were its inverse, then the above linear system would have a unique solution $\bar{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \ddots \\ \lambda_{n-1} \\ \lambda_n \end{pmatrix}$ given by $\bar{\lambda} = A^{-1} \cdot \bar{b}$.

Indeed, using the associativity of multiplication too, we can write

$$A\bar{x} = \bar{b} \;\Rightarrow\; A^{-1}(A\bar{x}) = A^{-1}\bar{b} \;\Rightarrow\; \bar{x} = (A^{-1}A)\bar{x} = A^{-1}\bar{b},$$

and also clearly $A(A^{-1}\bar{b}) = (AA^{-1})\bar{b} = \bar{b}$.

Back to relating matrices and systems of linear equations

Suppose we have a system $\mathcal{LS}_1$ of m linear equations in n unknowns with coefficients from a field $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Definition. The matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

is called the *coefficient matrix* of $\mathcal{LS}_1$, while the matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ \ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{pmatrix}$$

is called the *augmented matrix* of $\mathcal{LS}_1$.

Suppose we have a system $\mathcal{LS}_1$ of m linear equations in n unknowns with coefficients from a field $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Definition. The matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \end{pmatrix}$$

is called the *coefficient matrix* of $\mathcal{LS}_1$, while the matrix

$$\left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} & b_2 \\ \ddots & \ddots & \ddots\ddots & \ddots & \ddots & \ddots \\ a_{m-1,1} & a_{m-1,2} & \cdots\cdots & a_{m-1,n-1} & a_{m-1,n} & b_{m-1} \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} & b_m \end{array}\right)$$

is called the *augmented matrix* of $\mathcal{LS}_1$.

Gaussian elimination for matrices?

Recall the three ways of manipulating a linear system that we have considered so far, which constitute **the method of Gaussian elimination (or alternatively of row reduction)**:

1. Multiplying (both sides of) a linear equation of the system **by a non-zero constant**.
2. Adding a multiple of one equation of the system, say $Eq(i)$, to another equation, say $Eq(j)$, and replacing $Eq(j)$ by the new equation we get *(note that, here, $Eq(i)$ is kept where it was, unchanged).*
3. Swapping two equations (that is, writing $Eq(i)$ where we had $Eq(j)$ before, and writing $Eq(j)$ where we had $Eq(i)$).

Gaussian elimination for matrices

Completely analogously we could 'manipulate' a matrix *(given that we can view any matrix we consider as the augmented matrix (or the coefficient matrix) of some linear system)*:

1. We can multiply a row of the matrix **by a non-zero constant**.
2. We can add a multiple of one row of the matrix, say Row_i, to another row, say Row_j, and replace Row_j by the new row we get *(note that, here, Row_i is kept where it was, unchanged).*
3. We can swap two rows (that is, start writing Row_i where we had Row_j before, while writing Row_j where we had Row_i).

Additional terminology. We call such operations on a matrix *elementary row operations*.

Row equivalent matrices

Definition

Let A, B be matrices in $\mathbb{F}^{m\times n}$. We say that A and B are *row equivalent* if we can get the matrix B by finitely many applications of Gaussian elimination on the matrix A.

If this holds, we write $A \sim B$.

Example. The matrices

$$A = \begin{pmatrix} 4 & 1 & 0 \\ 5 & 6 & 1 \\ 3 & 2 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 5 \\ 5 & 6 & 1 \end{pmatrix}$$

from $\mathbb{Z}_7^{3\times 3}$ are row equivalent (why?). Because we have

$$A = \begin{pmatrix} 4 & 1 & 0 \\ 5 & 6 & 1 \\ 3 & 2 & 4 \end{pmatrix} \xrightarrow{2R_1 \to R_1'} \begin{pmatrix} 1 & 2 & 0 \\ 5 & 6 & 1 \\ 3 & 2 & 4 \end{pmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & 0 \\ 3 & 2 & 4 \\ 5 & 6 & 1 \end{pmatrix} \xrightarrow{R_2 + R_3 \to R_2'} \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 5 \\ 5 & 6 & 1 \end{pmatrix} = B\,.$$

Row equivalent matrices

Proposition 1

Suppose that $C = \left(A \mid \bar{b}\right)$ and $C' = \left(A' \mid \bar{b}'\right)$ are the **augmented** matrices of two linear systems, which are linear systems in the same set of unknowns, with the same number of linear equations (that is, the systems $A\bar{x} = \bar{b}$ and $A'\bar{x} = \bar{b}'$ respectively).

If $C \sim C'$, then the two linear systems are equivalent, or in other words they have the same set of solutions.

Justification. We have already discussed that applications of Gaussian elimination do not change the set of solutions of a linear system. Recall now that $C \sim C'$ means that we can get the matrix C' by finitely many applications of Gaussian elimination on the matrix C, which in this case is just an alternative way of saying that we can get the linear system $A'\bar{x} = \bar{b}'$ via finitely many applications of Gaussian elimination on the system $A\bar{x} = \bar{b}$.

Row Echelon Form and Reduced Row Echelon Form

Definition 1

Let A be a matrix in $\mathbb{F}^{m\times n}$. We say that A is in *Row Echelon Form* (REF for short) if the following hold:

1. if A has any zero rows, they are below any non-zero rows;
2. the first non-zero entry of every non-zero row is found (strictly) to the right of the first non-zero entry of any previous row.

If these hold, then we call the first non-zero entry of each non-zero row of the matrix A a *pivot* of A. Also, we call any column of A which contains a pivot a *pivot column*.

Definition 2

Suppose that $B \in \mathbb{F}^{m\times n}$ is a matrix in REF. We say that B is in *Reduced Row Echelon Form* (RREF for short) if the following also hold:

3. every pivot of B is equal to 1 (that is, the first non-zero entry of each non-zero row of B is equal to 1);
4. each pivot column of B contains **exactly one** non-zero entry (that is, its only non-zero entry is the pivot of B found in that column).

Examples and non-examples

Question. Are any of the following matrices in REF? Are any of them in RREF?

$$A_1 = \begin{pmatrix} 1 & 1 & -4 & 0 \\ 0 & 1 & 7 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 4}, \qquad A_2 = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{Z}_{11}^{3\times 3},$$

$$A_3 = \begin{pmatrix} 2 & -4 & 1 & -0.5 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0.6 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{Q}^{4\times 4}, \qquad A_4 = \begin{pmatrix} 1 & -2 & 0.5 & -0.25 \\ 0 & 0 & 1 & 2/3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

$$A_5 = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4}.$$

Two Very Important Theorems

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

1. Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

2. Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Two Very Important Theorems (cont.)

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n\times n}$. The following are equivalent:

- A is invertible.
- Every REF of A has exactly n pivots.
- At least one REF of A has exactly n pivots.
- The unique RREF of A is the identity matrix I_n.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 29

Monday October 26

Recall: Row Echelon Form and Reduced Row Echelon Form

Definition 1

Let A be a matrix in $\mathbb{F}^{m \times n}$. We say that A is in *Row Echelon Form* (REF for short) if the following hold:

1. if A has any zero rows, they are below any non-zero rows;
2. the first non-zero entry of every non-zero row is found (strictly) to the right of the first non-zero entry of any previous row.

If these hold, then we call the first non-zero entry of each non-zero row of the matrix A a *pivot* of A. Also, we call any column of A which contains a pivot a *pivot column*.

Definition 2

Suppose that $B \in \mathbb{F}^{m \times n}$ is a matrix in REF. We say that B is in *Reduced Row Echelon Form* (RREF for short) if the following also hold:

3. every pivot of B is equal to 1 (that is, the first non-zero entry of each non-zero row of B is equal to 1);
4. each pivot column of B contains **exactly one** non-zero entry (that is, its only non-zero entry is the pivot of B found in that column).

Examples and non-examples

Question. Are any of the following matrices in REF? Are any of them in RREF?

$$A_1 = \begin{pmatrix} 1 & 1 & -4 & 0 \\ 0 & 1 & 7 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 4}, \qquad A_2 = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{Z}_{11}^{3\times 3},$$

$$A_3 = \begin{pmatrix} 2 & -4 & 1 & -0.5 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0.6 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{Q}^{4\times 4}, \qquad A_4 = \begin{pmatrix} 1 & -2 & 0.5 & -0.25 \\ 0 & 0 & 1 & 2/3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

$$A_5 = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4}.$$

Examples and non-examples

Answer:

$$A_1 = \begin{pmatrix} \boxed{1} & 1 & -4 & 0 \\ 0 & \boxed{1} & 7 & 8 \\ 0 & 0 & 0 & \boxed{1} \end{pmatrix} \in \mathbb{R}^{3\times 4},$$

in REF, not in RREF; pivots highlighted

$$A_2 = \begin{pmatrix} \boxed{-4} & 0 & 0 \\ 0 & \boxed{2} & 0 \\ 0 & 0 & \boxed{8} \end{pmatrix} \in \mathbb{Z}_{11}^{3\times 3},$$

in REF, not in RREF; pivots highlighted

$$A_3 = \begin{pmatrix} 2 & -4 & 1 & -0.5 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0.6 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

upper triangular, but **not** in REF

$$A_4 = \begin{pmatrix} \boxed{1} & -2 & 0.5 & -0.25 \\ 0 & 0 & \boxed{1} & 2/3 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

in REF, not in RREF; pivots highlighted

$$A_5 = \begin{pmatrix} \boxed{1} & -2 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4}$$

in RREF; pivots highlighted

Note also that $A_3 \sim A_4 \sim A_5$, and thus A_4 and A_5 are REFs of A_3, while A_5 is the (unique) RREF of A_3 and A_4.

Reminder: Two Very Important Theorems

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Two Very Important Theorems (cont.)

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n\times n}$. The following are equivalent:

- A is invertible.
- Every REF of A has exactly n pivots.
- At least one REF of A has exactly n pivots.
- The unique RREF of A is the identity matrix I_n.

Applying Theorem 2

Recall that we said that the matrix

$$A = \begin{pmatrix} 4 & 5 & 1 \\ 4 & 1 & 3 \\ 1 & 6 & 4 \end{pmatrix} \in \mathbb{Z}_7^{3\times 3}$$

is not invertible. Let's verify this: by Theorem 2, if A is not invertible, then there exists at least one REF of A which does not have 3 pivots (note that here $n = 3$); this is because $(I) \Leftrightarrow (II)$ is equivalent to $\neg(I) \Leftrightarrow \neg(II)$, while the statement in red is exactly $\neg(II)$.

Moreover, both negated statements are equivalent to $\neg(III)$, that is, the statement that every REF of A will have fewer than 3 pivots. This shows that we don't need to be careful about which REF of A to consider when trying to confirm $\neg(II)$.

Looking now for one REF of A: we have that

$$A = \begin{pmatrix} 4 & 5 & 1 \\ 4 & 1 & 3 \\ 1 & 6 & 4 \end{pmatrix} \underbrace{\begin{matrix} R_2 - R_1 \to R_2' \\ R_3 - 2R_1 \to R_3' \end{matrix}} \begin{pmatrix} 4 & 5 & 1 \\ 0 & 3 & 2 \\ 0 & 3 & 2 \end{pmatrix}$$

$$\underbrace{R_3 - R_2 \to R_3'} \begin{pmatrix} 4 & 5 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

The last matrix is in REF, and has only 2 pivots. Therefore, A is not invertible.

Applying Theorem 2: one more example

Recall the past homework problem we discussed in Lecture 27:
Consider the following matrices:

$$A = \begin{pmatrix} 1 & 1 & -4 \\ 11 & -5 & 7 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{pmatrix} -1 & -2 \\ 10 & 9 \\ 8 & -6 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \quad \bar{c} = \begin{pmatrix} 1 & -5 \end{pmatrix} \in \mathbb{R}^{1\times 2},$$

$$D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad E = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 1 & 0 & 3 \end{pmatrix} \in \mathbb{Z}_5^{3\times 3}, \quad \bar{u} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{Z}_5^{3\times 1}.$$

Find all products of any two (different) matrices from above that make sense, as well as the following expressions, again if they are defined: $AB + I_2$, $A(I_3 + B)$, $AB\bar{c}$, $I_3 + (E^2 - E) \cdot (I_3 - E)^{-1}$.

Recall also that we gave a partial solution to this problem; the only thing that remained to check was whether the matrix $I_3 - E$ is invertible (in which case we said that the expression $I_3 + (E^2 - E) \cdot (I_3 - E)^{-1}$ would make sense, and it would be possible to simplify it to $I_3 - E$; see Lecture 27 for this).

To check whether $I_3 - E$ is invertible, we look for a REF of it:

$$I_3 - E = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & -2 \\ -1 & 0 & -2 \end{pmatrix} \sim \begin{pmatrix} -1 & 0 & -2 \\ 0 & -1 & -2 \\ 0 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} -1 & 0 & -2 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{pmatrix}.$$

We conclude that there is a REF of $I_3 - E$ with 3 pivots, and thus $I_3 - E$ is invertible.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 30

Tuesday October 27

Recall: Row Echelon Form and Reduced Row Echelon Form

Definition 1

Let A be a matrix in $\mathbb{F}^{m \times n}$. We say that A is in *Row Echelon Form* (REF for short) if the following hold:

1. if A has any zero rows, they are below any non-zero rows;
2. the first non-zero entry of every non-zero row is found (strictly) to the right of the first non-zero entry of any previous row.

If these hold, then we call the first non-zero entry of each non-zero row of the matrix A a *pivot* of A. Also, we call any column of A which contains a pivot a *pivot column*.

Definition 2

Suppose that $B \in \mathbb{F}^{m \times n}$ is a matrix in REF. We say that B is in *Reduced Row Echelon Form* (RREF for short) if the following also hold:

3. every pivot of B is equal to 1 (that is, the first non-zero entry of each non-zero row of B is equal to 1);
4. each pivot column of B contains **exactly one** non-zero entry (that is, its only non-zero entry is the pivot of B found in that column).

Examples and non-examples

Question. Are any of the following matrices in REF? Are any of them in RREF?

$$A_1 = \begin{pmatrix} \boxed{1} & 1 & -4 & 0 \\ 0 & \boxed{1} & 7 & 8 \\ 0 & 0 & 0 & \boxed{1} \end{pmatrix} \in \mathbb{R}^{3\times 4},$$

in REF, not in RREF; pivots highlighted

$$A_2 = \begin{pmatrix} \boxed{-4} & 0 & 0 \\ 0 & \boxed{2} & 0 \\ 0 & 0 & \boxed{8} \end{pmatrix} \in \mathbb{Z}_{11}^{3\times 3},$$

in REF, not in RREF; pivots highlighted

$$A_3 = \begin{pmatrix} 2 & -4 & 1 & -0.5 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0.6 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

upper triangular, but **not** in REF

$$A_4 = \begin{pmatrix} \boxed{1} & -2 & 0.5 & -0.25 \\ 0 & 0 & \boxed{1} & 2/3 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4},$$

in REF, not in RREF; pivots highlighted

$$A_5 = \begin{pmatrix} \boxed{1} & -2 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Q}^{4\times 4}$$

in RREF; pivots highlighted

Note also that $A_3 \sim A_4 \sim A_5$, and thus A_4 and A_5 are REFs of A_3, while A_5 is the (unique) RREF of A_3 and A_4.

Reminder: Row Echelon Form of a Matrix

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Corollary of Theorem 1a

Combining the following

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

Proposition 1

Suppose that $C = \left(A \mid \bar{b}\right)$ and $C' = \left(A' \mid \bar{b}'\right)$ are the **augmented** matrices of two linear systems, which are linear systems in the same set of unknowns, with the same number of linear equations (that is, the systems $A\bar{x} = \bar{b}$ and $A'\bar{x} = \bar{b}'$ respectively).

If $C \sim C'$, then the two linear systems are equivalent, or in other words they have the same set of solutions.

we obtain

Corollary

Let $\mathcal{LS}_1$ be a system of $\widetilde{m}$ linear equations in $\widetilde{n}$ unknowns $x_1, x_2, \ldots, x_{\widetilde{n}}$. Then we can replace $\mathcal{LS}_1$ by an **equivalent** **staircase** system $\mathcal{LS}_2$ (with $\widetilde{m}$ linear equations again, in the same set of unknowns) **via finitely many applications of the method of Gaussian elimination.**

Proving Theorem 1

We need to rely on the method of mathematical induction.
In fact, the proof of Theorem 1a will proceed via induction in the number m of rows of the matrix.

Method of Mathematical Induction

Suppose we have a mathematical statement involving a parameter n that takes values in the positive integers.

Suppose also that we want to show that this statement $P(n)$ holds true for every such value of n.

A proof via mathematical induction requires the following two steps:

(I) Proving that $P(1)$ holds true. *This is called the "Base Case" or "Base of the Induction".*

Remark. *Sometimes $P(n)$ is not true, due to a simple reason perhaps, for some initial values of n, in which case we can choose the Base of the Induction to be proving $P(k_0)$ for some $k_0 > 1$.* **This will then allow us to proceed to show that $P(n)$ is true for every $n \geqslant k_0$.**

[If $P(n')$ is true also for some $n' < k_0$, we have to check these separately; mathematical induction will only handle the parameter range $n \geqslant k_0$ where k_0 corresponds to the base case.]

Method of Mathematical Induction (cont.)

A proof via mathematical induction requires the following two steps:

(I) Proving that $P(1)$ holds true. *This is called the "Base Case" or "Base of the Induction".*

(II) Proving that

- Knowing $P(1)$ is true implies $P(2)$ is true
- Knowing $P(2)$ is true implies $P(3)$ is true

$$\vdots \quad \vdots$$

- Knowing $P(10)$ is true implies $P(11)$ is true . . . . . .

In short, we have to show, for an arbitrary $n \geqslant 1$ *(or $n \geqslant k_0$, if the base case is different)*, that

- if we assume $P(n)$ holds true *(this is called the "Inductive Hypothesis")*
- then $P(n+1)$ holds true as well.

Showing "$P(n) \Rightarrow P(n+1)$" is called the "Induction Step".

Method of Mathematical Induction: when to use the method

Recall that it makes sense to consider using the method when:
— we have a mathematical statement involving a parameter n that takes values in the positive integers,

— and we want to show that this statement $P(n)$ holds true for every such value of n (or for every value of $n \geqslant k_0$).

We then use mathematical induction when:

- we cannot show $P(n)$ for an arbitrary n directly (or it's too complicated / requires advanced tools),
- but we can see how to show that $P(n)$ implies $P(n+1)$.

Method of Mathematical Induction: some standard examples

Example 1: Binomial Theorem. For every $x, y \in \mathbb{R}$, and for every integer $n \geqslant 2$, we have

$$(x+y)^n = \sum_{k=0}^{n} \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k} x^{n-k}y^k$$

$$= x^n + nx^{n-1}y + \frac{n(n-1)}{2}x^{n-2}y^2 + \cdots\cdots + nxy^{n-1} + y^n\,.$$

Method of Mathematical Induction: some standard examples

Example 2: Bernoulli's Inequality. For every integer $n \geqslant 0$, and for every $x \in \mathbb{R}$, $x \geqslant -2$,

$$(1+x)^n \geqslant 1 + nx\,.$$

Method of Mathematical Induction: some standard examples

Example 3. For every (???) positive integer n

$$2^n > n^3.$$

Method of Mathematical Induction in the case of linear systems/matrices?

Recall the following linear system with coefficients from $\mathbb{Z}_5$:

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\}$$

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 31

Wednesday October 28

Recall: Method of Mathematical Induction

Suppose we have a mathematical statement involving a parameter n that takes values in the positive integers.

Suppose also that we want to show that this statement $P(n)$ holds true for every such value of n.

A proof via mathematical induction requires the following two steps:

(I) Proving that $P(1)$ holds true. *This is called the "Base Case" or "Base of the Induction".*

Remark. *Sometimes $P(n)$ is not true, due to a simple reason perhaps, for some initial values of n, in which case we can choose the Base of the Induction to be proving $P(k_0)$ for some $k_0 > 1$.* **This will then allow us to proceed to show that $P(n)$ is true for every $n \geqslant k_0$.**

[If $P(n')$ is true also for some $n' < k_0$, we have to check these separately; mathematical induction will only handle the parameter range $n \geqslant k_0$ where k_0 corresponds to the base case.]

Method of Mathematical Induction (cont.)

A proof via mathematical induction requires the following two steps:

(I) Proving that $P(1)$ holds true. *This is called the "Base Case" or "Base of the Induction".*

(II) Proving that

- Knowing $P(1)$ is true implies $P(2)$ is true
- Knowing $P(2)$ is true implies $P(3)$ is true

$$\vdots \quad \vdots$$

- Knowing $P(10)$ is true implies $P(11)$ is true . . . . . .

In short, we have to show, for an arbitrary $n \geqslant 1$ *(or $n \geqslant k_0$, if the base case is different)*, that

- if we assume $P(n)$ holds true *(this is called the "Inductive Hypothesis")*
- then $P(n+1)$ holds true as well.

Showing "$P(n) \Rightarrow P(n+1)$" is called the "Induction Step".

Method of Mathematical Induction: when to use the method

Recall that it makes sense to consider using the method when:
— we have a mathematical statement involving a parameter n that takes values in the positive integers,

— and we want to show that this statement $P(n)$ holds true for every such value of n (or for every value of $n \geqslant k_0$).

We then use mathematical induction when:

- we cannot show $P(n)$ for an arbitrary n directly (or it's too complicated / requires advanced tools),
- but we can see how to show that $P(n)$ implies $P(n+1)$.

Method of Mathematical Induction: some standard examples

Example 2: Bernoulli's Inequality. For every integer $n \geqslant 0$, and for every $x \in \mathbb{R}$, $x \geqslant -2$,

$$(1+x)^n \geqslant 1 + nx\,.$$

Method of Mathematical Induction: some standard examples

Example 2: Bernoulli's Inequality. For every integer $n \geqslant 0$, and for every $x \in \mathbb{R}$, $x \geqslant -2$,

$$(1+x)^n \geqslant 1+nx\,.$$

Proof. Via mathematical induction in n; in fact, we will set $n = 2k$ or $n = 2k+1$, and we will do mathematical induction in k.
Fix some $x \geqslant -2$.
Base of Induction: $k = 0$, and hence $n = 2k = 0$ or $n = 2k+1 = 1$.
— For $n = 0$, we have $(1+x)^0 = 1 = 1+0x$.
— For $n = 1$, we have $(1+x)^1 = 1+x = 1+1x$.

Induction Step: Assume that, for some $k \geqslant 0$,

$$\text{we have that } (1+x)^{2k} \geqslant 1+(2k)x \quad \text{and} \quad (1+x)^{2k+1} \geqslant 1+(2k+1)x.$$

(In green we have the Induction Hypothesis.)

We need to show that we also have

$$(1+x)^{2(k+1)} \geqslant 1+2(k+1)x \quad \text{and} \quad (1+x)^{2(k+1)+1} \geqslant 1+(2(k+1)+1)x.$$

The rest of the induction step left as an exercise.

Method of Mathematical Induction: some standard examples

Example 3. For every (???) positive integer n

$$2^n > n^3.$$

Method of Mathematical Induction: some standard examples

Example 3. For every $n \geqslant 10$

$$2^n > n^3.$$

Base of Induction: $n = 10$. We have that $2^{10} = 1024$, while $10^3 = 1000$. Therefore, $2^{10} > 10^3$.

Induction Step: Assume that, for some $n \geqslant 10$, we have that

$$2^n > n^3.$$

(In green we have the Induction Hypothesis.)

Then, we can write

$$\begin{aligned}
(n+1)^3 &= n^3 + 3n^2 + 3n + 1 \\
&< n^3 + 3n^2 + 3n^2 + n^2 && \text{because } n \geqslant 10 \Rightarrow 1 < n^2 \text{ and } n < n^2 \\
&= n^3 + 7n^2 \\
&< n^3 + n^3 && \text{because } n \geqslant 10 \Rightarrow 7n^2 < n^3 \\
&< 2^n + 2^n = 2^{n+1} && \text{by the Induction Hypothesis}
\end{aligned}$$

This completes the proof of the Induction Step, and hence the proof of the given statement.

Method of Mathematical Induction: some standard examples

An example similar to Example 3. Show that, for every $n \geqslant 2$

$$2^{n-1} > \sqrt{n}.$$

Method of Mathematical Induction: an incorrect example!

Example 0. Show that, for every $n \geqslant 1$ we have that, if $(a_1, a_2, \ldots, a_n)$ is a sequence of n real numbers, then the sequence is constant (that is, all its terms are equal).
Note that this is an obviously false claim.

An attempt at a proof. Base of Induction: $n = 1$. Then any given sequence (a_1) of 1 real number has only one term, so the claim is (trivially) true for this sequence.

Induction Step: Assume that, for some $n > 1$, we have that

if $(b_1, b_2, \ldots, b_n)$ is a sequence of n real numbers, then the sequence is constant.

(In green we have the Induction Hypothesis.)

We need to show that

if $(a_1, a_2, \ldots, a_n, a_{n+1})$ is a sequence of $n+1$ real numbers, then the sequence is constant.

But if we consider such a sequence $(a_1, a_2, \ldots, a_n, a_{n+1})$, then the subsequences $(a_1, a_2, \ldots, a_n)$ and $(a_2, \ldots, a_n, a_{n+1})$ are sequences of n real numbers, and thus, by the Induction Hypothesis, they are constant sequences. We thus have $a_i = a_2$ for every $i \in \{1, 2, 3 \ldots, n\}$, but also $a_j = a_2$ for every $j \in \{2, 3, \ldots, n, n+1\}$. In other words, we have $a_s = a_2$ for every $s \in \{1, 2, 3, \ldots, n, n+1\}$.

We have thus completed the proof of both the Base of the Induction and the Induction Step, and hence we have proven (???) the claim.

Method of Mathematical Induction: an incorrect example!

Example 0. Show that, for every $n \geqslant 1$ we have that, if $(a_1, a_2, \ldots, a_n)$ is a sequence of n real numbers, then the sequence is constant (that is, all its terms are equal). **Note that this is an obviously false claim.**

An attempt at a proof. Base of Induction: $n = 1$. Then any given sequence (a_1) of 1 real number has only one term, so the claim is (trivially) true for this sequence.

Induction Step: Assume that, for some $n > 1$, we have that

if $(b_1, b_2, \ldots, b_n)$ is a sequence of n real numbers, then the sequence is constant.

(In green we have the Induction Hypothesis.)

This is an **incorrectly stated** Induction Hypothesis, because we only made it for $n > 1$, and thus our proof scheme is now supposed to give

- Knowing $P(2)$ is true implies $P(3)$ is true
- Knowing $P(3)$ is true implies $P(4)$ is true

⋮ ⋮

- Knowing $P(10)$ is true implies $P(11)$ is true . . . . .

However, we are **not guaranteed to know** whether any of the statements/premises here is true, since none of them is the same as the Base of the Induction.

Theorem 1 from last lectures

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Proving Theorem 1

We need to rely on the method of mathematical induction.
In fact, the proof of Theorem 1a will proceed via induction in the number m of rows of the matrix.

Method of Mathematical Induction in the case of linear systems/matrices?

Recall the following linear system with coefficients from $\mathbb{Z}_5$:

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\}$$

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 32

Friday October 30

Method of Mathematical Induction in the case of linear systems/matrices?

Recall the following linear system with coefficients from $\mathbb{Z}_5$:

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\}$$

We want to prove the following mathematical statement (and demonstrate on this specific example):

$P(m)$: for every $n \geqslant 1$ and every linear system $\mathcal{LS}_1$ in n unknowns with coefficients from $\mathbb{Z}_5$ which has m equations, we can find an equivalent linear system $\mathcal{LS}_2$ which is staircase via finitely many applications of Gaussian elimination.

Method of Mathematical Induction in the case of linear systems/matrices?

$P(m)$: for every $n \geqslant 1$ and every linear system $\mathcal{LS}_1$ in n unknowns with coefficients from $\mathbb{Z}_5$ which has m equations, we can find an equivalent linear system $\mathcal{LS}_2$ which is staircase via finitely many applications of Gaussian elimination.

Base of Induction: $m = 1$. One linear system we could consider here is:

$$2x_2 + 0x_3 + x_4 = 3.$$

This is already a staircase system!

Induction Step: We need to show that $P(1) \Rightarrow P(2)$: one instance in which we need to verify this implication is the following linear system:

$$\left\{ \begin{array}{rcrcrcl} x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_2 & + & 0x_3 & + & 2x_4 & = & 1 \end{array} \right\} \quad \overset{E_2 - 2E_1 \to E_2'}{\Longleftrightarrow}$$

$$\left\{ \begin{array}{rcrcrcl} x_2 & + & 4x_3 & - & x_4 & = & 1 \\ & & 2x_3 & + & 4x_4 & = & 4 \end{array} \right\}.$$

The definition of 'staircase system' is now satisfied with respect to the first pivot of the system; the remaining pivots will be found in equations below the 1st equation, so we can look at a subsystem of the last system which has $m - 1 = 1$ linear equation in this case $\rightsquigarrow$ our current Inductive Hypothesis applies to this subsystem!

Method of Mathematical Induction in the case of linear systems/matrices?

We also need to show that $P(2) \Rightarrow P(3)$:

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xLeftrightarrow{E_1 \leftrightarrow E_2}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xLeftrightarrow{E_3+3E_1 \to E_3'}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ & & 2x_2 & & & + & 2x_4 & = & 1 \end{array}\right\}.$$

Method of Mathematical Induction in the case of linear systems/matrices?

We also need to show that $P(2) \Rightarrow P(3)$:

$$\left\{\begin{array}{rcrcrcrcl} & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xleftrightarrow{E_1 \leftrightarrow E_2}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ -x_1 & + & 3x_2 & - & 3x_3 & + & 4x_4 & = & 3 \end{array}\right\} \xleftrightarrow{E_3+3E_1 \to E_3'}$$

$$\left\{\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & = & 1 \\ & & x_2 & + & 4x_3 & - & x_4 & = & 1 \\ & & 2x_2 & & & + & 2x_4 & = & 1 \end{array}\right\}.$$

The definition of 'staircase system' is now satisfied with respect to the first pivot of the system; the remaining pivots will be found in equations below the 1st equation, so we can look at a subsystem of the last system which has $m - 1 = 2$ linear equations in this case $\rightsquigarrow$ our current Inductive Hypothesis applies to this subsystem!

Note also that, by replacing the highlighted subsystem by an equivalent staircase system with 2 linear equations (in the unknowns x_2, x_3, x_4), and then writing down the new system in the highlighted space, will give us a staircase system with 3 linear equations which is equivalent to the system we started with (why?).

Let us now apply this scheme to give a rigorous proof of Theorem 1 from last times.

Theorem 1 from last lectures

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Proof of Theorem 1a

We fix the number n of columns of the matrix, and we proceed by induction in the number m of rows of the matrix.

Base of Induction: $m = 1$. Then the matrix A looks like this

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n} \end{pmatrix},$$

and is already in REF.

Induction Step: Assume that, for some $m \geqslant 1$,
we have that every matrix $\widetilde{A} \in \mathbb{F}^{m \times n}$ is row equivalent
to a matrix $\widetilde{B} \in \mathbb{F}^{m \times n}$ which is in REF.

(In green we have the Induction Hypothesis.)

Consider a matrix $A \in \mathbb{F}^{(m+1) \times n}$:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots\cdots & a_{m,n-1} & a_{m,n} \\ a_{m+1,1} & a_{m+1,2} & \cdots\cdots & a_{m+1,n-1} & a_{m+1,n} \end{pmatrix}.$$

We look for the first column of A which has a non-zero entry; let's say that this is column j_0. We then look for the first non-zero entry of this column (moving from the top); let's say that this is in row i_0.

This tells us that the entry a_{i_0,j_0} of A will be the first pivot of A (which we now must move to the first row). In other words...

Proof of Theorem 1a (cont.)

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1,j_0+1} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & 0 & \cdots & 0 & a_{2,j_0+1} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{m,j_0} & a_{m,j_0+1} & \cdots & a_{m,n-1} & a_{m,n} \\ 0 & 0 & \cdots & a_{m+1,j_0} & a_{m+1,j_0+1} & \cdots & a_{m+1,n-1} & a_{m+1,n} \end{pmatrix} \underbrace{\qquad R_1 \leftrightarrow R_{i_0}}$$

$$\begin{pmatrix} 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ 0 & 0 & \cdots & 0 & a_{2,j_0+1} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & a_{1,j_0+1} & \cdots & a_{1,n-1} & a_{1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{m,j_0} & a_{m,j_0+1} & \cdots & a_{m,n-1} & a_{m,n} \\ 0 & 0 & \cdots & a_{m+1,j_0} & a_{m+1,j_0+1} & \cdots & a_{m+1,n-1} & a_{m+1,n} \end{pmatrix} \underbrace{\qquad \forall i \neq i_0, \;\; R'_i - \frac{a_{i,j_0}}{a_{i_0,j_0}} R'_{i_0} \to R''_i}$$

$$\begin{pmatrix} 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{2,j_0+1} & \cdots & \widetilde{a}_{2,n-1} & \widetilde{a}_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{1,j_0+1} & \cdots & \widetilde{a}_{1,n-1} & \widetilde{a}_{1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{m,j_0+1} & \cdots & \widetilde{a}_{m,n-1} & \widetilde{a}_{m,n} \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{m+1,j_0+1} & \cdots & \widetilde{a}_{m+1,n-1} & \widetilde{a}_{m+1,n} \end{pmatrix}.$$

Proof of Theorem 1a (cont.)

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1,j_0+1} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & 0 & \cdots & 0 & a_{2,j_0+1} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{m,j_0} & a_{m,j_0+1} & \cdots & a_{m,n-1} & a_{m,n} \\ 0 & 0 & \cdots & a_{m+1,j_0} & a_{m+1,j_0+1} & \cdots & a_{m+1,n-1} & a_{m+1,n} \end{pmatrix} \underbrace{\qquad}_{R_1 \leftrightarrow R_{i_0}}$$

$$\begin{pmatrix} 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ 0 & 0 & \cdots & 0 & a_{2,j_0+1} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & a_{1,j_0+1} & \cdots & a_{1,n-1} & a_{1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{m,j_0} & a_{m,j_0+1} & \cdots & a_{m,n-1} & a_{m,n} \\ 0 & 0 & \cdots & a_{m+1,j_0} & a_{m+1,j_0+1} & \cdots & a_{m+1,n-1} & a_{m+1,n} \end{pmatrix} \underbrace{\qquad}_{\forall i \neq i_0,\ R'_i - \frac{a_{i,j_0}}{a_{i_0,j_0}} R'_{i_0} \to R''_i}$$

$$\begin{pmatrix} 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{2,j_0+1} & \cdots & \widetilde{a}_{2,n-1} & \widetilde{a}_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{1,j_0+1} & \cdots & \widetilde{a}_{1,n-1} & \widetilde{a}_{1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{m,j_0+1} & \cdots & \widetilde{a}_{m,n-1} & \widetilde{a}_{m,n} \\ 0 & 0 & \cdots & 0 & \widetilde{a}_{m+1,j_0+1} & \cdots & \widetilde{a}_{m+1,n-1} & \widetilde{a}_{m+1,n} \end{pmatrix}.$$

Proof of Theorem 1a (cont.)

Finally the highlighted submatrix $\widetilde{A}$ is in $\mathbb{F}^{m\times n}$, so we can apply the Inductive Hypothesis to it and find a matrix $\widetilde{B} \in \mathbb{F}^{m\times n}$ which is row equivalent to $\widetilde{A}$ and is in REF.

Note finally that, because the first j_0 columns of $\widetilde{A}$ are zero columns, every matrix which is row equivalent to $\widetilde{A}$ will also have this property (why? check that elementary row operations cannot transform a zero column into a non-zero column!).

Thus the entire matrix

$$\begin{pmatrix} 0 & 0 & \cdots & a_{i_0,j_0} & a_{i_0,j_0+1} & \cdots & a_{i_0,n-1} & a_{i_0,n} \\ & & & & \widetilde{B} & & & \end{pmatrix}$$

will be in Row Echelon Form, and will also be row equivalent to the original matrix A.

This completes the proof of the Induction Step, and hence also the proof of Theorem 1a.

Theorem 1 again

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Proof of Theorem 1b?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 33

Monday November 2

Theorem 1 again

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

A remark about the proof of Theorem 1a

Note that, in the induction step, we start by writing the Inductive Hypothesis:

given $m \geqslant 1$, we have that every matrix $\widetilde{A} \in \mathbb{F}^{m \times n}$
is row equivalent to a matrix $\widetilde{B} \in \mathbb{F}^{m \times n}$ which is in REF

and then we consider a matrix $A \in \mathbb{F}^{(m+1) \times n}$.

Then, to be able to make use of the Inductive Hypothesis, we are essentially trying to find the first pivot of a REF of the matrix A, move it to the first row, and then also replace all entries below the first pivot by zero entries (via elementary row operations of course).

How do we find this first pivot of A? It will be the first non-zero entry (from the top) in the first non-zero column of A; in other words, the first non-zero column of A will be its first pivot column.

But what if we can't find such a non-zero column of A? Then $A = \overline{\boldsymbol{O}}$, which is in REF (and RREF) already *(that is, in this special case we don't even need to make use of the Inductive Hypothesis).*

Proving Theorem 1b?

In this lecture we will deal with the first part of the statement:

Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF).

Let us begin with some remarks again:

- By Theorem 1a, we know that, given a matrix $A \in \mathbb{F}^{m \times n}$, we can find a Row Echelon Form $B \in \mathbb{F}^{m \times n}$ of A. **Thus, it suffices to show that this matrix B in REF has a Reduced Row Echelon Form $C \in \mathbb{F}^{m \times n}$,** since knowing that
$$A \sim B \qquad \text{and} \qquad B \sim C$$
gives us that $A \sim C$ as well. In other words, C will be a RREF of A as well.

- We will prove the statement in green via induction in the number of pivots of B (or equivalently in the number of pivot columns of B). In fact, we will prove the stronger statement:

Theorem 1b′: Every matrix $B \in \mathbb{F}^{m \times n}$ in REF
is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF,
and also has exactly as many pivots as B
and in the same positions as B.

Proof of Theorem 1b′

We prove that

Every matrix $B \in \mathbb{F}^{m\times n}$ in REF is row equivalent
to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF,
and also has exactly as many pivots as B
and in the same positions as B

using induction in the number s of pivots of B (note that $s \leqslant \min\{m, n\}$).

Base of Induction: $s = 0$ or $s = 1$. **Note** that if $s = 0$, then B has no non-zero rows, and hence $B = \overline{\boldsymbol{O}}$, which is a matrix in RREF already.

On the other hand, if $s = 1$, then B has exactly one pivot. Since B is a matrix in REF, its only pivot must be in the 1st row, while all the other rows of B must be zero; let's say that the only pivot of B is the entry a_{1,j_0}:

$$B = \begin{pmatrix} 0 & \cdots & 0 & a_{1,j_0} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

Then, by multiplying the 1st row of B by $(a_{1,j_0})^{-1}$, we get a new matrix C which satisfies the following:
– C is row equivalent to B.
– C has only one non-zero row, its 1st row. This also shows that C is in REF and has only one pivot.
– The only pivot of C is in position $(1, j_0)$ and is equal to 1.
Combining the above, we see that C is in RREF and has the desired properties.

Proof of Theorem 1b′

Induction Step: Assume that, for some $s \geqslant 1$, we have that

given $m \geqslant s$ and $n \geqslant s$, and a matrix $\widetilde{B} \in \mathbb{F}^{m \times n}$ in REF which has s pivots, we can find a matrix $\widetilde{C} \in \mathbb{F}^{m \times n}$ which is row equivalent to $\widetilde{B}$ and in RREF, **and has s pivots too, in the same positions as $\widetilde{B}$ does.**

We now consider $m \geqslant s+1$, $n \geqslant s+1$, and a matrix $B \in \mathbb{F}^{m \times n}$ in REF which has $s+1$ pivots; let's suppose that its last pivot column is the j_0-th column (note that necessarily the last pivot of B, that is, its $(s+1)$-th pivot, will be in the $(s+1)$-th row).

Then we have

$$B = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,j_0} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,j_0} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ a_{s,1} & a_{s,2} & \cdots\cdots & a_{s,j_0} & \cdots & a_{s,n-1} & a_{s,n} \\ 0 & 0 & \cdots\cdots & a_{s+1,j_0} & \cdots & a_{s+1,n-1} & a_{s+1,n} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

We can then write

$$B = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,j_0} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,j_0} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ a_{s,1} & a_{s,2} & \cdots\cdots & a_{s,j_0} & \cdots & a_{s,n-1} & a_{s,n} \\ 0 & 0 & \cdots\cdots & a_{s+1,j_0} & \cdots & a_{s+1,n-1} & a_{s+1,n} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix} \underbrace{\qquad (a_{s+1,j_0})^{-1} R_{s+1} \to R'_{s+1}}$$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & a_{1,j_0} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots\cdots & a_{2,j_0} & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ a_{s,1} & a_{s,2} & \cdots\cdots & a_{s,j_0} & \cdots & a_{s,n-1} & a_{s,n} \\ 0 & 0 & \cdots\cdots & 1 & \cdots & \frac{a_{s+1,n-1}}{a_{s+1,j_0}} & \frac{a_{s+1,n}}{a_{s+1,j_0}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix} \underbrace{\qquad \forall i \leqslant s, \quad R_i - a_{i,j_0} R_{s+1} \to R'_i}$$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{1,n-1}} & \widetilde{a_{1,n}} \\ a_{2,1} & a_{2,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{2,n-1}} & \widetilde{a_{2,n}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ a_{s,1} & a_{s,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{s,n-1}} & \widetilde{a_{s,n}} \\ 0 & 0 & \cdots\cdots & 1 & \cdots & \frac{a_{s+1,n-1}}{a_{s+1,j_0}} & \frac{a_{s+1,n}}{a_{s+1,j_0}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Proof of Theorem 1b′

We finally apply our Inductive Hypothesis to the upper $s \times n$ submatrix of

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{1,n-1}} & \widetilde{a_{1,n}} \\ a_{2,1} & a_{2,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{2,n-1}} & \widetilde{a_{2,n}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ a_{s,1} & a_{s,2} & \cdots\cdots & 0 & \cdots & \widetilde{a_{s,n-1}} & \widetilde{a_{s,n}} \\ 0 & 0 & \cdots\cdots & 1 & \cdots & \frac{a_{s+1,n-1}}{a_{s+1,j_0}} & \frac{a_{s+1,n}}{a_{s+1,j_0}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Note that this submatrix $\widetilde{B}$ has exactly s pivots, all before the j_0-th column, so this will give a matrix $\widetilde{C} \in \mathbb{F}^{s \times n}$ which is row equivalent to $\widetilde{B}$, is in RREF, and has s pivots as well, in the same positions as $\widetilde{B}$. *Also, its j_0-th column is a zero column (since the j_0-th column of $\widetilde{B}$ has only zero entries, the entries above 1 in the above matrix).*
But then the matrix

$$\begin{pmatrix} & & & \widetilde{C} & & & \\ 0 & 0 & \cdots\cdots & 1 & \cdots & \frac{a_{s+1,n-1}}{a_{s+1,j_0}} & \frac{a_{s+1,n}}{a_{s+1,j_0}} \\ \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots\cdots & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

is an RREF of the original matrix B with the desired properties.

Theorem 1 again

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Important Corollary of Theorem 1 and all previous arguments

Given that we proved the stronger Theorem 1b′, and given that Theorem 1 has one remaining statement (which we haven't proven yet, but we can already use in other arguments), namely that for every matrix $A \in \mathbb{F}^{m\times n}$ **its RREF is unique**, we can also conclude the following:

Let A be a matrix in $\mathbb{F}^{m\times n}$,
and let $B_1, B_2 \in \mathbb{F}^{m\times n}$ be two REFs of A.
Then B_1 and B_2 have the same number of pivots,
and the positions of the pivots are the same in both matrices.

Justification. Let $C \in \mathbb{F}^{m\times n}$ be the unique RREF of A.

– Then since $B_1 \sim A$ and $A \sim C$, we get that $B_1 \sim C$. Thus C is an RREF of B_1, and hence it must be the unique RREF of B_1. But as we saw before, we can find an RREF of B_1 which has the same number of pivots as B_1, and in the same positions as B_1 does; since C is the unique RREF of B_1, C must have these properties.

– Similarly, we can see that C is the unique RREF of B_2, and hence C has the same number of pivots as B_2, and in the same positions as B_2 does.

Combining these two conclusions, we can see that B_1 and B_2 have the same number of pivots, which coincides with the number of pivots of C, and also have pivots in the same positions (which are the positions of pivots of C).

An example

Let A be the following matrix from $\mathbb{Z}_7^{3\times 3}$:

$$A = \begin{pmatrix} 4 & 5 & 1 \\ 4 & 1 & 3 \\ 1 & 6 & 4 \end{pmatrix}.$$

Both matrices below are REFs of A (check this yourselves):

$$B_1 = \begin{pmatrix} 4 & 5 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} 1 & 6 & 4 \\ 0 & 5 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that both matrices have 2 pivots, in positions $(1,1)$ and $(2,2)$.

Also the matrix

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

is the unique RREF of A (and again has pivots in positions $(1,1)$ and $(2,2)$).

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 34

Tuesday November 3

Recall: Theorem 1

Terminology

Let A, B be matrices in $\mathbb{F}^{m\times n}$. Suppose that

- $A \sim B$
- and B is in Row Echelon Form (or in Reduced Row Echelon Form).

Then we say that B is a REF of A (or an RREF of A respectively).

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m\times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m\times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m\times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m\times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m\times n}$, its RREF is unique.**

Recall: Important Corollary of Theorem 1 and the proofs we gave

Given that we proved the stronger Theorem 1b′, and given that Theorem 1 has one remaining statement (which we haven't proven yet, but we can already use in other arguments), namely that for every matrix $A \in \mathbb{F}^{m\times n}$ **its RREF is unique**, we can also conclude the following:

Let A be a matrix in $\mathbb{F}^{m\times n}$,
and let $B_1, B_2 \in \mathbb{F}^{m\times n}$ be two REFs of A.
Then B_1 and B_2 have the same number of pivots,
and the positions of the pivots are the same in both matrices.

Justification. Let $C \in \mathbb{F}^{m\times n}$ be the unique RREF of A.

– Then since $B_1 \sim A$ and $A \sim C$, we get that $B_1 \sim C$. Thus C is an RREF of B_1, and hence it must be the unique RREF of B_1. But as we saw before, we can find an RREF of B_1 which has the same number of pivots as B_1, and in the same positions as B_1 does; since C is the unique RREF of B_1, C must have these properties.

– Similarly, we can see that C is the unique RREF of B_2, and hence C has the same number of pivots as B_2, and in the same positions as B_2 does.

Combining these two conclusions, we can see that B_1 and B_2 have the same number of pivots, which coincides with the number of pivots of C, and also have pivots in the same positions (which are the positions of pivots of C).

An example

Let A be the following matrix from $\mathbb{Z}_7^{3\times 3}$:

$$A = \begin{pmatrix} 4 & 5 & 1 \\ 4 & 1 & 3 \\ 1 & 6 & 4 \end{pmatrix}.$$

Both matrices below are REFs of A (check this yourselves):

$$B_1 = \begin{pmatrix} 4 & 5 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} 1 & 6 & 4 \\ 0 & 5 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that both matrices have 2 pivots, in positions $(1,1)$ and $(2,2)$.

Also the matrix

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

is the unique RREF of A (and again has pivots in positions $(1,1)$ and $(2,2)$).

Corollary of Theorem 1a

Combining the following

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

Proposition 1

Suppose that $C = \left(A \mid \bar{b}\right)$ and $C' = \left(A' \mid \bar{b}'\right)$ are the **augmented** matrices of two linear systems, which are linear systems in the same set of unknowns, with the same number of linear equations (that is, the systems $A\bar{x} = \bar{b}$ and $A'\bar{x} = \bar{b}'$ respectively).

If $C \sim C'$, then the two linear systems are equivalent, or in other words they have the same set of solutions.

we obtain

Corollary

Let $\mathcal{LS}_1$ be a system of $\widetilde{m}$ linear equations in $\widetilde{n}$ unknowns $x_1, x_2, \ldots, x_{\widetilde{n}}$. Then we can replace $\mathcal{LS}_1$ by an **equivalent** **staircase** system $\mathcal{LS}_2$ (with $\widetilde{m}$ linear equations again, in the same set of unknowns) **via finitely many applications of the method of Gaussian elimination.**

Concluding remarks about solving linear systems via Gaussian elimination

Terminology. Since every REF of a matrix $A \in \mathbb{F}^{m\times(n+1)}$ has pivots in the same positions, there is no risk of confusion if we call the pivots of the REF of A, which the proof of Theorem 1a is guaranteed to give us, as the pivots of A as well (the main emphasis here is what the positions of these pivots will be, and this, as we saw, does not depend on the algorithm/proof we rely on to find a REF of A, rather than what the value of each such pivot is).

Similarly, for every linear system $\mathcal{LS}_1$ with coefficients from $\mathbb{F}$ and m linear equations in n unknowns, we call the pivots of the equivalent staircase system $\mathcal{LS}_0$, which the Corollary of Theorem 1 is guaranteed to give us, the pivots of $\mathcal{LS}_1$ as well.

We also call pivot variables of $\mathcal{LS}_1$ (respectively free variables) those variables that are pivot (respectively free) in the equivalent staircase system $\mathcal{LS}_0$.

With this terminology in mind, we now have...

1. The number of pivots is $\leqslant$ the number of rows of the matrix (or equations of the linear system). Similarly, it is $\leqslant$ the number of columns.
   Thus, the number of pivots is $\leqslant \min\{m, n+1\}$.
2. In a staircase system (analogously, in a matrix in REF), the number of pivots $=$ the number of non-trivial equations (the number of non-zero rows, respectively).
   Attention. This statement is not always true when the linear system we consider is not staircase (analogously, when the matrix we consider is not in REF).

Concluding remarks about solving linear systems via Gaussian elimination

3. If $m < n$, then the system has free variables (more generally, if the system is underdetermined, then there are free variables).
4. If all variables are pivot variables, then we must have $m \geqslant n$.
5. If there are exactly n pivots, then
   - either the system is inconsistent,
   - or it has a unique solution.
6. If the number of pivots is $< n$, then
   - either the system is inconsistent,
   - or it has more than one solutions.
7. If there are more than n pivots, then the system is inconsistent *(how many pivots will we have in such a case?).*
8. If there is a pivot in the last column of any equivalent staircase system we can find for the system $\mathcal{LS}_1$ that we are considering, then the system $\mathcal{LS}_1$ is inconsistent.
9. The numbers of pivot variables and of free variables are not affected by what the last column is, that is, the column of constant terms. **In other words, if we change only the column of constant terms, but do not change the LHS of the linear equations of the system, then the numbers of pivot variables and of free variables will remain the same.**

Proof of Theorem 2?

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n\times n}$. The following are equivalent:

- A is invertible.
- Every REF of A has exactly n pivots.
- At least one REF of A has exactly n pivots.
- The unique RREF of A is the identity matrix I_n.

Elementary Row Operations, but using Matrix Multiplication

Recall the three different ways we have to apply the method of Gaussian elimination to the rows of a matrix:

1. We can multiply a row of the matrix **by a non-zero constant**.
2. We can add a multiple of one row of the matrix, say Row_i, to another row, say Row_j, and replace Row_j by the new row we get *(note that, here, Row_i is kept where it was, unchanged).*
3. We can swap two rows (that is, start writing Row_i where we had Row_j before, while writing Row_j where we had Row_i).

Additional terminology. We call such operations on a matrix *elementary row operations*.

Elementary Row Operations, but using Matrix Multiplication

A very useful idea: Given a matrix A, could we perform elementary row operations on A (of each of the three different types above) by simply mutliplying A by another (suitably chosen) matrix E from the left?

This leads to what we call **Elementary Matrices.**

Question. How do we find what these matrices should be?

Answer. We perform the corresponding elementary operation on the identity matrix.

1. E.g. try an operation of Type 1 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{4R_2 \to R_2'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix we got is an elementary matrix. Note also that, for this type of operation, we will always end up with a diagonal matrix, so we can denote it by $D_{2;4}$ (where the first number in the subscript gives the row that is to be multiplied, and the second number in the subscript gives the constant we should multiply by).

Elementary matrices

1. E.g. try an operation of Type 1 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{4R_2 \to R_2'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix we got is an elementary matrix. Note also that, for this type of operation, we will always end up with a diagonal matrix, so we can denote it by $D_{2;4}$ (where the first number in the subscript gives the row that is to be multiplied, and the second number in the subscript gives the constant we should multiply by).

What do we get if we multiply a matrix $A \in \mathbb{R}^{3\times n}$ by $D_{2;4}$ from the left? E.g. if

$$A = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix}?$$

We have

$$D_{2;4} \cdot A = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 8 & -16 & 0 & 4 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by multiplying the 2nd row of A by 4 (that is, the same elementary row operation we performed on I_3 initially).

Elementary matrices (cont.)

2. E.g. try an operation of Type 2 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 + (-2)R_1 \to R_3'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

The matrix we got is another elementary matrix. **We will use the notation $E_{3,1;-2}$ for this matrix** (where the first numbers in the subscript show which row we are changing and for which row we will take a multiple to add to the former row, while the last number in the subscript (which should be an element of the field $\mathbb{F}$) shows what multiple of the latter/auxiliary row we'll take).

What do we get if we multiply the matrix A from before by $E_{3,1;-2}$ from the left? We have

$$E_{3,1;-2} \cdot A = E_{3,1;-2} \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 0 & 1 & 3 & 3 & -2 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by multiplying the 1st row of A by -2 and adding it to its 3rd row, and then replacing the 3rd row by the result (that is, the same elementary row operation we performed on I_3 initially).

Elementary matrices (cont.)

3. E.g. try an operation of Type 3 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The matrix we got is another elementary matrix, which we call a *permutation matrix* (or sometimes *transposition matrix*). **We will use the notation $P_{1,3}$ for this matrix** (where the numbers in the subscript show which rows we are swapping).

What do we get if we multiply the matrix A by $E_{3,1;-2}$ from the left? We have

$$P_{1,3} \cdot A = P_{1,3} \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 5 & 1 & 0 \\ 0 & 2 & -4 & 0 & 1 \\ 2 & 1 & 1 & -1 & 1 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by swapping the 1st row and the 3rd row of A (that is, the same elementary row operation we performed on I_3 initially).

Important Observation

Proposition 1

Elementary matrices are invertible.

More specifically, let $\mathbb{F}$ be a field, and let $n \geqslant 1$. Every elementary matrix in $\mathbb{F}^{n\times n}$ has an inverse in $\mathbb{F}^{n\times n}$ (and its inverse is again an elementary matrix).

Proof. Let us consider an elementary matrix of Type 1 in $\mathbb{F}^{n\times n}$, that is, a matrix of the form $D_{i_0;\lambda}$ where $1 \leqslant i_0 \leqslant n$ and $\lambda \in \mathbb{F}$ is a non-zero element. **Note that this matrix corresponds to the elementary row operation of multiplying the i_0-th row by the non-zero constant λ.**
Since $\lambda \neq 0$, it has a multiplicative inverse λ^{-1}. Consider the matrix $D_{i_0;\lambda^{-1}}$; this corresponds to the elementary row operation of multiplying the i_0-th row by the non-zero constant λ^{-1}. Thus

$$D_{i_0;\lambda^{-1}} \cdot D_{i_0;\lambda} = D_{i_0;\lambda^{-1}} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 0 & \cdots\cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = I_n\,.$$

Similarly...

Important Observation (cont.)

Proposition 1

Elementary matrices are invertible.

More specifically, let $\mathbb{F}$ be a field, and let $n \geqslant 1$. Every elementary matrix in $\mathbb{F}^{n\times n}$ has an inverse in $\mathbb{F}^{n\times n}$ (and its inverse is again an elementary matrix).

Proof. Similarly, we have

$$D_{i_0;\lambda} \cdot D_{i_0;\lambda^{-1}} = D_{i_0;\lambda} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots\ddots & \ddots \\ 0 & 0 & \cdots & \lambda^{-1} & 0 & \cdots\cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots\cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots\ddots & \ddots \\ 0 & 0 & \cdots & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = I_n \,.$$

In the end, we see that $D_{i_0;\lambda}$ is an invertible matrix, and its inverse is the elementary matrix $D_{i_0;\lambda^{-1}}$ (which is of the same type).

Proof of Proposition 1 (cont.)

Consider now a permutation matrix $P_{i,j} \in \mathbb{F}^{n \times n}$, that is, an elementary matrix of Type 3 which corresponds to swapping, say, the i-th row and the j-th row. Then

$$P_{i,j} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

We can then check that $P_{i,j} \cdot P_{i,j} = I_n$, and thus $P_{i,j}$ is an invertible matrix, and furthermore is its own inverse.

Proof of Proposition 1 (cont.)

Finally, consider an elementary matrix of Type 2, that is, a matrix of the form $E_{i,j;\mu}$, where $1 \leqslant i, j \leqslant n$, $i \neq j$, and $\mu \in \mathbb{F}$. We can assume here that $i < j$ because the other case is completely analogous. **Note that this matrix corresponds to the elementary row operation of multiplying the j-th row by μ and adding this to the i-th row, and then replacing the i-th row by the result.**

Let $-\mu$ be the additive inverse of μ in $\mathbb{F}$. We consider also the elementary matrix $E_{i,j;-\mu}$ which corresponds to the elementary row operation of multiplying the j-th row by $-\mu$ and adding this to the i-th row, and then replacing the i-th row by the result. We then have

$$E_{i,j;-\mu} \cdot E_{i,j;\mu} = E_{i,j;-\mu} \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots\cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots\cdots & 0 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots\cdots & \mu & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots\cdots & 1 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots\cdots & 1 \end{pmatrix} = I_n.$$

Similarly $E_{i,j;\mu} \cdot E_{i,j;-\mu} = I_n$. Thus $E_{i,j;\mu}$ is an invertible matrix, and its inverse is an elementary matrix of the same type.

Proof of Theorem 2

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n\times n}$. The following are equivalent:

- A is invertible.
- Every REF of A has exactly n pivots.
- At least one REF of A has exactly n pivots.
- The unique RREF of A is the identity matrix I_n.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 35

Wednesday November 4

Elementary Row Operations, but using Matrix Multiplication

A very useful idea: Given a matrix A, could we perform elementary row operations on A (of each of the three different types above) by simply mutliplying A by another (suitably chosen) matrix E from the left?

This leads to what we call **Elementary Matrices.**

Question. How do we find what these matrices should be?

Answer. We perform the corresponding elementary operation on the identity matrix.

1. E.g. try an operation of Type 1 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{4R_2 \to R_2'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix we got is an elementary matrix. Note also that, for this type of operation, we will always end up with a diagonal matrix, so we can denote it by $D_{2;4}$ (where the first number in the subscript gives the row that is to be multiplied, and the second number in the subscript gives the constant we should multiply by).

Elementary matrices

1. E.g. try an operation of Type 1 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{4R_2 \to R_2'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix we got is an elementary matrix. Note also that, for this type of operation, we will always end up with a diagonal matrix, so we can denote it by $D_{2;4}$ (where the first number in the subscript gives the row that is to be multiplied, and the second number in the subscript gives the constant we should multiply by).

What do we get if we multiply a matrix $A \in \mathbb{R}^{3\times n}$ by $D_{2;4}$ from the left? E.g. if

$$A = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix}?$$

We have

$$D_{2;4} \cdot A = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 8 & -16 & 0 & 4 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by multiplying the 2nd row of A by 4 (that is, the same elementary row operation we performed on I_3 initially).

Elementary matrices (cont.)

2. E.g. try an operation of Type 2 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 + (-2)R_1 \to R_3'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

The matrix we got is another elementary matrix. **We will use the notation $E_{3,1;-2}$ for this matrix** (where the first numbers in the subscript show which row we are changing and for which row we will take a multiple to add to the former row, while the last number in the subscript (which should be an element of the field $\mathbb{F}$) shows what multiple of the latter/auxiliary row we'll take).

What do we get if we multiply the matrix A from before by $E_{3,1;-2}$ from the left? We have

$$E_{3,1;-2} \cdot A = E_{3,1;-2} \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 0 & 1 & 3 & 3 & -2 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by multiplying the 1st row of A by -2 and adding it to its 3rd row, and then replacing the 3rd row by the result (that is, the same elementary row operation we performed on I_3 initially).

Elementary matrices (cont.)

3. E.g. try an operation of Type 3 on the identity matrix in $\mathbb{R}^{3\times 3}$:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The matrix we got is another elementary matrix, which we call a *permutation matrix* (or sometimes *transposition matrix*). **We will use the notation $P_{1,3}$ for this matrix** (where the numbers in the subscript show which rows we are swapping).

What do we get if we multiply the matrix A by $E_{3,1;-2}$ from the left? We have

$$P_{1,3} \cdot A = P_{1,3} \begin{pmatrix} 2 & 1 & 1 & -1 & 1 \\ 0 & 2 & -4 & 0 & 1 \\ 4 & 3 & 5 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 5 & 1 & 0 \\ 0 & 2 & -4 & 0 & 1 \\ 2 & 1 & 1 & -1 & 1 \end{pmatrix}.$$

Thus we get the same matrix we would obtain by swapping the 1st row and the 3rd row of A (that is, the same elementary row operation we performed on I_3 initially).

Important Observation

Proposition 1

Elementary matrices are invertible.

More specifically, let $\mathbb{F}$ be a field, and let $n \geqslant 1$. Every elementary matrix in $\mathbb{F}^{n\times n}$ has an inverse in $\mathbb{F}^{n\times n}$ (and its inverse is again an elementary matrix).

Proof. Let us consider an elementary matrix of Type 1 in $\mathbb{F}^{n\times n}$, that is, a matrix of the form $D_{i_0;\lambda}$ where $1 \leqslant i_0 \leqslant n$ and $\lambda \in \mathbb{F}$ is a non-zero element. **Note that this matrix corresponds to the elementary row operation of multiplying the i_0-th row by the non-zero constant λ.**
Since $\lambda \neq 0$, it has a multiplicative inverse λ^{-1}. Consider the matrix $D_{i_0;\lambda^{-1}}$; this corresponds to the elementary row operation of multiplying the i_0-th row by the non-zero constant λ^{-1}. Thus

$$D_{i_0;\lambda^{-1}} \cdot D_{i_0;\lambda} = D_{i_0;\lambda^{-1}} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 0 & \cdots\cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = I_n\,.$$

Similarly...

Important Observation (cont.)

Proposition 1

Elementary matrices are invertible.

More specifically, let $\mathbb{F}$ be a field, and let $n \geqslant 1$. Every elementary matrix in $\mathbb{F}^{n\times n}$ has an inverse in $\mathbb{F}^{n\times n}$ (and its inverse is again an elementary matrix).

Proof. Similarly, we have

$$D_{i_0;\lambda} \cdot D_{i_0;\lambda^{-1}} = D_{i_0;\lambda} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots\cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots\ddots & \ddots \\ 0 & 0 & \cdots & \lambda^{-1} & 0 & \cdots\cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots\cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots\ddots & \ddots \\ 0 & 0 & \cdots & 0 & 0 & \cdots\cdots & 1 \end{pmatrix} = I_n \,.$$

In the end, we see that $D_{i_0;\lambda}$ is an invertible matrix, and its inverse is the elementary matrix $D_{i_0;\lambda^{-1}}$ (which is of the same type).

Proof of Proposition 1 (cont.)

Consider now a permutation matrix $P_{i,j} \in \mathbb{F}^{n \times n}$, that is, an elementary matrix of Type 3 which corresponds to swapping, say, the i-th row and the j-th row. Then

$$P_{i,j} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

We can then check that $P_{i,j} \cdot P_{i,j} = I_n$, and thus $P_{i,j}$ is an invertible matrix, and furthermore is its own inverse.

Proof of Proposition 1 (cont.)

Finally, consider an elementary matrix of Type 2, that is, a matrix of the form $E_{i,j;\mu}$, where $1 \leqslant i, j \leqslant n$, $i \neq j$, and $\mu \in \mathbb{F}$. We can assume here that $i < j$ because the other case is completely analogous. **Note that this matrix corresponds to the elementary row operation of multiplying the j-th row by μ and adding this to the i-th row, and then replacing the i-th row by the result.**

Let $-\mu$ be the additive inverse of μ in $\mathbb{F}$. We consider also the elementary matrix $E_{i,j;-\mu}$ which corresponds to the elementary row operation of multiplying the j-th row by $-\mu$ and adding this to the i-th row, and then replacing the i-th row by the result. We then have

$$E_{i,j;-\mu} \cdot E_{i,j;\mu} = E_{i,j;-\mu} \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & \mu & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} = I_n.$$

Similarly $E_{i,j;\mu} \cdot E_{i,j;-\mu} = I_n$. Thus $E_{i,j;\mu}$ is an invertible matrix, and its inverse is an elementary matrix of the same type.

Examples

Recall the following elementary matrices from $\mathbb{R}^{3\times 3}$ that we considered last time:

1. The inverse of the matrix

$$D_{2;4} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the matrix

$$D_{2;\,0.25} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Examples (cont.)

2. The inverse of the matrix

$$E_{3,1;\,-2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

is the matrix

$$E_{3,1;\,2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

3. Finally, the inverse of the matrix

$$P_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is itself.

Proof of Theorem 2

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n\times n}$. The following are equivalent:

(I) A is invertible.

(II) Every REF of A has exactly n pivots.

(III) At least one REF of A has exactly n pivots.

(IV) The unique RREF of A is the identity matrix I_n.

Proof. **We need to establish each of the following equivalences:**

$$(I) \Leftrightarrow (II), \quad (I) \Leftrightarrow (III), \quad (I) \Leftrightarrow (IV),$$
$$(II) \Leftrightarrow (III), \quad (II) \Leftrightarrow (IV), \quad (III) \Leftrightarrow (IV).$$

We can prove all these if we show the following implications:

$$(I) \Rightarrow (II), \quad (II) \Rightarrow (III), \quad (III) \Rightarrow (IV), \quad \textbf{and} \quad (IV) \Rightarrow (I),$$

or, to write it more simply, we need to show that

$$(I) \Rightarrow (II) \Rightarrow (III) \Rightarrow (IV) \Rightarrow (I).$$

Proof of Theorem 2 (cont.)

$(I) \Rightarrow (II)$ Let us assume that A is invertible. We consider the linear system

$$A \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Note that the augmented matrix of this system is $\left(A \mid \bar{0}\right)$.

Since A is invertible, this system has a unique solution, the solution $A^{-1} \cdot \bar{0} = \bar{0}$.

Consider now a Row Echelon Form B of A. Then $A \sim B$, and hence we can also see that $\left(A \mid \bar{0}\right) \sim \left(B \mid \bar{0}\right)$ (in fact, the same elementary row operations that transform A to B will transform $\left(A \mid \bar{0}\right)$ to $\left(B \mid \bar{0}\right)$, given that each elementary row operation leaves the last zero column unchanged, and acts on the previous columns as before).

Since $\left(A \mid \bar{0}\right) \sim \left(B \mid \bar{0}\right)$, we get that the linear systems $A \cdot \bar{x} = \bar{0}$ and $B \cdot \bar{x} = \bar{0}$ are equivalent, so they both have a unique solution.

But, as we have seen, in the case of $B \cdot \bar{x} = \bar{0}$ this can only happen if all variables of the system are pivot variables, or in other words if B has n pivots.

We conclude that the arbitrary Row Echelon Form B of A has n pivots.

Proof of Theorem 2 (cont.)

$(II) \Rightarrow (III)$ Let us assume that every REF of A has n pivots.

By Theorem 1a we know that there exists at least one REF B_0 of A. Then this matrix B_0 must have n pivots, and its existence confirms statement (III).

$(III) \Rightarrow (IV)$ Suppose that there exists a REF B_0 of A with exactly n pivots. Let C be the unique RREF of A. Then, as we have seen, this is also the unique RREF of B_0 (indeed, we have that $B_0 \sim A$ and $A \sim C$, which combined imply that $B_0 \sim C$; hence, C is an RREF of B_0, and thus it must the unique RREF of B_0).

But as we established in the proof of Theorem 1b, B_0 and C must have the same number of pivots. Thus C is a square matrix in RREF with n columns, all of which are pivot columns. This necessarily implies that $C = I_n$. In other words, the unique RREF of A is the identity matrix I_n.

Proof of Theorem 2 (cont.)

$(IV) \Rightarrow (I)$ Let us assume that the unique RREF of A is the identity matrix I_n.

Then we have that $A \sim I_n$, or in other words we can get from A to I_n doing finitely many elementary row operations.

Equivalently, there are k elementary matrices $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ (for some $k \geqslant 1$) such that

$$\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \cdot A = I_n.$$

But we have seen that elementary matrices are invertible, thus the product $\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1$ is invertible (recall HW3, Problem 1), and its inverse is the matrix

$$\mathcal{E}_1^{-1} \mathcal{E}_2^{-1} \cdots \mathcal{E}_k^{-1}.$$

We conclude that

$$\begin{aligned} A &= \left(\mathcal{E}_1^{-1} \mathcal{E}_2^{-1} \cdots \mathcal{E}_k^{-1}\right) \cdot \left(\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \cdot A\right) \\ &= \left(\mathcal{E}_1^{-1} \mathcal{E}_2^{-1} \cdots \mathcal{E}_k^{-1}\right) \cdot I_n = \mathcal{E}_1^{-1} \mathcal{E}_2^{-1} \cdots \mathcal{E}_k^{-1}. \end{aligned}$$

In other words, A is an invertible matrix as it can be written as the product of invertible matrices.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 36

Monday November 16

Reminders from last few lectures

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m \times n}$, its RREF is unique.**

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n \times n}$. The following are equivalent:

(I) A is invertible.

(II) Every REF of A has exactly n pivots.

(III) At least one REF of A has exactly n pivots.

(IV) The unique RREF of A is the identity matrix I_n.

How can we use the theorems?

2nd Midterm Exam Problem. Consider the following matrices from $\mathbb{Z}_5^{4\times 3}$. Find all the pairs of row equivalent matrices. Justify your answer.

$$A_1 = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & 4 \\ 2 & 0 & 4 \\ 0 & 4 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 3 & 0 & 1 \\ 2 & 1 & 3 \\ 3 & 3 & 2 \\ 0 & 4 & 1 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \qquad A_5 = \begin{pmatrix} 2 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 1 & 3 \\ 2 & 1 & 4 \end{pmatrix}, \qquad A_6 = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 1 & 4 \\ 2 & 0 & 3 \\ 1 & 3 & 4 \end{pmatrix}.$$

Suggested solution

We recall that each of the matrices A_i above has a **unique** RREF, which we can denote by $\mathrm{RREF}(A_i)$.

- Suppose now that the matrices A_i and A_j are row equivalent. Then, we will have

$$A_i \sim A_j, \quad \text{and at the same time} \quad A_j \sim \mathrm{RREF}(A_j),$$
$$\text{which will imply} \quad A_i \sim \mathrm{RREF}(A_j) \text{ too.}$$

Therefore, $\mathrm{RREF}(A_j)$ will be an RREF of A_i, and given that this should be unique, we will have $\mathrm{RREF}(A_i) = \mathrm{RREF}(A_j)$.

- Conversely, if we have that $\mathrm{RREF}(A_i) = \mathrm{RREF}(A_j)$, then we can write

$$A_i \sim \mathrm{RREF}(A_i) \quad \text{and} \quad A_j \sim \mathrm{RREF}(A_j) = \mathrm{RREF}(A_i)$$
$$\Rightarrow \quad A_i \sim \mathrm{RREF}(A_i) \quad \text{and} \quad \mathrm{RREF}(A_i) \sim A_j$$
$$\Rightarrow \quad A_i \sim A_j.$$

Conclusion

Two matrices A_i, A_j from $\mathbb{F}^{m\times n}$ are row equivalent **if and only if** they have the same RREF, that is, $\mathrm{RREF}(A_i) = \mathrm{RREF}(A_j)$.

Applying this to the given matrices

$$A_1 = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & 4 \\ 2 & 0 & 4 \\ 0 & 4 & 1 \end{pmatrix} \xrightarrow[4R_4 \to R_4']{R_3 - R_1 \to R_2'} \begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 3 & 1 \\ 0 & 1 & 4 \end{pmatrix} \xrightarrow[3R_1 \to R_1',\ 2R_3 \to R_3']{R_4 - R_2 \to R_4'} \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow[R_3 - R_2 \to R_3']{R_1 - R_2 \to R_1'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{2R_3 \to R_3'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_2 - 4R_3 \to R_2'} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Similarly,

$$A_2 = \begin{pmatrix} 3 & 0 & 1 \\ 2 & 1 & 3 \\ 3 & 3 & 2 \\ 0 & 4 & 1 \end{pmatrix} \xrightarrow[R_3 - R_1 \to R_3']{R_2 - 4R_1 \to R_2'} \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & 4 \\ 0 & 3 & 1 \\ 0 & 4 & 1 \end{pmatrix} \xrightarrow{2R_1 \to R_1',\ 4R_4 \to R_4'} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 4 \\ 0 & 3 & 1 \\ 0 & 1 & 4 \end{pmatrix}$$

$$\cdots\cdots \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Applying this to the given matrices

Similarly,

$$A_3 = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{\substack{2R_1 \to R_1', \\ 4R_2 \to R_2'}} \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_1 - 4R_2 \to R_1'} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

From what we have seen so far, $\mathrm{RREF}(A_1) = \mathrm{RREF}(A_2)$, while $\mathrm{RREF}(A_3) \neq \mathrm{RREF}(A_1) = \mathrm{RREF}(A_2)$. Therefore, we can conclude that $A_1 \sim A_2$, while A_3 is **not** row equivalent to A_1 or to A_2.

We can continue like this in order to determine whether any of the matrices A_4, A_5 and A_6 are also pairwise row equivalent and/or row equivalent to any of the matrices we have already seen.

Theorem 1 again

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m \times n}$, its RREF is unique.**

Proving the statement in red? Recall that

- we proved Theorem 1a using mathematical induction in the number m of rows of the matrix A.
- we proved the first part of Theorem 1b using mathematical induction in the number s of pivots of the REF of A that we had already found when 'algorithmically' verifying part (a).

Proof of the uniqueness of the RREF

This time, we will use mathematical induction in the number n of columns of A.

What we need to show is the following: if $A_1, A_2 \in \mathbb{F}^{m\times n}$ are two matrices in RREF which are both row equivalent to A, then $A_1 = A_2$.

Note that by our assumptions we have the following:

$$\begin{aligned} & A \sim A_1 \quad \text{and} \quad A \sim A_2 \\ \Rightarrow \quad & A_1 \sim A \quad \text{and} \quad A \sim A_2 \\ \Rightarrow \quad & A_1 \sim A_2. \end{aligned}$$

Thus, we will obtain the desired conclusion if we prove the following

Claim. Let $B, C \in \mathbb{F}^{m\times n}$ be matrices **in RREF**.
If we have $B \sim C$, then necessarily $B = C$.

We now prove the claim using induction in n.

A proposition needed for the induction step

Auxiliary proposition

Let $\widetilde{Q} \in \mathbb{F}^{m \times (n+1)}$ be a matrix in RREF, and write

$$\widetilde{Q} = (Q \mid \bar{q})$$

where Q is the matrix formed by the **first** n columns of $\widetilde{Q}$.

Then Q is also a matrix in RREF.

Proof of the proposition. We can distinguish two cases:

Case 1: $\bar{q}$, that is, the last column of $\widetilde{Q}$ is **not** a pivot column of $\widetilde{Q}$. Then all the pivots of $\widetilde{Q}$ are found in its first n columns, that is, the columns of Q, and hence Q has as many non-zero rows as $\widetilde{Q}$ does (and in the same positions). Therefore, all non-zero rows of Q are above any zero rows.

Moreover, the pivot columns of Q are the same as the pivot columns of $\widetilde{Q}$, and thus each of them has exactly one non-zero entry, which will be equal to 1.

Finally, the first non-zero entry of any non-zero row of Q is precisely the first non-zero entry of the corresponding row in $\widetilde{Q}$ (since we only ignore the last entry of that row in $\widetilde{Q}$ when looking at Q). Thus, each first non-zero entry is to the right of previous such entries.

A proposition needed for the induction step

Auxiliary proposition

Let $\widetilde{Q} \in \mathbb{F}^{m\times(n+1)}$ be a matrix in RREF, and write

$$\widetilde{Q} = (Q \mid \bar{q})$$

where Q is the matrix formed by the **first** n columns of $\widetilde{Q}$.

Then Q is also a matrix in RREF.

Proof of the proposition. We can distinguish two cases:

Case 2: $\bar{q}$, the last column of $\widetilde{Q}$, is a pivot column of $\widetilde{Q}$.

In this case, we know that $\bar{q}$ contains the last pivot of $\widetilde{Q}$, that is, **the pivot found in the last non-zero row of $\widetilde{Q}$;** moreover, all other pivots of $\widetilde{Q}$ *(if any exist)* are not only above this last pivot, but also to the left of it, **and hence they must be entries of Q** *(note that, if these were not true, then the condition from the definition of RREF, that each pivot must be to the right of previous pivots, would be violated).*

Therefore, the non-zero rows of Q are the same as the non-zero rows of $\widetilde{Q}$, **except for the last non-zero row of $\widetilde{Q}$, whose part in Q is a zero row.** But then we can check analogously to above that all conditions from the definition of RREF are satisfied for Q.

This completes the proof of the proposition.

Applying the proposition in examples

Consider the matrices

$$\widetilde{Q_1} = \begin{pmatrix} 1 & 2 & 0 & 0 & 5 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{4\times 5}, \qquad \widetilde{Q_2} = \begin{pmatrix} 1 & 0 & 3 & 4 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Z}_5^{4\times 5}.$$

Observe that both matrices are in RREF.

Note that the pivot columns of $\widetilde{Q_1}$ are $\mathrm{Col}_1(\widetilde{Q_1})$, $\mathrm{Col}_3(\widetilde{Q_1})$ and $\mathrm{Col}_4(\widetilde{Q_1})$. Therefore, the last column of $\widetilde{Q_1}$ is **not** a pivot column, so the first case of the proof of the proposition applies here. As expected (based on what we proved), the matrix

$$Q_1 = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is in RREF too.

On the other hand, the last column of $\widetilde{Q_2}$ is a pivot column of $\widetilde{Q_2}$, so this example falls into the second case of the proof of the proposition. Again as expected, the matrix

$$Q_2 = \begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is in RREF.

Applying the proposition in examples (cont.)

Question 1. What kind of matrix do we get if we 'delete' the third column of $\widetilde{Q_1}$ instead of its last column?

Answer. We get the following matrix, which is not even in REF:

$$\begin{pmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Question 2. *(analogous to the above question)* What kind of matrix do we get if we 'delete' the first column of $\widetilde{Q_2}$?

Answer. We get the following matrix, which again fails to be a matrix in REF:

$$\begin{pmatrix} 0 & 3 & 4 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Conclusion. The above examples show that it was crucial that we considered working with/'deleting' the last column of a given matrix in RREF for the conclusion of the proposition to be true in all cases.

A related observation. If we already know the matrix $\widetilde{Q_0}$ in RREF that we will work with, then we can also consider 'deleting' any column of $\widetilde{Q_0}$ that is **not** a pivot column. It can be checked that the resulting submatrix will also be in RREF *(however in this case we need to know first which are the pivot columns of $\widetilde{Q_0}$ and which are not).*

Proof of the claim leading to the uniqueness of RREF

We need to establish the claim

Let $B, C \in \mathbb{F}^{m \times n}$ be matrices **in RREF**.
If we have $B \sim C$, then necessarily $B = C$.

We will do so using induction in n.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 37

Tuesday November 17

Reminders from last few lectures

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m \times n}$, its RREF is unique.**

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n \times n}$. The following are equivalent:

(I) A is invertible.

(II) Every REF of A has exactly n pivots.

(III) At least one REF of A has exactly n pivots.

(IV) The unique RREF of A is the identity matrix I_n.

Proof of the uniqueness of the RREF?

What we need to show is the following: if $A_1, A_2 \in \mathbb{F}^{m\times n}$ are two matrices in RREF which are both row equivalent to A, then $A_1 = A_2$.

Note that by our assumptions we have the following:

$$\begin{aligned} & A \sim A_1 \quad \text{and} \quad A \sim A_2 \\ \Rightarrow \quad & A_1 \sim A \quad \text{and} \quad A \sim A_2 \\ \Rightarrow \quad & A_1 \sim A_2. \end{aligned}$$

Thus, we will obtain the desired conclusion if we prove the following

Claim. Let $B, C \in \mathbb{F}^{m\times n}$ be matrices **in RREF**.
If we have $B \sim C$, then necessarily $B = C$.

We now prove the claim using induction in the number n of columns of the matrices.

Proof of the uniqueness of RREF

We need to establish the claim

Let $B, C \in \mathbb{F}^{m \times n}$ be matrices **in RREF**.
If we have $B \sim C$, then necessarily $B = C$.

Base of the induction: $n = 1$. Then $B = \bar{b}$ is a column vector in $\mathbb{F}^m \equiv \mathbb{F}^{m \times 1}$, and similarly $C = \bar{c}$ is a column vector in $\mathbb{F}^m$.

Given that $\bar{b} \sim \bar{c}$, we must have that

- either both $\bar{b}$ and $\bar{c}$ equal the zero vector $\bar{0}$ in $\mathbb{F}^m$,
- or both $\bar{b}$ and $\bar{c}$ are non-zero.

In the former case, we immediately get that $\bar{b} = \bar{c}$.

In the latter case, we have that $\bar{b}$ has at least one non-zero row, and therefore it has at least one pivot. But then its only column is a pivot column, and therefore, given that $\bar{b}$ is in RREF, its only pivot column equals $\bar{e}_1$ *(note that the only non-zero entry of $\bar{b}$ must be at the top, so that the condition that all zero rows are below any non-zero rows is satisfied).*

Similarly, we can argue that $\bar{c}$ must equal $\bar{e}_1$, which finally gives us that $\bar{b} = \bar{c}$ in this case too.

Proof of the induction step

We now assume that, for some $n \geqslant 1$, we have that

$$\text{for every two matrices } B, C \in \mathbb{F}^{m \times n} \text{ which are } \textbf{in RREF}.$$
$$\text{if } B \sim C, \text{ then necessarily } B = C.$$

Note that the statement in green is the Induction Hypothesis.

Consider now two matrices $\widetilde{B}, \widetilde{C} \in \mathbb{F}^{m \times (n+1)}$ which are in RREF. Let us write

$$\widetilde{B} = (B \mid \bar{b}) \quad \text{and} \quad \widetilde{C} = (C \mid \bar{c})$$

where B is the submatrix formed by the first n columns of $\widetilde{B}$, and similarly C is the submatrix formed by the first n columns of $\widetilde{C}$.

Reminder from last time...

Recall here a proposition that we proved in the previous lecture, which will help us invoke the Induction Hypothesis:

Proposition

Let $\widetilde{Q} \in \mathbb{F}^{m \times (n+1)}$ be a matrix in RREF, and write

$$\widetilde{Q} = (Q \mid \bar{q})$$

where Q is the matrix formed by the **first** n columns of $\widetilde{Q}$.

Then Q is also a matrix in RREF.

Proof of the induction step

We now assume that, for some $n \geqslant 1$, we have that

$$\text{for every two matrices } B, C \in \mathbb{F}^{m\times n} \text{ which are } \textbf{in RREF}.$$
$$\text{if } B \sim C, \text{ then necessarily } B = C.$$

Note that the statement in green is the Induction Hypothesis.

Consider now two matrices $\widetilde{B}, \widetilde{C} \in \mathbb{F}^{m\times(n+1)}$ which are in RREF. Let us write

$$\widetilde{B} = (B \mid \bar{b}) \quad \text{and} \quad \widetilde{C} = (C \mid \bar{c})$$

where B is the submatrix formed by the first n columns of $\widetilde{B}$, and similarly C is the submatrix formed by the first n columns of $\widetilde{C}$.

- By this proposition we have that both submatrices $B, C \in \mathbb{F}^{m\times n}$ are in RREF.
- We also have that $B \sim C$. Indeed, since we have assumed that $\widetilde{B} \sim \widetilde{C}$, we can find elementary matrices $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k \in \mathbb{F}^{m\times m}$ so that

$$\widetilde{C} = \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \widetilde{B}$$
$$\Rightarrow \quad (C \mid \bar{c}) = \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 (B \mid \bar{b}) = \left(\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 B \mid \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \bar{b}\right)$$
$$\Rightarrow \quad C = \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 B.$$

Thus by the Induction Hypothesis we get that $B = C$.
It remains to show that $\bar{b} = \bar{c}$.

Completing the proof of the induction step

1st way. Assume towards a contradiction that $\bar{b} \neq \bar{c}$. In this case, we can find $i \in \{1, 2, \ldots, m\}$ such that $b_i \neq c_i \Leftrightarrow b_i - c_i \neq 0_{\mathbb{F}}$.

We now consider the linear systems $\widetilde{B}\bar{x} = \bar{0}$ and $\widetilde{C}\bar{x} = \bar{0}$, where $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix}$ and $\bar{0} \in \mathbb{F}^m$. Both systems are consistent, since both admit the trivial solution.

At the same time, since we have assumed that $\widetilde{C} = \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \widetilde{B}$ for some elementary matrices $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k \in \mathbb{F}^{m \times m}$, we also obtain that

$$\left(\widetilde{C} \mid \bar{0}\right) = \left(\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \widetilde{B} \mid \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \bar{0}\right) = \mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 \left(\widetilde{B} \mid \bar{0}\right)$$

$$\Rightarrow \quad \left(\widetilde{B} \mid \bar{0}\right) \sim \left(\widetilde{C} \mid \bar{0}\right).$$

Thus the two systems are equivalent, which shows that they have exactly the same solutions.

Completing the proof of the induction step

Consider now an arbitrary common solution of the two systems, that is, $\bar{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \\ u_{n+1} \end{pmatrix} \in \mathbb{F}^{n+1}$ such that $\widetilde{B}\bar{u} = \bar{0} = \widetilde{C}\bar{u}$. Then we will have

$$\left(\widetilde{B} - \widetilde{C}\right)\bar{u} = \widetilde{B}\bar{u} - \widetilde{C}\bar{u} = \bar{0}.$$

At the same time, the i-th component of $\left(\widetilde{B} - \widetilde{C}\right)\bar{u}$ is

$$\left\langle \mathrm{Row}_i\left(\widetilde{B} - \widetilde{C}\right), \bar{u} \right\rangle = (b_i - c_i) \cdot u_{n+1},$$

with equality here holding because, as we have already discussed, the first n columns of $\widetilde{B} - \widetilde{C}$ are zero, or equivalently the first n components of each row of $\widetilde{B} - \widetilde{C}$ are zero.

Therefore, $(b_i - c_i) \cdot u_{n+1} = 0$, and since we have assumed that $b_i - c_i \neq 0$, we must have u_{n+1}. In other words, for every solution $\bar{u}$ of the two systems, we must have $u_{n+1} = 0$.

This shows that x_{n+1} cannot be a free variable of either of the two systems, and hence both the last column of $\widetilde{B}$ and the last column of $\widetilde{C}$ must be pivot columns. But since $\widetilde{B}$ is in RREF, any of its pivot columns must be equal to a standard basis vector of $\mathbb{F}^m$; say its last column is equal to $\bar{e}_{j_1}$. Similarly, the last column of $\widetilde{C}$ must be equal to a standard basis vector $\bar{e}_{j_2}$.

Since we assumed $\bar{b} \neq \bar{c}$, we have $j_1 \neq j_2$. We can assume that $j_1 < j_2$ (the other case can be handled very analogously). But then, the j_1-th row of $\widetilde{C}$ will be zero, while its j_2-th row will be non-zero. This contradicts the assumption that $\widetilde{C}$ is in RREF, and hence all its non-zero rows are above any zero rows.

Thus, the assumption that the last columns of $\widetilde{B}$ and $\widetilde{C}$ are different was incorrect.

Completing the proof of the induction step

Alternative way (suggested by a student during the lecture). Again, we assume towards a contradiction that $\bar{b} \neq \bar{c}$, and find $i \in \{1, 2, \ldots, m\}$ such that $b_i \neq c_i$.

We now consider the linear systems $B\bar{y} = \bar{b}$ and $C\bar{y} = \bar{c}$. Since

$$(B \mid \bar{b}) = \widetilde{B} \;\sim\; \widetilde{C} = (C \mid \bar{c}),$$

these two systems are equivalent, and hence

- either they are both inconsistent,
- or they are both consistent.

Since $(B \mid \bar{b})$ is in RREF, we have that, if $B\bar{y} = \bar{b}$ is inconsistent, then the last column of $(B \mid \bar{b})$ is a pivot column. Similarly, we see that in this case the last column of $(C \mid \bar{c}) = (B \mid \bar{c})$ is a pivot column (recall that we have already seen that $B = C$).

At this point, we can reach a contradiction in the same way as in the 1st proof we gave.

In the other case that we must consider now, that is, when both systems are consistent, we find one common solution $\bar{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{F}^n$. Then we can write

$$\bar{b} = B\bar{w} \quad \text{and} \quad \bar{c} = C\bar{w} = B\bar{w} \qquad \Rightarrow \qquad \bar{b} = \bar{c}.$$

This contradicts the assumption that $\bar{b} \neq \bar{c}$, and thus we conclude again that this assumption was incorrect.

Theorems 1 and 2 again

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m \times n}$, its RREF is unique.**

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n \times n}$. The following are equivalent:

(I) A is invertible.

(II) Every REF of A has exactly n pivots.

(III) At least one REF of A has exactly n pivots.

(IV) The unique RREF of A is the identity matrix I_n.

How can we use the theorems? (cont.)

Recall the following matrix from HW3:

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 11 & 4 & 2 \\ 2 & -2 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Question 1. How did we check that A is invertible?

Question 2. How can we find the inverse of A?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 38

Wednesday November 18

Theorems 1 and 2 from last lectures again

Theorem 1

(a) Every matrix $A \in \mathbb{F}^{m \times n}$ has a REF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $B \in \mathbb{F}^{m \times n}$ which is in REF).

(b) Every matrix $A \in \mathbb{F}^{m \times n}$ has an RREF (that is, every matrix $A \in \mathbb{F}^{m \times n}$ is row equivalent to a matrix $C \in \mathbb{F}^{m \times n}$ which is in RREF). **Moreover, for each matrix $A \in \mathbb{F}^{m \times n}$, its RREF is unique.**

Theorem 2

Let A be a square matrix in $\mathbb{F}^{n \times n}$. The following are equivalent:

(I) A is invertible.

(II) Every REF of A has exactly n pivots.

(III) At least one REF of A has exactly n pivots.

(IV) The unique RREF of A is the identity matrix I_n.

How can we use the theorems? (cont.)

Recall the following matrix from HW3:

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 11 & 4 & 2 \\ 2 & -2 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Question 1. How did we check that A is invertible?

Answer. Recall that we can rely on equivalent conditions (II) and (III) from Theorem 2 to argue as follows: to show that A is invertible, by condition (III) it suffices to find a REF of A which has 3 pivots.

Moreover, because of condition (II), we don't have to be careful about which REF of A to consider; we can simply use Gaussian elimination to find one such matrix in REF, and the number of pivots of this matrix will definitively tell us if A is invertible or not.

Recall that for this specific example we can write

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 11 & 4 & 2 \\ 2 & -2 & 3 \end{pmatrix} \xrightarrow[R_3-2R_1\to R_3']{R_2-11R_1\to R_2'} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 15 & -9 \\ 0 & 0 & 1 \end{pmatrix}.$$

The last matrix is in REF, and has 3 pivots. Hence A is invertible.

How can we use the theorems? (cont.)

Recall the following matrix from HW3:

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 11 & 4 & 2 \\ 2 & -2 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Question 2. How can we find the inverse of A?

Answer. We now recall how we proved that condition (IV) of Theorem 2 is equivalent to A being invertible:

we said that, if the RREF of A is the identity matrix I_3, then we have $A \sim I_3$, and hence we can find some $k \geqslant 1$ and some elementary matrices $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k \in \mathbb{R}^{3\times 3}$ such that

$$\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1 A = I_3 \quad \Rightarrow \quad A = \mathcal{E}_1^{-1} \mathcal{E}_2^{-1} \cdots \mathcal{E}_k^{-1}.$$

This shows that A is invertible (as the product of invertible matrices).
It also shows that the inverse of A is the product $\mathcal{E}_k \cdots \mathcal{E}_2 \mathcal{E}_1$.

Question 3. How can we (efficiently) keep track of which elementary matrices we use to transform A to its RREF, and even more specifically of what their product is?

Gauss-Jordan elimination

Let us write the matrix A and the identity matrix I_3 next to each other as follows:

$$(A \mid I_3) = \left(\begin{array}{rrr|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 11 & 4 & 2 & 0 & 1 & 0 \\ 2 & -2 & 3 & 0 & 0 & 1 \end{array}\right).$$

- If we then try to apply an elementary row operation on A with the aim of moving towards a REF of A, but make sure we apply this operation on the entire matrix above, then the rightmost columns will preserve what the elementary matrix we had to use is. Indeed, applying an elementary row operation on A corresponds to multiplying A from the left by a suitable elementary matrix $\mathcal{E}_1 \in \mathbb{R}^{3\times 3}$; but then we have

$$\mathcal{E}_1(A \mid I_3) = (\mathcal{E}_1 A \mid \mathcal{E}_1 I_3) = (\mathcal{E}_1 A \mid \mathcal{E}_1).$$

- If we try to apply a second elementary row operation, or in other words multiply what we just got by a second elementary matrix $\mathcal{E}_2 \in \mathbb{R}^{3\times 3}$, then the rightmost columns will be capturing the product of the elementary matrices we have used so far:

$$\mathcal{E}_2(\mathcal{E}_1 A \mid \mathcal{E}_1) = (\mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_2\mathcal{E}_1).$$

Gauss-Jordan elimination

Let us write the matrix A and the identity matrix I_3 next to each other as follows:

$$(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 11 & 4 & 2 & 0 & 1 & 0 \\ 2 & -2 & 3 & 0 & 0 & 1 \end{array}\right).$$

- Continuing like this until we end up with the RREF of A within the leftmost columns, we can see that the rightmost columns will be capturing **the product of the elementary matrices that we needed to use, or in other words the inverse of A:**

$$(A \mid I_3) \to (\mathcal{E}_1 A \mid \mathcal{E}_1) \to (\mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_2\mathcal{E}_1) \to (\mathcal{E}_3\mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_3\mathcal{E}_2\mathcal{E}_1) \to \cdots\cdots$$
$$\to (\mathcal{E}_k \cdots \mathcal{E}_3\mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_k \cdots \mathcal{E}_3\mathcal{E}_2\mathcal{E}_1) = (\mathrm{RREF}(A) \mid \mathcal{E}_k \cdots \mathcal{E}_3\mathcal{E}_2\mathcal{E}_1) = (I_3 \mid A^{-1}),$$

  with the last equality being true if and only if $\mathrm{RREF}(A)$ equals the identity matrix, as in this case.

We usually refer to the process (series of applications of the method of Gaussian elimination) that takes us from a REF of A to the RREF of A, and also gives us the inverse of A in the cases that $\mathrm{RREF}(A) = I_n$, as the method of Gauss-Jordan elimination.

For this specific example...

we have

$$(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 11 & 4 & 2 & 0 & 1 & 0 \\ 2 & -2 & 3 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2 - 11R_1 \to R_2'} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 15 & -9 & -11 & 1 & 0 \\ 2 & -2 & 3 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_3 - 2R_1 \to R_3'} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 15 & -9 & -11 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right) \xrightarrow{R_2 + 9R_3 \to R_2'} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 15 & 0 & -29 & 1 & 9 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_1 - R_3 \to R_1'} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 3 & 0 & -1 \\ 0 & 15 & 0 & -29 & 1 & 9 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right) \xrightarrow{\frac{1}{15}R_2 \to R_2'} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 3 & 0 & -1 \\ 0 & 1 & 0 & -\frac{29}{15} & \frac{1}{15} & \frac{3}{5} \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_1 + R_2 \to R_1'} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{16}{15} & \frac{1}{15} & -\frac{2}{5} \\ 0 & 1 & 0 & -\frac{29}{15} & \frac{1}{15} & \frac{3}{5} \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right).$$

We can conclude that

$$A^{-1} = \left(\begin{array}{ccc} \frac{16}{15} & \frac{1}{15} & -\frac{2}{5} \\ -\frac{29}{15} & \frac{1}{15} & \frac{3}{5} \\ -2 & 0 & 1 \end{array}\right)$$

(and we can double check this, if we want, by directly computing that $AA^{-1} = I_3$, or equivalently that $A^{-1}A = I_3$).

A final, very important observation about linear systems

Homogeneous systems

Terminology

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ be a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with coefficients from $\mathbb{F}$ (including the constant terms b_i):

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ \ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Recall that we can equivalently write this system using matrix notation as follows: $A\bar{x} = \bar{b}$, where $A \in \mathbb{F}^{m\times n}$ is the coefficient matrix of the system, and $\bar{b}$ is the column of constant terms.

The system is called *homogeneous* if all constant terms are equal to 0, that is, if for every $i \in \{1, 2, \ldots, m\}$ we have $b_i = 0$. Equivalently if $\bar{b} = \bar{0}_{\mathbb{F}^m}$.

Useful to note. The solution set of a homogeneous system (in n unknowns) is always **non-empty** (in other words, a homogeneous system is always consistent). Indeed, the zero vector $\bar{0}_{\mathbb{F}^n}$ is always a solution to such a system, and is called *the trivial solution*.

Then, what's interesting to ask is whether there are more solutions besides the trivial solution.

Important theorem about homogeneous systems

Reminder

A system of m linear equations in n unknowns is called underdetermined if the number n of unknowns is larger than the number of non-trivial (non-zero) equations of the system.

This in particular holds true if $n > m$.

Theorem 3

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a homogeneous underdetermined system of linear equations with coefficients from $\mathbb{F}$.

Then $\mathcal{LS}_0$ has more than one solution, or in other words it has solutions besides the trivial one. In fact, its solution set is at least as big as the set $\mathbb{F}$.

Proof of Theorem 3

Observe first that, if $\mathcal{LS}_0$ has any trivial equations (that is, equations where all the coefficients are equal to 0), then if we delete them, the solution set to the system of the remaining equations is still the same as the solution set to the entire $\mathcal{LS}_0$. Therefore, without loss of generality, we can assume that the number m of equations of $\mathcal{LS}_0$ coincides with the number of its non-trivial equations, and thus that $m < n$, where n is the number of unknowns.

Let A be the coefficient matrix of $\mathcal{LS}_0$. Then $\mathcal{LS}_0 : A\bar{x} = \bar{0}$, where $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and where $\bar{0} \in \mathbb{F}^m$. Consider a REF B of A. Then the augmented matrices $(A \mid \bar{0})$ and $(B \mid \bar{0})$ are equivalent. Indeed, we know that there are $k \geqslant 1$ and some elementary matrices $\mathcal{E}_i$, $1 \leqslant i \leqslant k$, such that $B = \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A$; this then implies that

$$(B \mid \bar{0}) = (\mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A \mid \bar{0}) = (\mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1\bar{0}) = \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1(A \mid \bar{0}).$$

But this shows that the system $B\bar{x} = \bar{0}$ is equivalent to $\mathcal{LS}_0$, and thus, to analyse the desired solution set, it suffices to examine the system $B\bar{x} = \bar{0}$.

Moreover, since B is in REF, we can check that $(B \mid \bar{0})$ is in REF too. Thus the cardinality of the solution set of the corresponding system depends only on how many pivots $(B \mid \bar{0})$ has, **or equivalently on how many non-zero rows it has.** Given now that

$$\#\{\text{non-zero rows of } (B \mid \bar{0})\} \leqslant m < n,$$

we see that not all variables of the system will be pivot variables, and hence $B\bar{x} = \bar{0}$ has at least one free variable. It follows that the system has at least $|\mathbb{F}|$ different solutions.

Why understanding homogeneous linear systems is helpful

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ be the following linear system with coefficients from $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Let's also use matrix notation to describe it: $\mathcal{LS}_1 : A\bar{x} = \bar{b}$, where $A \in \mathbb{F}^{m\times n}$ is the coefficient matrix of the system, and $\bar{b}$ is the column of constant terms.

Consider also the corresponding homogeneous linear system $\mathcal{LS}_0$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & 0 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & 0 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & 0 \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & 0 \end{array}\right\},$$

or equivalently the system $A\bar{x} = \bar{0}$. Recall that $\mathcal{LS}_0$ will definitely be consistent; let S_0 be its solution set (which is a subset of $\mathbb{F}^n$, and certainly contains the zero vector $\bar{0}_{\mathbb{F}^n}$).

If $\mathcal{LS}_1$ is consistent as well, then we can find a specific solution $\bar{d}_0$ to it (note that $\bar{d}_0 \in \mathbb{F}^n$ should be such that $A\bar{d}_0 = \bar{b}$ holds).

Then the entire solution set S_1 of $\mathcal{LS}_1$ is given by $S_1 = \left\{\bar{d}_0 + \bar{u} : \bar{u} \in S_0\right\}$.

Examples where we have already seen this in practice

What have we said a line in $\mathbb{R}^2$ is? It is the set of all points (x, y) in $\mathbb{R}^2$ whose coordinates satisfy an equation of the form

$$ax + by = c$$

(that is, a linear system of 1 equation in 2 unknowns with coefficients from $\mathbb{R}$).

Recall that a vector equation for this line is of the form

$$\binom{x}{y} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1P_2}, \qquad r \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the line (and has initial point the origin), while $\overrightarrow{P_1P_2}$ is a vector parallel to the line.

Observe now that the coordinates of P_1 give a specific solution to the equation $ax + by = c$, while

$$\left\{ r \cdot \overrightarrow{P_1P_2} : r \in \mathbb{R} \right\}$$

is the solution set to the corresponding homogeneous equation $ax + by = 0$.

Other similar examples?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 39

Friday November 20

Recall: Important theorem about homogeneous systems

Reminder

A system of m linear equations in n unknowns is called underdetermined if the number n of unknowns is larger than the number of non-trivial (non-zero) equations of the system.

This in particular holds true if $n > m$.

Theorem 3

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a homogeneous underdetermined system of linear equations with coefficients from $\mathbb{F}$.

Then $\mathcal{LS}_0$ has more than one solution, or in other words it has solutions besides the trivial one. In fact, its solution set is at least as big as the set $\mathbb{F}$.

Proof of Theorem 3

Observe first that, if $\mathcal{LS}_0$ has any trivial equations (that is, equations where all the coefficients are equal to 0), then if we delete them, the solution set to the system of the remaining equations is still the same as the solution set to the entire $\mathcal{LS}_0$. Therefore, without loss of generality, we can assume that the number m of equations of $\mathcal{LS}_0$ coincides with the number of its non-trivial equations, and thus that $m < n$, where n is the number of unknowns.

Let A be the coefficient matrix of $\mathcal{LS}_0$. Then $\mathcal{LS}_0 : A\bar{x} = \bar{0}$, where $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and

where $\bar{0} \in \mathbb{F}^m$. Consider a REF B of A. Then the augmented matrices $(A \mid \bar{0})$ and $(B \mid \bar{0})$ are equivalent. Indeed, we know that there are $k \geqslant 1$ and some elementary matrices $\mathcal{E}_i$, $1 \leqslant i \leqslant k$, such that $B = \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A$; this then implies that

$$(B \mid \bar{0}) = (\mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A \mid \bar{0}) = (\mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1 A \mid \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1\bar{0}) = \mathcal{E}_k \cdots \mathcal{E}_2\mathcal{E}_1(A \mid \bar{0}).$$

But this shows that the system $B\bar{x} = \bar{0}$ is equivalent to $\mathcal{LS}_0$, and thus, to analyse the desired solution set, it suffices to examine the system $B\bar{x} = \bar{0}$.

Moreover, since B is in REF, we can check that $(B \mid \bar{0})$ is in REF too. Thus the cardinality of the solution set of the corresponding system depends only on how many pivots $(B \mid \bar{0})$ has, **or equivalently on how many non-zero rows it has.** Given now that

$$\#\{\text{non-zero rows of } (B \mid \bar{0})\} \leqslant m < n,$$

we see that not all variables of the system will be pivot variables, and hence $B\bar{x} = \bar{0}$ has at least one free variable. It follows that the system has at least $|\mathbb{F}|$ different solutions.

Why understanding homogeneous linear systems is helpful

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_1$ be the following linear system with coefficients from $\mathbb{F}$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & b_2 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & b_{m-1} \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & b_m \end{array}\right\}.$$

Let's also use matrix notation to describe it: $\mathcal{LS}_1 : A\bar{x} = \bar{b}$, where $A \in \mathbb{F}^{m\times n}$ is the coefficient matrix of the system, and $\bar{b}$ is the column of constant terms.

Consider also the corresponding homogeneous linear system $\mathcal{LS}_0$:

$$\left\{\begin{array}{rcrcccrcr} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots\cdots & + & a_{1,n}x_n & = & 0 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots\cdots & + & a_{2,n}x_n & = & 0 \\ \ddots & & \ddots & & \ddots\ddots & & \ddots & & \ddots \\ a_{m-1,1}x_1 & + & a_{m-1,2}x_2 & + & \cdots\cdots & + & a_{m-1,n}x_n & = & 0 \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots\cdots & + & a_{m,n}x_n & = & 0 \end{array}\right\},$$

or equivalently the system $A\bar{x} = \bar{0}$. Recall that $\mathcal{LS}_0$ will definitely be consistent; let S_0 be its solution set (which is a subset of $\mathbb{F}^n$, and certainly contains the zero vector $\bar{0}_{\mathbb{F}^n}$).

If $\mathcal{LS}_1$ is consistent as well, then we can find a specific solution $\bar{d}_0$ to it (note that $\bar{d}_0 \in \mathbb{F}^n$ should be such that $A\bar{d}_0 = \bar{b}$ holds).

Then the entire solution set S_1 of $\mathcal{LS}_1$ is given by $S_1 = \left\{\bar{d}_0 + \bar{u} : \bar{u} \in S_0\right\}$.

Justification

If $\mathcal{LS}_1$ is consistent as well, then we can find a specific solution $\bar{d}_0$ to it (note that $\bar{d}_0 \in \mathbb{F}^n$ should be such that $A\bar{d}_0 = \bar{b}$ holds).

Then the entire solution set S_1 of $\mathcal{LS}_1$ is given by $S_1 = \{\bar{d}_0 + \bar{u} : \bar{u} \in S_0\}$.

- Indeed, if $\bar{u}_1$ is a solution of the homogeneous system, that is, if $\bar{u}_1 \in S_0$, then we will have

$$A(\bar{d}_0 + \bar{u}_1) = A\bar{d}_0 + A\bar{u}_1 = \bar{b} + \bar{0} = \bar{b},$$

and hence $\bar{d}_0 + \bar{u}_1 \in S_1$.

- Conversely, if $\bar{d}_1$ is a solution to $\mathcal{LS}_1$, that is, if $\bar{d}_1 \in S_1$, then we will have

$$A(\bar{d}_1 - \bar{d}_0) = A\bar{d}_1 - A\bar{d}_0 = \bar{b} - \bar{b} = \bar{0},$$

and hence $\bar{u}_2 = \bar{d}_1 - \bar{d}_0$ will be a solution of the homogeneous system, or in other words we will be able to write $\bar{d}_1 = \bar{d}_0 + \bar{u}_2$ with $\bar{u}_2 \in S_0$.

Examples where we have already seen this in practice

What have we said a line in $\mathbb{R}^2$ is? It is the set of all points (x, y) in $\mathbb{R}^2$ whose coordinates satisfy an equation of the form

$$ax + by = c$$

(that is, a linear system of 1 equation in 2 unknowns with coefficients from $\mathbb{R}$*, with those coefficients satisfying* $(a, b) \neq (0, 0)$*).*

Recall that a vector equation for this line is of the form

$$\binom{x}{y} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1P_2}, \qquad r \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the line (and has initial point the origin), while $\overrightarrow{P_1P_2}$ is a vector parallel to the line.

Observe now that the coordinates of P_1 give a specific solution to the equation $ax + by = c$, while

$$\left\{ r \cdot \overrightarrow{P_1P_2} : r \in \mathbb{R} \right\}$$

is the solution set to the corresponding homogeneous equation $ax + by = 0$.

Other similar examples?

What have we said a line in $\mathbb{R}^3$ is? It is the set of all points (x, y, z) in $\mathbb{R}^3$ whose coordinates satisfy a linear system of the form

$$\left\{ \begin{array}{rcrcrcl} a_1 x & + & b_1 y & + & c_1 z & = & d_1 \\ a_2 x & + & b_2 y & + & c_2 z & = & d_2 \end{array} \right\}$$

(except in those special cases where such a system is inconsistent or has fewer than 2 pivots, and thus the solution set cannot be a line).

Recall that a vector equation for this line is of the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1 P_2}, \qquad r \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the line (and has initial point the origin), while $\overrightarrow{P_1 P_2}$ is a vector parallel to the line.

Observe now that the coordinates of P_1 give a specific solution to the given linear system, while

$$\left\{ r \cdot \overrightarrow{P_1 P_2} : r \in \mathbb{R} \right\}$$

is the solution set to the corresponding homogeneous system:

$$\left\{ \begin{array}{rcrcrcl} a_1 x & + & b_1 y & + & c_1 z & = & 0 \\ a_2 x & + & b_2 y & + & c_2 z & = & 0 \end{array} \right\}.$$

Other similar examples? (cont.)

What have we said a plane in $\mathbb{R}^3$ is? It is the set of all points (x, y, z) in $\mathbb{R}^3$ whose coordinates satisfy an equation of the form

$$ax + by + cz = d$$

(that is, a linear system of 1 equation in 3 unknowns with coefficients from $\mathbb{R}$*, with those coefficients satisfying* $(a, b, c) \neq (0, 0, 0)$*).*

Recall that a vector equation for this plane is of the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{OP_1} + \lambda \cdot \overrightarrow{P_1P_2} + \mu \cdot \overrightarrow{P_1P_3}, \qquad \lambda, \mu \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the plane (and has initial point the origin), while $\overrightarrow{P_1P_2}$ and $\overrightarrow{P_1P_3}$ are vectors parallel to the plane, which are not parallel to each other.

Observe now that the coordinates of P_1 give a specific solution to the equation $ax + by + cz = d$, while

$$\left\{\lambda \cdot \overrightarrow{P_1P_2} + \mu \cdot \overrightarrow{P_1P_3} : \lambda, \mu \in \mathbb{R}\right\}$$

is the solution set to the homogeneous equation $ax + by + cz = 0$.

Next, we look at other important constructions / notions that can be associated with matrices.

Important subspaces associated with a given matrix

Definition

Let $\mathbb{F}$ be a field, and consider a matrix $A \in \mathbb{F}^{m \times n}$. Observe that each row of A has n components, that is, each row of A is a vector in $\mathbb{F}^n$. Similarly, each column of A has m components, thus it is a vector in $\mathbb{F}^m$.

1. We define the *Row Space* of A to be the linear span of all rows of A, and we denote it by $RS(A)$.

   Observe that $RS(A)$ is a subspace of $\mathbb{F}^n$.

2. We define the *Column Space* of A to be the linear span of all columns of A, and we denote it by $CS(A)$.

   Observe that $CS(A)$ is a subspace of $\mathbb{F}^m$.

3. The *Nullspace* of A is defined to be the solution set to the system $A\bar{x} = \bar{0}$. We denote it by $N(A)$.

   In other words, $N(A) = \left\{ \bar{u} \in \mathbb{F}^n : A\bar{u} = \bar{0}_{\mathbb{F}^m} \right\}$.

Important subspaces associated with a given matrix

Clearly $RS(A)$ and $CS(A)$ are subspaces of the corresponding vector spaces (since we have seen that linear spans are subspaces). What about $N(A)$?

Theorem

Given a matrix $A \in \mathbb{F}^{m \times n}$, the Nullspace $N(A)$ of A is a subspace of $\mathbb{F}^n$.

Proof. Recall that we need to check that

(i) $\bar{0}_{\mathbb{F}^n} \in N(A)$;

(ii) $N(A)$ is closed under vector addition;

(iii) $N(A)$ is closed under scalar multiplication.

Regarding (i), we can immediately see that $A\bar{0}_{\mathbb{F}^n} = \bar{0}_{\mathbb{F}^m}$, and thus $\bar{0}_{\mathbb{F}^n} \in N(A)$.

Regarding (ii), consider $\bar{u}_1, \bar{u}_2 \in N(A)$. Then we have $A\bar{u}_1 = A\bar{u}_2 = \bar{0}_{\mathbb{F}^m}$, and hence

$$A(\bar{u}_1 + \bar{u}_2) = A\bar{u}_1 + A\bar{u}_2 = \bar{0}_{\mathbb{F}^m} + \bar{0}_{\mathbb{F}^m} = \bar{0}_{\mathbb{F}^m}.$$

We conclude that $\bar{u}_1 + \bar{u}_2 \in N(A)$ too.

Finally, regarding (iii), consider again $\bar{u}_1 \in N(A)$, as well as $r \in \mathbb{F}$. Then, we have $A\bar{u}_1 = \bar{0}_{\mathbb{F}^m}$, and hence

$$A(r\bar{u}_1) = r\big(A\bar{u}_1\big) = r\bar{0}_{\mathbb{F}^m} = \bar{0}_{\mathbb{F}^m}.$$

We conclude that $r\bar{u}_1 \in N(A)$ too.

Recall once more...

What have we said a line in $\mathbb{R}^2$ is? It is the set of all points (x, y) in $\mathbb{R}^2$ whose coordinates satisfy an equation of the form

$$ax + by = c$$

(that is, a linear system of 1 equation in 2 unknowns with coefficients from $\mathbb{R}$*, with those coefficients satisfying* $(a, b) \neq (0, 0)$*).*

Recall that a vector equation for this line is of the form

$$\binom{x}{y} = \overrightarrow{OP_1} + r \cdot \overrightarrow{P_1P_2}, \qquad r \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the line (and has initial point the origin), while $\overrightarrow{P_1P_2}$ is a vector parallel to the line.

Observe now that the coordinates of P_1 give a specific solution to the equation $ax + by = c$, while

$$\left\{ r \cdot \overrightarrow{P_1P_2} : r \in \mathbb{R} \right\}$$

is the solution set to the corresponding homogeneous equation $ax + by = 0$.

Thus, a line in $\mathbb{R}^2$ (according to the geometric meaning of the notion) is the translate of an appropriate **nullspace / subspace.**

Recall...

What have we said a plane in $\mathbb{R}^3$ is? It is the set of all points (x, y, z) in $\mathbb{R}^3$ whose coordinates satisfy an equation of the form

$$ax + by + cz = d$$

(that is, a linear system of 1 equation in 3 unknowns with coefficients from $\mathbb{R}$*, with those coefficients satisfying* $(a, b, c) \neq (0, 0, 0)$*).*

Recall that a vector equation for this plane is of the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{OP_1} + \lambda \cdot \overrightarrow{P_1P_2} + \mu \cdot \overrightarrow{P_1P_3}, \qquad \lambda, \mu \in \mathbb{R},$$

where $\overrightarrow{OP_1}$ is a vector whose terminal point is found on the plane (and has initial point the origin), while $\overrightarrow{P_1P_2}$ and $\overrightarrow{P_1P_3}$ are vectors parallel to the plane, which are not parallel to each other.

Observe now that the coordinates of P_1 give a specific solution to the equation $ax + by + cz = d$, while

$$\left\{\lambda \cdot \overrightarrow{P_1P_2} + \mu \cdot \overrightarrow{P_1P_3} : \lambda, \mu \in \mathbb{R}\right\}$$

is the solution set to the homogeneous equation $ax + by + cz = 0$.

Thus, a plane in $\mathbb{R}^3$ (according to the geometric meaning of the notion) is the translate of an appropriate **nullspace / subspace.**

What about functions in this 'universe'?

Since we are not working with simple sets, but rather with sets that have an associated structure, the most important functions to consider are functions that respect these structures.

Functions that preserve / 'respect' a vector space structure

Definition

Let $\mathbb{F}$ be a field, and let $V_1 = \big(\{\text{elements in } V_1\}, +_1, \cdot_1\big)$ and $V_2 = \big(\{\text{elements in } V_2\}, +_2, \cdot_2\big)$ be two vector spaces **over $\mathbb{F}$.**

A function $f : V_1 \to V_2$ is called a *linear function*, or a *linear map*, or a *vector space homomorphism*, if f satisfies the following two properties:

1. for every $\bar{x}, \bar{y} \in V_1$, we have that
$$f(\bar{x} +_1 \bar{y}) = f(\bar{x}) +_2 f(\bar{y}).$$
2. for every $\lambda \in \mathbb{F}$ and $\bar{x} \in V_1$, we have that
$$f(\lambda \cdot_1 \bar{x}) = \lambda \cdot_2 f(\bar{x}).$$

Question. Is requiring these two properties enough in order to conclude that f respects the vector space structure? And what should we understand by this?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 40

Monday November 23

Recall: Functions that preserve / 'respect' a vector space structure

Definition

Let $\mathbb{F}$ be a field, and let $V_1 = \big(\{\text{elements in } V_1\}, +_1, \cdot_1\big)$ and $V_2 = \big(\{\text{elements in } V_2\}, +_2, \cdot_2\big)$ be two vector spaces **over $\mathbb{F}$.**

A function $f : V_1 \to V_2$ is called a *linear function*, or a *linear map*, or a *vector space homomorphism*, if f satisfies the following two properties:

1. for every $\bar{x}, \bar{y} \in V_1$, we have that
$$f(\bar{x} +_1 \bar{y}) = f(\bar{x}) +_2 f(\bar{y}).$$
2. for every $\lambda \in \mathbb{F}$ and $\bar{x} \in V_1$, we have that
$$f(\lambda \cdot_1 \bar{x}) = \lambda \cdot_2 f(\bar{x}).$$

Question. Is requiring these two properties enough in order to conclude that f respects the vector space structure? And what should we understand by this?

What are the absolutely necessary concepts / elements / properties of a vector space structure?

We recall that, to be able to have the structure of a vector space, we need

- an operation of vector addition,
- an operation of scalar multiplication,
- a zero vector (that is, a vector which acts as the neutral element of vector addition),
- the existence of additive inverses

and of course some additional properties that govern how these operations will behave.

We have already seen that a linear function is required by its definition to respect the operations of vector addition and of scalar multiplication. We will now see that it also preserves the zero vector and the additive inverses: in other words,

- $f(\bar{0}_{V_1}) = \bar{0}_{V_2}$.
- for every $\bar{x} \in V_1$, we have that $f(-\bar{x}) = -f(\bar{x})$.

Verifying the additional properties of a linear function

— We need to show that $f(\bar{0}_{V_1}) = \bar{0}_{V_2}$. Indeed, we can write

$$f(\bar{0}_{V_1}) = f\left(0_{\mathbb{F}} \cdot \bar{0}_{V_1}\right) = 0_{\mathbb{F}} \cdot f(\bar{0}_{V_1}) = \bar{0}_{V_2},$$

where the second equality follows from the definition of a linear function, which requires that the linear function respects scalar multiplication.

— We need to show that, for every $\bar{x} \in V_1$, we have that $f(-\bar{x}) = -f(\bar{x})$ *(in other words, the image of the additive inverse of $\bar{x}$ is the additive inverse of the image of $\bar{x}$)*.

1st way. We recall that, for every vector $\bar{z}$, $-\bar{z} = (-1_{\mathbb{F}}) \cdot \bar{z}$, and thus

$$f(-\bar{x}) = f\left((-1_{\mathbb{F}}) \cdot \bar{x}\right) = (-1_{\mathbb{F}}) \cdot f(\bar{x}) = -f(\bar{x}).$$

2nd way. By the linearity of f, which requires that f respects vector addition, and also by what we just showed above, we have that

$$\bar{0}_{V_2} = f(\bar{0}_{V_1}) = f\left(\bar{x} + (-\bar{x})\right) = f(\bar{x}) + f(-\bar{x}).$$

Clearly, we also have that

$$\bar{0}_{V_2} = f(\bar{x}) + (-f(\bar{x})).$$

Thus

$$f(\bar{x}) + f(-\bar{x}) = f(\bar{x}) + (-f(\bar{x})) \quad \Rightarrow \quad f(-\bar{x}) = -f(\bar{x}).$$

Important notions related with functions

Formal Definition

— Let A, B be two sets (not necessarily vector spaces, or any other specific structure). A function $f : A \to B$ is *formally* a subset of $A \times B$ such that

- for every $a \in A$, there is exactly one $b \in B$ such that $(a, b) \in f$.

Most commonly, we express this by writing $b = f(a)$ (and we say that b is the *image* of a under f, and that a is a *preimage* of b).

— The set A is called the *domain* of f, while the set B is called the *codomain* of f. Sometimes we will write $\mathrm{Dom}(f)$ for the domain of f.
Moreover, the subset

$$\{b \in B : \exists\, a \in A \text{ such that } b = f(a)\} \;=\; \{f(a) : a \in A\}$$

of B is called the *range* of f. We will write $\mathrm{Range}(f)$ for it.

Important notions related with functions

Definitions

— Let A, B be two sets (not necessarily vector spaces, or any other specific structure). A function $f : A \to B$ is *formally* a subset of $A \times B$ such that

- for every $a \in A$, there is exactly one $b \in B$ such that $(a, b) \in f$.

Most commonly, we express this by writing $b = f(a)$.

— A function $g : A \to B$ is called *injective* or an *injection* or *1-1* (with the latter being read 'one-to-one') if

- for every $a_1, a_2 \in A$ we have that, if $g(a_1) = g(a_2)$, then $a_1 = a_2$,
  - or equivalently, if, whenever $a_1 \neq a_2$, we have that $g(a_1) \neq g(a_2)$.

  In other words, g is an injective function if
  - for every $b \in B$, there is at most one $a \in A$ such that $(a, b) \in g$.

— A function $h : A \to B$ is called *surjective* or a *surjection* or *onto* if

- for every $b \in B$, there exists $a \in A$ such that $b = f(a)$.

In other words, if $\mathrm{Range}(h) = B$.

— A function $u : A \to B$ is called *bijective* or a *bijection* if u is both injective and surjective (in other words, if u is 1-1 **and** onto).

Important notions related with functions

Definitions

— Let A, B, C be three sets. Consider also a subset B_0 of B, and suppose we have functions $f : A \to B$ and $g : B_0 \to C$.

If $\mathbf{Range}(f) \subseteq B_0$, then we define the *composition* of g with f as follows:

$$g \circ f \text{ is a function from } A \text{ to } C \text{ such that,}$$

$$\text{for every } a \in A, \quad (g \circ f)(a) := g(f(a)) = \text{the image of } f(a) \text{ under } g.$$

Remark. We have that

$$\begin{aligned} \mathrm{Range}(g \circ f) &= \{g(f(a)) : a \in A\} \\ &= \{g(b) : b \in \mathrm{Range}(f)\} \\ &\subseteq \{g(b') : b' \in B_0\} \\ &= \mathrm{Range}(g). \end{aligned}$$

Examples and non-examples of compositions

- Recall that the function $\cos : \mathbb{R} \to \mathbb{R}$ takes values in $[-1, 1]$, that is, $\mathrm{Range}(\cos) = [-1, 1]$.

  Now, if $g_1(x) = \frac{1}{x-2}$, then its (natural) domain is the set $\mathbb{R} \setminus \{2\}$ (that is, $\mathrm{Dom}\big(1/(x-2)\big) = \mathbb{R} \setminus \{2\}$), and thus we can consider the composition of g_1 with $\cos$: we have that $g_1 \circ \cos : \mathbb{R} \to \mathbb{R}$ and

$$(g_1 \circ \cos)(x) = \frac{1}{\cos(x) - 2}$$

  for each $x \in \mathbb{R}$.

- On the other hand, if $g_2(x) = \frac{x^2}{x+1}$, then the (natural) domain of g_2 is the set $\mathbb{R} \setminus \{-1\}$, and thus $g_2 \circ \cos$ is **not** defined.

- Note that, on the other hand, both $\cos \circ g_1$ and $\cos \circ g_2$ are defined, since $\mathrm{Dom}(\cos) = \mathbb{R} \supseteq \mathrm{Range}(g_i)$ for both $i = 1$ and $i = 2$.

Important notions related with functions

Definitions

— Given a set A, we can define the identity function id_A on A to be the function $\mathrm{id}_A : A \to A$ which satisfies $\mathrm{id}_A(a) = a$ for every $a \in A$.

— Suppose that B is a second set, and that $f : A \to B$ is a **bijective** function. We have already seen that this means that

$$\text{for every } b \in B, \text{ there is a unique } a \in A \text{ such that } b = f(a).$$

Thus, we could pair each $b \in B$ with its unique preimage in A, and in this way we could obtain a function $g : B \to A$.

Note now that this function satisfies the following

$$\text{(i) } g \circ f = \mathrm{id}_A \qquad \text{and} \qquad \text{(ii) } f \circ g = \mathrm{id}_B.$$

— Given a function $f : A \to B$, if there exists a function $g : B \to A$ which satisifes (i) and (ii), then we say that this function g is the *inverse of f*, and we usually denote it by f^{-1}. **Note that, if such a function g exists, then it is unique.**

Remark

Observe that we have just explained that, if a function $f : A \to B$ is bijective, then it does have an inverse.

It can be checked that the converse is also true: that is, if $h : A \to B$ has an inverse, then h is a bijection from A to B.

Examples and non-examples of invertible functions

- The function $\cos : \mathbb{R} \to \mathbb{R}$ is not bijective (in fact, it is neither injective nor surjective). Thus, it cannot have an inverse.
- On the other hand, the function $\cos : [0, \pi] \to [-1, 1]$ is bijective, so it does have an inverse function

$$\cos^{-1} : [-1, 1] \to [0, \pi]$$

(most commonly written as arccos*).*

- The conjugate function $\psi : \mathbb{C} \to \mathbb{C}$, which maps each $z \in \mathbb{C}$ to its conjugate $\overline{z} = \mathrm{Re}(z) - \mathrm{Im}(z)\,\boldsymbol{i}$, has an inverse function: in fact, it is an inverse of itself, $\psi \circ \psi = \mathrm{id}_{\mathbb{C}}$.

Back to linear maps

Question 1. Is the function $f_1 : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$f_1\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 \\ 2x_1^2 - 2x_2^2 \end{pmatrix}$$

linear? Justify your answer.

Question 2. Is the function $f_2 : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$f_2\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} 3x_1 + 2x_2 + x_3 \\ x_1 - 2x_3 \\ 2x_1 - 2x_3 \end{pmatrix}$$

linear? Justify your answer.

Relating with concepts from previous chapters

Past Homework Problem. Explain why (a) through (d) are the same problem essentially, and choose whichever way you want to solve the problem.

(a) Solve the following system of linear equations with coefficients from $\mathbb{R}$:

$$\left\{\begin{array}{rcl} 3x_1 + 2x_2 & +x_3 = & 9 \\ x_1 & -2x_3 = & -8 \\ 2x_1 & -2x_3 = & -6 \end{array}\right\}.$$

(b) Consider the vectors $\bar{u} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$, $\bar{v} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$, $\bar{w} = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix}$, and $\bar{b} = \begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix}$ in $\mathbb{R}^3$.

Determine whether $\bar{b}$ is in the linear span of $\bar{u}$, $\bar{v}$, and $\bar{w}$. If yes, write explicitly the linear combination(s) showing this.

(c) Let

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & -2 \\ 2 & 0 & -2 \end{pmatrix} \in \mathbb{R}^{3\times 3} \quad \text{and} \quad \bar{b} = \begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix} \in \mathbb{R}^3.$$

Solve the equation $A\vec{x} = \bar{b}$ for the unknown vector $\vec{x} \in \mathbb{R}^3$.

(d) Define $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} 3x_1+2x_2+x_3 \\ x_1-2x_3 \\ 2x_1-2x_3 \end{pmatrix}$. Determine whether $\begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix}$ is in the range of f, and, if it is, find all its preimages.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 41

Tuesday November 24

Recall: Functions that preserve / 'respect' a vector space structure

Definition

Let $\mathbb{F}$ be a field, and let $V_1 = \big(\{\text{elements in } V_1\}, +_1, \cdot_1\big)$ and $V_2 = \big(\{\text{elements in } V_2\}, +_2, \cdot_2\big)$ be two vector spaces **over $\mathbb{F}$.**

A function $f : V_1 \to V_2$ is called a *linear function*, or a *linear map*, or a *vector space homomorphism*, if f satisfies the following two properties:

1. for every $\bar{x}, \bar{y} \in V_1$, we have that
$$f(\bar{x} +_1 \bar{y}) = f(\bar{x}) +_2 f(\bar{y}).$$
2. for every $\lambda \in \mathbb{F}$ and $\bar{x} \in V_1$, we have that
$$f(\lambda \cdot_1 \bar{x}) = \lambda \cdot_2 f(\bar{x}).$$

Question. Is requiring these two properties enough in order to conclude that f respects the vector space structure? And what should we understand by this?

What are the absolutely necessary concepts / elements / properties of a vector space structure?

We recall that, to be able to have the structure of a vector space, we need

- an operation of vector addition,
- an operation of scalar multiplication,
- a zero vector (that is, a vector which acts as the neutral element of vector addition),
- the existence of additive inverses

and of course some additional properties that govern how these operations will behave.

We have already seen that a linear function is required by its definition to respect the operations of vector addition and of scalar multiplication. We will now see that it also preserves the zero vector and the additive inverses: in other words,

- $f(\bar{0}_{V_1}) = \bar{0}_{V_2}$.
- for every $\bar{x} \in V_1$, we have that $f(-\bar{x}) = -f(\bar{x})$.

Verifying the additional properties of a linear function

— We need to show that $f(\bar{0}_{V_1}) = \bar{0}_{V_2}$. Indeed, we can write

$$f(\bar{0}_{V_1}) = f\left(0_{\mathbb{F}} \cdot \bar{0}_{V_1}\right) = 0_{\mathbb{F}} \cdot f(\bar{0}_{V_1}) = \bar{0}_{V_2},$$

where the second equality follows from the definition of a linear function, which requires that the linear function respects scalar multiplication.

— We need to show that, for every $\bar{x} \in V_1$, we have that $f(-\bar{x}) = -f(\bar{x})$ *(in other words, the image of the additive inverse of $\bar{x}$ is the additive inverse of the image of $\bar{x}$)*.

1st way. We recall that, for every vector $\bar{z}$, $-\bar{z} = (-1_{\mathbb{F}}) \cdot \bar{z}$, and thus

$$f(-\bar{x}) = f\left((-1_{\mathbb{F}}) \cdot \bar{x}\right) = (-1_{\mathbb{F}}) \cdot f(\bar{x}) = -f(\bar{x}).$$

2nd way. By the linearity of f, which requires that f respects vector addition, and also by what we just showed above, we have that

$$\bar{0}_{V_2} = f(\bar{0}_{V_1}) = f\left(\bar{x} + (-\bar{x})\right) = f(\bar{x}) + f(-\bar{x}).$$

Clearly, we also have that

$$\bar{0}_{V_2} = f(\bar{x}) + (-f(\bar{x})).$$

Thus

$$f(\bar{x}) + f(-\bar{x}) = f(\bar{x}) + (-f(\bar{x})) \quad \Rightarrow \quad f(-\bar{x}) = -f(\bar{x}).$$

Some examples

Question 1. Is the function $f_1 : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$f_1\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 \\ 2x_1^2 - 2x_2^2 \end{pmatrix}$$

linear? Justify your answer.

Question 2. Is the function $f_2 : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$f_2\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} 3x_1 + 2x_2 + x_3 \\ x_1 - 2x_3 \\ 2x_1 - 2x_3 \end{pmatrix}$$

linear? Justify your answer.

Relating with concepts from previous chapters

Past Homework Problem. Explain why (a) through (d) are the same problem essentially, and choose whichever way you want to solve the problem.

(a) Solve the following system of linear equations with coefficients from $\mathbb{R}$:

$$\left\{\begin{array}{rcl} 3x_1 + 2x_2 & +x_3 = & 9 \\ x_1 & -2x_3 = & -8 \\ 2x_1 & -2x_3 = & -6 \end{array}\right\}.$$

(b) Consider the vectors $\bar{u} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$, $\bar{v} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$, $\bar{w} = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix}$, and $\bar{b} = \begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix}$ in $\mathbb{R}^3$.

Determine whether $\bar{b}$ is in the linear span of $\bar{u}$, $\bar{v}$, and $\bar{w}$. If yes, write explicitly the linear combination(s) showing this.

(c) Let

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & -2 \\ 2 & 0 & -2 \end{pmatrix} \in \mathbb{R}^{3\times 3} \quad \text{and} \quad \bar{b} = \begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix} \in \mathbb{R}^3.$$

Solve the equation $A\vec{x} = \bar{b}$ for the unknown vector $\vec{x} \in \mathbb{R}^3$.

(d) Define $f : \mathbb{R}^3 \to \mathbb{R}^3$ by $f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} 3x_1+2x_2+x_3 \\ x_1-2x_3 \\ 2x_1-2x_3 \end{pmatrix}$. Determine whether $\begin{pmatrix} 9 \\ -8 \\ -6 \end{pmatrix}$ is in the range of f, and, if it is, find all its preimages.

Important subspaces associated with a linear function

Let $\mathbb{F}$ be a field, let V_1, V_2 be two vector spaces over $\mathbb{F}$, and consider a linear function $g : V_1 \to V_2$.

Note that we don't necessarily have that $\mathrm{Range}(g) = V_2$; this would be true only if g is surjective. However...

Theorem 1

Given a linear function $g : V_1 \to V_2$, its range is always a subspace of V_2, that is, $\mathrm{Range}(g) \leqslant V_2$.

Proof. We need to check that

(i) $\bar{0}_{V_2} \in \mathrm{Range}(g)$;

(ii) $\mathrm{Range}(g)$ is closed under vector addition;

(iii) $\mathrm{Range}(g)$ is closed under scalar multiplication.

Regarding (i), we have already seen that $g(\bar{0}_{V_1}) = \bar{0}_{V_2}$ since g is linear, and thus $\bar{0}_{V_2} \in \mathrm{Range}(g)$.

Continuing the proof of Theorem 1

Regarding (ii), consider $\bar{w}_1, \bar{w}_2 \in \mathrm{Range}(g)$. Then we can find $\bar{u}_1 \in V_1$ so that $\bar{w}_1 = g(\bar{u}_1)$. Similarly, we can find $\bar{u}_2 \in V_1$ so that $\bar{w}_2 = g(\bar{u}_2)$.

But then

$$\bar{w}_1 + \bar{w}_2 = g(\bar{u}_1) + g(\bar{u}_2) = g(\bar{u}_1 + \bar{u}_2)$$

since g is linear. This shows that $\bar{w}_1 + \bar{w}_2 \in \mathrm{Range}(g)$, as it is the image under g of the vector $\bar{u}_1 + \bar{u}_2$.

Finally, regarding (iii), consider $\bar{w}_1 \in \mathrm{Range}(g)$ and $r \in \mathbb{F}$. Then we can find $\bar{u}_1 \in V_1$ so that $\bar{w}_1 = g(\bar{u}_1)$; we will then be able to write

$$r \cdot \bar{w}_1 = r \cdot g(\bar{u}_1) = g\big(r \cdot \bar{u}_1\big)$$

since g is linear. This shows that $r \cdot \bar{w}_1 \in \mathrm{Range}(g)$, as it is the image under g of the vector $r \cdot \bar{u}_1$.

One more subspace associated with a linear function

Definition

Let $\mathbb{F}$ be a field, let V_1, V_2 be two vector spaces over $\mathbb{F}$, and let $g : V_1 \to V_2$ be a linear function.

The set $\{\bar{x} \in V_1 : g(\bar{x}) = \bar{0}_{V_2}\}$ (that is, the set of preimages of $\bar{0}_{V_2}$ under g, which sometimes is also denoted by $g^{-1}(\{\bar{0}_{V_2}\})$) is called the *Kernel* of g. We write $\mathrm{Ker}(g)$ for it.

Theorem 2

Given a linear function $g : V_1 \to V_2$, its kernel is always a subspace of V_1, that is, $\mathrm{Ker}(g) \leqslant V_1$.

Proof. We need to check that

(i) $\bar{0}_{V_1} \in \mathrm{Ker}(g)$;

(ii) $\mathrm{Ker}(g)$ is closed under vector addition;

(iii) $\mathrm{Ker}(g)$ is closed under scalar multiplication.

Regarding (i), we have already seen that $g(\bar{0}_{V_1}) = \bar{0}_{V_2}$, and thus $\bar{0}_{V_1} \in \mathrm{Ker}(g)$.

Continuing the proof of Theorem 2

Regarding (ii), consider $\bar{v}_1, \bar{v}_2 \in \mathrm{Ker}(g)$. Then we know that $g(\bar{v}_1) = \bar{0}_{V_2} = g(\bar{v}_2)$, and hence

$$\begin{aligned} g(\bar{v}_1 + \bar{v}_2) &= g(\bar{v}_1) + g(\bar{v}_2) && \text{(since } g \text{ is linear)} \\ &= \bar{0}_{V_2} + \bar{0}_{V_2} = \bar{0}_{V_2}. \end{aligned}$$

This shows that $\bar{v}_1 + \bar{v}_2 \in \mathrm{Ker}(g)$ too.

Finally, regarding (iii), consider $\bar{v}_1 \in \mathrm{Range}(g)$ and $r \in \mathbb{F}$. Then, we will again have $g(\bar{v}_1) = \bar{0}_{V_2}$. But then

$$\begin{aligned} g\big(r \cdot \bar{v}_1\big) &= r \cdot g(\bar{v}_1) && \text{(since } g \text{ is linear)} \\ &= r \cdot \bar{0}_{V_2} = \bar{0}_{V_2}. \end{aligned}$$

This shows that $r \cdot \bar{v}_1 \in \mathrm{Ker}(g)$ too.

An important observation

Proposition 3

Let $\mathbb{F}$ be a field, let V_1, V_2 be two vector spaces over $\mathbb{F}$, and let $g : V_1 \to V_2$ be a **linear** function.

Then g is injective **if and only if** $\mathrm{Ker}(g) = \{\bar{0}_{V_1}\}$ (that is, **if and only if** we ensure that $\bar{0}_{V_2}$ has a unique preimage).

Proof. *See next lecture.*

Linear function defined via a matrix

We have already, earlier in the term, come very close to defining linear functions between vector spaces of the form $\mathbb{F}^s$, where $\mathbb{F}$ is some field and $s \geqslant 1$ is an integer.

Indeed, if $A \in \mathbb{F}^{m \times n}$ is a matrix with entries from $\mathbb{F}$, then the operation $A\bar{u}$ makes sense for every vector $\bar{u} \in \mathbb{F}^n$.

Question. What is the output of this operation for a given vector $\bar{u}_0 \in \mathbb{F}^n$?

Answer. It is a vector with m components from $\mathbb{F}$, that is, a vector in $\mathbb{F}^m$.

In other words, the rule $\bar{u} \mapsto A\bar{u}$ gives a function from $\mathbb{F}^n$ to $\mathbb{F}^m$.

Useful Observation

Such a function is a linear function from $\mathbb{F}^n$ to $\mathbb{F}^m$.

Verification. *See next lecture.*

What about the reverse?
Matrix representations of linear functions

It shouldn't be hard to convince ourselves that the converse to the above observation would be very useful (given that being able to have a 'formula' for a linear function given via a matrix would allow us more 'concrete' access to what this function does). Thus it makes sense to ask the following

Question. Given a linear function $f : \mathbb{F}^n \to \mathbb{F}^m$, can we find a matrix $A = A_f \in \mathbb{F}^{m \times n}$ such that,

$$\text{for all } \bar{u} \in \mathbb{F}^n, \text{ we will have } f(\bar{u}) = A\bar{u}\,?$$

A matrix A satisfying this would be called a *matrix representation* of f.

How can we make a 'good' guess about what the entries of such a matrix A should be?

Applying this to an example

Find a matrix representation for the function $f_3 : \mathbb{R}^4 \to \mathbb{R}^5$ given by

$$f_3\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\right) = \begin{pmatrix} 2x_2 + x_3 + x_4 \\ x_1 - 2x_3 \\ 2x_1 - 2x_3 \\ x_1 - 2x_2 - 3x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 \end{pmatrix}.$$

Interesting question

Recall that we have said that, given a matrix $A \in \mathbb{F}^{m \times n}$, there are subspaces of $\mathbb{F}^m$ and of $\mathbb{F}^n$ that we can naturally associate with A.

1. The *Row Space* of A is the linear span of all rows of A. We denote it by $RS(A)$.

   Observe that $RS(A)$ is a subspace of $\mathbb{F}^n$.

2. The *Column Space* of A is the linear span of all columns of A. We denote it by $CS(A)$.

   Observe that $CS(A)$ is a subspace of $\mathbb{F}^m$.

3. The *Nullspace* of A is defined to be the solution set to the system $A\bar{x} = \bar{0}$. We denote it by $N(A)$.

   In other words, $N(A) = \left\{ \bar{u} \in \mathbb{F}^n : A\bar{u} = \bar{0}_{\mathbb{F}^m} \right\}$.

Question. If we now consider the linear function $f : \mathbb{F}^n \to \mathbb{F}^m$ which has matrix representation A, and we also recall the important subspaces associated with this function f, that is, the subspaces $\mathrm{Range}(f)$ and $\mathrm{Ker}(f)$, what is the relation between these subspaces and the above subspaces? Are there any connections / identifications we can make?

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 42

Wednesday November 25

Recall: Linear Functions

Definition

Let $\mathbb{F}$ be a field, and let $V_1 = \big(\{\text{elements in } V_1\}, +_1, \cdot_1\big)$ and $V_2 = \big(\{\text{elements in } V_2\}, +_2, \cdot_2\big)$ be two vector spaces **over $\mathbb{F}$.**

A function $f : V_1 \to V_2$ is called a *linear function*, or a *linear map*, or a *vector space homomorphism*, if f satisfies the following two properties:

1. for every $\bar{x}, \bar{y} \in V_1$, we have that
$$f(\bar{x} +_1 \bar{y}) = f(\bar{x}) +_2 f(\bar{y}).$$
2. for every $\lambda \in \mathbb{F}$ and $\bar{x} \in V_1$, we have that
$$f(\lambda \cdot_1 \bar{x}) = \lambda \cdot_2 f(\bar{x}).$$

Important Remark

If $f : V_1 \to V_2$ is a linear function, then we also have

- $f(\bar{0}_{V_1}) = \bar{0}_{V_2}$,
- and $f(-\bar{x}) = -f(\bar{x})$ for every $\bar{x} \in V_1$.

Important subspaces associated with a linear function

Let $\mathbb{F}$ be a field, let V_1, V_2 be two vector spaces over $\mathbb{F}$, and consider a linear function $g : V_1 \to V_2$.

Note that we don't necessarily have that $\mathrm{Range}(g) = V_2$; this would be true only if g is surjective. However...

Theorem 1

Given a linear function $g : V_1 \to V_2$, its range is always a subspace of V_2, that is, $\mathrm{Range}(g) \leqslant V_2$.

Furthermore...

Definition

The set $\{\bar{x} \in V_1 : g(\bar{x}) = \bar{0}_{V_2}\}$ (that is, the set of preimages of $\bar{0}_{V_2}$ under g, which sometimes is also denoted by $g^{-1}(\{\bar{0}_{V_2}\})$) is called the *Kernel* of g. We write $\mathrm{Ker}(g)$ for it.

Theorem 2

Given a linear function $g : V_1 \to V_2$, its kernel is always a subspace of V_1, that is, $\mathrm{Ker}(g) \leqslant V_1$.

An important observation

Proposition 3

Let $\mathbb{F}$ be a field, let V_1, V_2 be two vector spaces over $\mathbb{F}$, and let $g : V_1 \to V_2$ be a **linear** function.

Then g is injective **if and only if** $\mathrm{Ker}(g) = \{\bar{0}_{V_1}\}$ (that is, **if and only if** we ensure that $\bar{0}_{V_2}$ has a unique preimage).

Proof. Suppose first that g is injective. Then, if we have $\bar{u} \in \mathrm{Ker}(g)$, we can write

$$g(\bar{u}) = \bar{0}_{V_2} = g(\bar{0}_{V_1}),$$

which implies that $\bar{u} = \bar{0}_{V_1}$ since g is injective. Thus, $\mathrm{Ker}(g)$ contains only one element, the zero vector $\bar{0}_{V_1}$.

Assume now that $\mathrm{Ker}(g) = \{\bar{0}_{V_1}\}$. Consider $\bar{u}_1, \bar{u}_2 \in V_1$ such that $g(\bar{u}_1) = g(\bar{u}_2)$; we need to show that $\bar{u}_1 = \bar{u}_2$.

We have seen that, because of the linearity of g, $-g(\bar{u}_2) = g(-\bar{u}_2)$. Thus we can write

$$\bar{0}_{V_2} = g(\bar{u}_1) - g(\bar{u}_2) = g(\bar{u}_1) + g(-\bar{u}_2) = g(\bar{u}_1 - \bar{u}_2).$$

This shows that $\bar{u}_1 - \bar{u}_2 \in \mathrm{Ker}(g) = \{\bar{0}_{V_1}\} \Rightarrow \bar{u}_1 - \bar{u}_2 = \bar{0}_{V_1} \Rightarrow \bar{u}_1 = \bar{u}_2$, as we wanted.

Given that $\bar{u}_1, \bar{u}_2 \in V_1$ were arbitrary elements satisfying $g(\bar{u}_1) = g(\bar{u}_2)$, we conclude that g is injective.

Linear function defined via a matrix

We have already, earlier in the term, come very close to defining linear functions between vector spaces of the form $\mathbb{F}^s$, where $\mathbb{F}$ is some field and $s \geqslant 1$ is an integer.

Indeed, if $A \in \mathbb{F}^{m \times n}$ is a matrix with entries from $\mathbb{F}$, then the operation $A\bar{u}$ makes sense for every vector $\bar{u} \in \mathbb{F}^n$.

Question. What is the output of this operation for a given vector $\bar{u}_0 \in \mathbb{F}^n$?

Answer. It is a vector with m components from $\mathbb{F}$, that is, a vector in $\mathbb{F}^m$.

In other words, the rule $\bar{u} \mapsto A\bar{u}$ gives a function from $\mathbb{F}^n$ to $\mathbb{F}^m$.

Useful Observation

Such a function is a linear function from $\mathbb{F}^n$ to $\mathbb{F}^m$.

Verification. Let's consider a matrix $A \in \mathbb{F}^{m \times n}$, and let us denote by f_A the function that is defined according to the above rule.
Then, for every two vectors $\bar{u}_1, \bar{u}_2 \in \mathbb{F}^n$, we have that

$$f_A(\bar{u}_1 + \bar{u}_2) = A(\bar{u}_1 + \bar{u}_2) = A\bar{u}_1 + A\bar{u}_2 = f_A(\bar{u}_1) + f_A(\bar{u}_2).$$

Similarly, for every vector $\bar{u}_1 \in \mathbb{F}^n$ and every $r \in \mathbb{F}$, we have that

$$f_A(r \cdot \bar{u}_1) = A(r \cdot \bar{u}_1) = r \cdot (A\bar{u}_1) = r \cdot f_A(\bar{u}_1).$$

The above combined show that f_A is linear.

What about the reverse?
Matrix representations of linear functions

It shouldn't be hard to convince ourselves that the converse to the above observation would be very useful (given that being able to have a 'formula' for a linear function given via a matrix would allow us more 'concrete' access to what this function does). Thus it makes sense to ask the following

Question. Given a linear function $f : \mathbb{F}^n \to \mathbb{F}^m$, can we find a matrix $A = A_f \in \mathbb{F}^{m \times n}$ such that,

$$\text{for all } \bar{u} \in \mathbb{F}^n, \text{ we will have } f(\bar{u}) = A\bar{u}\,?$$

A matrix A satisfying this would be called a *matrix representation* of f.

How can we make a 'good' guess about what the entries of such a matrix A should be?

First, we need to get a good idea of what matrix (or matrices) in $\mathbb{F}^{m\times n}$ could satisfy $f(\bar{u}) = A\bar{u}$ for every $\bar{u} \in \mathbb{F}^n$.

If we examine what the required property gives when $\bar{u}$ is one of the standard basis vectors of $\mathbb{F}^n$, then we get that, for every $i \in \{1, 2, \ldots, n\}$,

$$f(\bar{e}_i) \text{ should be equal to } A\bar{e}_i, \text{ with the latter being equal to } \mathrm{Col}_i(A).$$

But since we consider all $i \in \{1, 2, \ldots, n\}$, this 'prescribes' what **all the columns of A should be,** or in other words, it shows that only the matrix

$$A_f := \begin{pmatrix} | & | & | & & | \\ f(\bar{e}_1) & f(\bar{e}_2) & f(\bar{e}_3) & \cdots & f(\bar{e}_n) \\ | & | & | & & | \end{pmatrix}$$

is a 'good' candidate here.

Finally we check that this matrix works: for every $\bar{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{F}^n$, recall that we can write

$$\begin{aligned} A_f\bar{u} &= u_1 \cdot \mathrm{Col}_1(A_f) + u_2 \cdot \mathrm{Col}_2(A_f) + \cdots + u_n \cdot \mathrm{Col}_n(A_f) \\ &= u_1 \cdot f(\bar{e}_1) + u_2 \cdot f(\bar{e}_2) + \cdots + u_n \cdot f(\bar{e}_n) \\ &= f\big(u_1 \cdot \bar{e}_1 + u_2 \cdot \bar{e}_2 + \cdots + u_n \cdot \bar{e}_n\big) \\ &= f(\bar{u}), \end{aligned}$$

where the third equality holds because of the linearity of f.

Applying this to an example

Find a matrix representation for the function $f_3 : \mathbb{R}^4 \to \mathbb{R}^5$ given by

$$f_3\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\right) = \begin{pmatrix} 2x_2 + x_3 + x_4 \\ x_1 - 2x_3 \\ 2x_1 - 2x_3 \\ x_1 - 2x_2 - 3x_3 + x_4 \\ x_1 + x_2 + x_3 + x_4 \end{pmatrix}.$$

Answer. We have that

$$f_3(\bar{e}_1) = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad f_3(\bar{e}_2) = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}, \quad f_3(\bar{e}_3) = \begin{pmatrix} 1 \\ -2 \\ -2 \\ -3 \\ 1 \end{pmatrix}, \quad \text{and} \quad f_3(\bar{e}_4) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Thus the desired matrix representation of f_3 is the matrix

$$A_{f_3} = \begin{pmatrix} 0 & 2 & 1 & 1 \\ 1 & 0 & -2 & 0 \\ 2 & 0 & -2 & 0 \\ 1 & -2 & -3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Interesting question

Recall that we have said that, given a matrix $A \in \mathbb{F}^{m \times n}$, there are subspaces of $\mathbb{F}^m$ and of $\mathbb{F}^n$ that we can naturally associate with A.

1. The *Row Space* of A is the linear span of all rows of A. We denote it by $RS(A)$.

   Observe that $RS(A)$ is a subspace of $\mathbb{F}^n$.

2. The *Column Space* of A is the linear span of all columns of A. We denote it by $CS(A)$.

   Observe that $CS(A)$ is a subspace of $\mathbb{F}^m$.

3. The *Nullspace* of A is defined to be the solution set to the system $A\bar{x} = \bar{0}$. We denote it by $N(A)$.

   In other words, $N(A) = \left\{ \bar{u} \in \mathbb{F}^n : A\bar{u} = \bar{0}_{\mathbb{F}^m} \right\}$.

Question. If we now consider the linear function $f : \mathbb{F}^n \to \mathbb{F}^m$ which has matrix representation A, and we also recall the important subspaces associated with this function f, that is, the subspaces $\mathrm{Range}(f)$ and $\mathrm{Ker}(f)$, what is the relation between these subspaces and the above subspaces? Are there any connections / identifications we can make? YES, we have that $\mathrm{Ker}(f) = N(A)$ and $\mathrm{Range}(f) = CS(A)$ (why?).

Isomorphic Mathematical Structures

In Mathematics, it's very useful to determine which structures out of those that we study are *isomorphic*. These would essentially be structures that we can think of as being identical, with their only differences being **superficial** (from a mathematical point of view): e.g. what the 'name' of the elements contained in the structure is, rather than how they interact with each other, which is most commonly what we want to focus on.

E.g. the structure formed by the set $\{Green, Yellow\}$ together with the operations

$+$	*Green*	*Yellow*
Green	*Yellow*	*Green*
Yellow	*Green*	*Yellow*

$\cdot$	*Green*	*Yellow*
Green	*Green*	*Yellow*
Yellow	*Yellow*	*Yellow*

is isomorphic to $\mathbb{Z}_2$ with the standard modulo 2 operations *(we'll properly define a bit later what we should understand by 'isomorphism' here, although for these two structures it should not be hard to see which 'renaming' of the elements we should do).*

Isomorphic Vector Spaces

Definition

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

We will say that V_1 and V_2 are *isomorphic*, and we will write $V_1 \cong V_2$, if there exist linear functions $f : V_1 \to V_2$ and $g : V_2 \to V_1$ such that

$$\text{(i)}\ \ g \circ f = \mathrm{id}_{V_1} \qquad \text{and} \qquad \text{(ii)}\ \ f \circ g = \mathrm{id}_{V_2}.$$

In other words, if there exists an **invertible** linear function $f : V_1 \to V_2$ such that its inverse $f^{-1} : V_2 \to V_1$ is also linear.
In such a case we call f an *isomorphism*, or more precisely a *linear isomorphism*, from V_1 to V_2.

Important Observation

We have already said that a function $h : V_1 \to V_2$ will have an inverse $h^{-1} : V_2 \to V_1$ **if and only if** h is bijective.
In other words, to even be able to consider the inverse of a linear function $f : V_1 \to V_2$ (and before we even ask whether this inverse is a linear function or not), we need to know that f is bijective, which, as we have already seen, is equivalent to

$$\mathrm{Range}(f) = V_2 \qquad \text{and} \qquad \mathrm{Ker}(f) = \{\bar{0}_{V_1}\}.$$

It turns out that this is also sufficient, that is, if a linear function $f : V_1 \to V_2$ is bijective, then it is an isomorphism from V_1 to V_2.

Linear isomorphisms coincide with bijective linear maps

This is because of the following

Theorem 4

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

Suppose that $f : V_1 \to V_2$ is a linear function, **and suppose that f is bijective.**

Then the function $f^{-1} : V_2 \to V_1$ is also linear.

Again, recall that, because we have assumed that f is bijective, we know that f has an inverse function, the function $f^{-1} : V_2 \to V_1$ which is defined by $f^{-1}(\bar{y}) = \bar{x} \Leftrightarrow f(\bar{x}) = \bar{y}$ for every $\bar{y} \in V_2$, $\bar{x} \in V_1$.

Proof. *See next lecture.*

One of the most important theorems in Linear Algebra

Theorem

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

V_1 and V_2 are isomorphic (that is, there exists a linear isomorphism from V_1 to V_2) **if and only if** V_1 and V_2 have the same **dimension** over $\mathbb{F}$.

Of course we still have to properly define the notion of 'dimension of a vector space', so that it's clear what this theorem says; we will do so in the coming lectures.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 43

Friday November 27

Recall: Linear Functions

Definition

Let $\mathbb{F}$ be a field, and let $V_1 = \big(\{\text{elements in } V_1\}, +_1, \cdot_1\big)$ and $V_2 = \big(\{\text{elements in } V_2\}, +_2, \cdot_2\big)$ be two vector spaces **over $\mathbb{F}$.**

A function $f : V_1 \to V_2$ is called a *linear function*, or a *linear map*, or a *vector space homomorphism*, if f satisfies the following two properties:

1. for every $\bar{x}, \bar{y} \in V_1$, we have that
$$f(\bar{x} +_1 \bar{y}) = f(\bar{x}) +_2 f(\bar{y}).$$
2. for every $\lambda \in \mathbb{F}$ and $\bar{x} \in V_1$, we have that
$$f(\lambda \cdot_1 \bar{x}) = \lambda \cdot_2 f(\bar{x}).$$

Isomorphic Mathematical Structures

In Mathematics, it's very useful to determine which structures out of those that we study are *isomorphic*. These would essentially be structures that we can think of as being identical, with their only differences being **superficial** (from a mathematical point of view): e.g. what the 'name' of the elements contained in the structure is, rather than how they interact with each other, which is most commonly what we want to focus on.

E.g. the structure formed by the set $\{Green, Yellow\}$ together with the operations

$+$	*Green*	*Yellow*
Green	*Yellow*	*Green*
Yellow	*Green*	*Yellow*

$\cdot$	*Green*	*Yellow*
Green	*Green*	*Yellow*
Yellow	*Yellow*	*Yellow*

is isomorphic to $\mathbb{Z}_2$ with the standard modulo 2 operations *(we'll properly define a bit later what we should understand by 'isomorphism' here, although for these two structures it should not be hard to see which 'renaming' of the elements we should do).*

Isomorphic Vector Spaces

Definition

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

We will say that V_1 and V_2 are *isomorphic*, and we will write $V_1 \cong V_2$, if there exist linear functions $f : V_1 \to V_2$ and $g : V_2 \to V_1$ such that

$$\text{(i)}\ \ g \circ f = \mathrm{id}_{V_1} \qquad \text{and} \qquad \text{(ii)}\ \ f \circ g = \mathrm{id}_{V_2}.$$

In other words, if there exists an **invertible** linear function $f : V_1 \to V_2$ such that its inverse $f^{-1} : V_2 \to V_1$ is also linear.
In such a case we call f an *isomorphism*, or more precisely a *linear isomorphism*, from V_1 to V_2.

Important Observation

We have already said that a function $h : V_1 \to V_2$ will have an inverse $h^{-1} : V_2 \to V_1$ **if and only if** h is bijective.
In other words, to even be able to consider the inverse of a linear function $f : V_1 \to V_2$ (and before we even ask whether this inverse is a linear function or not), we need to know that f is bijective, which, as we have already seen, is equivalent to

$$\mathrm{Range}(f) = V_2 \qquad \text{and} \qquad \mathrm{Ker}(f) = \{\bar{0}_{V_1}\}.$$

It turns out that this is also sufficient, that is, if a linear function $f : V_1 \to V_2$ is bijective, then it is an isomorphism from V_1 to V_2.

Linear isomorphisms coincide with bijective linear maps

This is because of the following

Theorem 4

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

Suppose that $f : V_1 \to V_2$ is a linear function, **and suppose that f is bijective.**

Then the function $f^{-1} : V_2 \to V_1$ is also linear.

Again, recall that, because we have assumed that f is bijective, we know that f has an inverse function, the function $f^{-1} : V_2 \to V_1$ which is defined by $f^{-1}(\bar{y}) = \bar{x} \Leftrightarrow f(\bar{x}) = \bar{y}$ for every $\bar{y} \in V_2$, $\bar{x} \in V_1$.

Proof of Theorem 4

Let $\bar{w}_1, \bar{w}_2 \in V_2$, and let $r \in \mathbb{F}$. We need to show that

$$f^{-1}(\bar{w}_1 + \bar{w}_2) = f^{-1}(\bar{w}_1) + f^{-1}(\bar{w}_2) \qquad \text{and} \qquad f^{-1}(r \cdot \bar{w}_1) = r \cdot f^{-1}(\bar{w}_1).$$

We first observe that we can find $\bar{x}_1, \bar{x}_2 \in V_1$ such that $f(\bar{x}_1) = \bar{w}_1 \Leftrightarrow \bar{x}_1 = f^{-1}(\bar{w}_1)$, and analogously $f(\bar{x}_2) = \bar{w}_2 \Leftrightarrow \bar{x}_2 = f^{-1}(\bar{w}_2)$. We can now write

$$\bar{w}_1 + \bar{w}_2 = f(\bar{x}_1) + f(\bar{x}_2) = f(\bar{x}_1 + \bar{x}_2)$$

because f is linear. This shows that $\bar{x}_1 + \bar{x}_2$ is the (unique) preimage of $\bar{w}_1 + \bar{w}_2$ under f, and hence we have $f^{-1}(\bar{w}_1 + \bar{w}_2) = \bar{x}_1 + \bar{x}_2$.

We can now conclude that

$$f^{-1}(\bar{w}_1 + \bar{w}_2) = \bar{x}_1 + \bar{x}_2 = f^{-1}(\bar{w}_1) + f^{-1}(\bar{w}_2),$$

as we wanted.

Similarly, we can observe that

$$r \cdot \bar{w}_1 = r \cdot f(\bar{x}_1) = f(r \cdot \bar{x}_1)$$

because f is linear. This shows that $r \cdot \bar{x}_1$ is the (unique) preimage of $r \cdot \bar{w}_1$ under f, and hence we have $f^{-1}(r \cdot \bar{w}_1) = r \cdot \bar{x}_1$.
We can now conclude that

$$f^{-1}(r \cdot \bar{w}_1) = r \cdot \bar{x}_1 = r \cdot f^{-1}(\bar{w}_1),$$

as we wanted.

One of the most important theorems in Linear Algebra

Theorem

Let $\mathbb{F}$ be a field, and let V_1, V_2 be two vector spaces over $\mathbb{F}$.

V_1 and V_2 are isomorphic (that is, there exists a linear isomorphism from V_1 to V_2) **if and only if** V_1 and V_2 have the same **dimension** over $\mathbb{F}$.

Of course we still have to properly define the notion of 'dimension of a vector space', so that it's clear what this theorem says; we will do so in the coming lectures.

Other subspaces associated with a linear function and/or its matrix representation

Definition: *Invariant Subspaces*

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Consider a linear function $f : V \to V$ (that is, a linear function from the vector space V to itself).

A subspace S of V is called an *invariant subspace* of f, or with respect to f, if S is **preserved by f**, that is, if

$$f(S) := \{f(\bar{x}) : \bar{x} \in S\} \subseteq S.$$

Remark. Recall that, if $V = \mathbb{F}^n$ for some $n \geqslant 1$, and we consider a function $g : \mathbb{F}^n \to \mathbb{F}^n$, then g can be also be represented by a matrix $A = A_g \in \mathbb{F}^{n \times n}$ (recall also that by this we mean that we will have $g(\bar{u}) = A\bar{u}$ for all $\bar{u} \in \mathbb{F}^n$).

But then in this setting we can also talk about **invariant subspaces with respect to the matrix A**: that is, for any invariant subspace T of the function g, we can also choose to call it an invariant subspace under the matrix A. Such a subspace T will need to satisfy the condition

$$\{A\bar{u} : \bar{u} \in T\} \subseteq T,$$

which in this setting is equivalent to the condition of the definition:

$$g(T) = \{g(\bar{u}) : \bar{u} \in T\} \subseteq T.$$

Examples of Invariant Subspaces

- The first examples of invariant subspaces of a linear function $f : V \to V$ are the two trivial subspaces of V, that is, V itself and the zero subspace $\{\bar{0}_V\}$.

— Indeed, $f(V) = \mathrm{Range}(f) \subseteq V$, so V is preserved by f.

— Moreover, $f\left(\{\bar{0}_V\}\right) = \left\{f(\bar{0}_V)\right\} = \{\bar{0}_V\}$.

- Another family of examples of invariant subspaces is related to the following notion.

Definition: *Eigenvalues and eigenvectors*

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$, and let $f : V \to V$ be a linear function.

A **non-zero** vector $\bar{u}$ of V is called an *eigenvector* of f if

$$\text{we can find } \lambda \in \mathbb{F} \text{ such that } f(\bar{u}) = \lambda \cdot \bar{u}.$$

In such a case λ is called an *eigenvalue* of f.

Invariant Subspaces and Eigenvectors

Consider a linear function $f : V \to V$ and assume that $\bar{u}_0 \in V$ is an eigenvector of f (corresponding to eigenvalue λ).

Consider now a vector $\bar{x}$ in $\mathrm{span}(\bar{u}_0)$, that is, $\bar{x} = \mu \cdot \bar{u}_0$ for some $\mu \in \mathbb{F}$. Then we can write

$$\begin{aligned} f(\bar{x}) = f(\mu \cdot \bar{u}_0) &= \mu \cdot f(\bar{u}_0) && \text{(by the linearity of } f\text{)} \\ &= \mu \cdot (\lambda \cdot \bar{u}_0) && (\bar{u}_0 \text{ is an eigenvector of } f) \\ &= (\mu \cdot \lambda) \cdot \bar{u}_0 \quad \in \mathrm{span}(\bar{u}_0). \end{aligned}$$

What we have just checked is that the subspace $\mathrm{span}(\bar{u}_0)$ is preserved by f. In other words,

a subspace of V which is spanned by a non-zero vector $\bar{v}$
is an invariant subspace of f **if and only if**
the vector $\bar{v}$ is an eigenvector of f.

An example

Consider the linear function $f : \mathbb{R}^4 \to \mathbb{R}^4$ given by

$$f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\right) = \begin{pmatrix} 2x_2 \\ 2x_1 \\ x_3 - 3x_4 \\ x_3 + x_4 \end{pmatrix}.$$

Claim. The vector $\bar{e}_1 + \bar{e}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector of f.

Verify this, and find what eigenvalue it corresponds to as well.

Solution. We can directly check that

$$f\left(\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 2 \cdot 1 \\ 2 \cdot 1 \\ 0 - 3 \cdot 0 \\ 0 + 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \\ 0 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Thus $\bar{e}_1 + \bar{e}_2$ is an eigenvector of f corresponding to eigenvalue 2.

Question 1. Are there any other eigenvectors of f? *To discuss next time.*

Question 2. Are there any other eigenvectors of f which are not scalar multiples of $\bar{e}_1 + \bar{e}_2$? *To discuss next time.*

Eigenvalues and eigenvectors of matrices

Analogously we define eigenvalues and eigenvectors of square matrices (that is, matrices which can be viewed as representations of linear functions $f : V \to V$ for a vector space V of the form $\mathbb{F}^n$).

Definition: *Eigenvalues and eigenvectors of a matrix*

Let $\mathbb{F}$ be a field, $n \geqslant 1$, and let A be a matrix in $\mathbb{F}^{n \times n}$.

A **non-zero** vector $\bar{v}$ of $\mathbb{F}^n$ is called an *eigenvector* of A if

$$\text{we can find } \lambda \in \mathbb{F} \text{ such that } A\bar{v} = \lambda \cdot \bar{v}.$$

In such a case λ is called an *eigenvalue* of A.

Eigenvalues and eigenvectors of matrices (cont.)

A very useful remark

Consider $A \in \mathbb{F}^{n \times n}$ and $\lambda \in \mathbb{F}$.

Then λ is an eigenvalue of A **if and only if**

the Nullspace $N(A - \lambda I_n)$ of the matrix $A - \lambda I_n$
contains non-zero vectors
(in other words, $N(A - \lambda I_n) \neq \{\bar{0}\}$) **if and only if**

the matrix $A - \lambda I_n$ is **not** invertible.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 44

Monday November 30

Subspaces associated with a linear function and/or its matrix representation

Definition: *Invariant Subspaces*

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Consider a linear function $f : V \to V$ (that is, a linear function from the vector space V to itself).

A subspace S of V is called an *invariant subspace* of f, or with respect to f, if S is **preserved by f**, that is, if

$$f(S) := \{f(\bar{x}) : \bar{x} \in S\} \subseteq S.$$

Remark. Recall that, if $V = \mathbb{F}^n$ for some $n \geqslant 1$, and we consider a function $g : \mathbb{F}^n \to \mathbb{F}^n$, then g can be also be represented by a matrix $A = A_g \in \mathbb{F}^{n \times n}$ (recall also that by this we mean that we will have $g(\bar{u}) = A\bar{u}$ for all $\bar{u} \in \mathbb{F}^n$).

But then in this setting we can also talk about **invariant subspaces with respect to the matrix A**: that is, for any invariant subspace T of the function g, we can also choose to call it an invariant subspace under the matrix A. Such a subspace T will need to satisfy the condition

$$\{A\bar{u} : \bar{u} \in T\} \subseteq T,$$

which in this setting is equivalent to the condition of the definition:

$$g(T) = \{g(\bar{u}) : \bar{u} \in T\} \subseteq T.$$

Examples of Invariant Subspaces

- The first examples of invariant subspaces of a linear function $f : V \to V$ are the two trivial subspaces of V, that is, V itself and the zero subspace $\{\bar{0}_V\}$.

— Indeed, $f(V) = \mathrm{Range}(f) \subseteq V$, so V is preserved by f.

— Moreover, $f\left(\{\bar{0}_V\}\right) = \left\{f(\bar{0}_V)\right\} = \{\bar{0}_V\}$.

- Another family of examples of invariant subspaces is related to the following notion.

Definition: *Eigenvalues and eigenvectors*

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$, and let $f : V \to V$ be a linear function.

A **non-zero** vector $\bar{u}$ of V is called an *eigenvector* of f if

$$\text{we can find } \lambda \in \mathbb{F} \text{ such that } f(\bar{u}) = \lambda \cdot \bar{u}.$$

In such a case λ is called an *eigenvalue* of f.

Invariant Subspaces and Eigenvectors

Consider a linear function $f : V \to V$ and assume that $\bar{u}_0 \in V$ is an eigenvector of f (corresponding to eigenvalue λ).

Consider now a vector $\bar{x}$ in $\mathrm{span}(\bar{u}_0)$, that is, $\bar{x} = \mu \cdot \bar{u}_0$ for some $\mu \in \mathbb{F}$. Then we can write

$$\begin{aligned} f(\bar{x}) = f(\mu \cdot \bar{u}_0) &= \mu \cdot f(\bar{u}_0) && \text{(by the linearity of } f\text{)} \\ &= \mu \cdot (\lambda \cdot \bar{u}_0) && (\bar{u}_0 \text{ is an eigenvector of } f) \\ &= (\mu \cdot \lambda) \cdot \bar{u}_0 \quad \in \mathrm{span}(\bar{u}_0). \end{aligned}$$

What we have just checked is that the subspace $\mathrm{span}(\bar{u}_0)$ is preserved by f. In other words,

a subspace of V which is spanned by a non-zero vector $\bar{v}$
is an invariant subspace of f **if and only if**
the vector $\bar{v}$ is an eigenvector of f.

An example

Consider the linear function $f : \mathbb{R}^4 \to \mathbb{R}^4$ given by

$$f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\right) = \begin{pmatrix} 2x_2 \\ 2x_1 \\ x_3 - 3x_4 \\ x_3 + x_4 \end{pmatrix}.$$

Claim. The vector $\bar{e}_1 + \bar{e}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector of f.

Verify this, and find what eigenvalue it corresponds to as well.

Solution. We can directly check that

$$f\left(\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 2 \cdot 1 \\ 2 \cdot 1 \\ 0 - 3 \cdot 0 \\ 0 + 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \\ 0 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Thus $\bar{e}_1 + \bar{e}_2$ is an eigenvector of f corresponding to eigenvalue 2.

Question 1. Are there any other eigenvectors of f?

Question 2. Are there any other eigenvectors of f which are not scalar multiples of $\bar{e}_1 + \bar{e}_2$?

Answer to Question 1. Yes, every **non-zero** scalar multiple of $\bar{e}_1 + \bar{e}_2$ is also an eigenvector of f (and again it corresponds to the eigenvalue 2). Indeed, if $\bar{x} = t \cdot (\bar{e}_1 + \bar{e}_2)$ for some $t \in \mathbb{R} \setminus \{0\}$, then, by the linearity of f, we can write

$$\begin{aligned} f(\bar{x}) = f\big(t(\bar{e}_1 + \bar{e}_2)\big) &= t \cdot f(\bar{e}_1 + \bar{e}_2) \\ &= t \cdot \big(2 \cdot (\bar{e}_1 + \bar{e}_2)\big) = 2 \cdot \big(t \cdot (\bar{e}_1 + \bar{e}_2)\big) = 2 \cdot \bar{x}. \end{aligned}$$

Answer to Question 2. We can check that this function f has one more eigenvalue, the eigenvalue -2. One eigenvector corresponding to this eigenvalue is the vector $\bar{e}_1 - \bar{e}_2$. Indeed,

$$f\left(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 2 \cdot (-1) \\ 2 \cdot 1 \\ 0 - 3 \cdot 0 \\ 0 + 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 0 \\ 0 \end{pmatrix} = (-2) \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

Of course, similarly to what we showed above, we also get that every non-zero scalar multiple of $\bar{e}_1 - \bar{e}_2$ is an eigenvector of f corresponding to eigenvalue -2.

On the other hand, we can check that f doesn't have any other eigenvalues, and also that the only eigenvectors corresponding to eigenvalue 2 are the non-zero scalar multiples of $\bar{e}_1 + \bar{e}_2$, while the only eigenvectors corresponding to eigenvalue -2 are the non-zero scalar multiples of $\bar{e}_1 - \bar{e}_2$.

Indeed, to show e.g. the second claim, we need to analyse/solve completely the linear system

$$\begin{pmatrix} 2x_2 \\ 2x_1 \\ x_3 - 3x_4 \\ x_3 + x_4 \end{pmatrix} = f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\right) = 2 \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \\ 2x_4 \end{pmatrix}.$$

Observe that the equations $x_3 - 3x_4 = 2x_3$ and $x_3 + x_4 = 2x_4$ considered together will give us $x_3 = x_4 = 0$, while the equations $2x_2 = 2x_1$ and $2x_1 = 2x_2$ are essentially the same and imply that $x_1 = x_2$. Thus, taking into account that we are only looking for non-zero vectors, we conclude that the only such solutions to this linear system are the non-zero scalar multiples of $\bar{e}_1 + \bar{e}_2$.

Eigenvalues and eigenvectors of matrices

Analogously we define eigenvalues and eigenvectors of square matrices (that is, matrices which can be viewed as representations of linear functions $f : V \to V$ for a vector space V of the form $\mathbb{F}^n$).

Definition: *Eigenvalues and eigenvectors of a matrix*

Let $\mathbb{F}$ be a field, $n \geqslant 1$, and let A be a matrix in $\mathbb{F}^{n \times n}$.

A **non-zero** vector $\bar{v}$ of $\mathbb{F}^n$ is called an *eigenvector* of A if

$$\text{we can find } \lambda \in \mathbb{F} \text{ such that } A\bar{v} = \lambda \cdot \bar{v}.$$

In such a case λ is called an *eigenvalue* of A.

Eigenvalues and eigenvectors of matrices (cont.)

A very useful remark

Consider $A \in \mathbb{F}^{n \times n}$ and $\lambda \in \mathbb{F}$. Then

(I) λ is an eigenvalue of A **if and only if**

(II) the Nullspace $N(A - \lambda I_n)$ of the matrix $A - \lambda I_n$ contains non-zero vectors
(in other words, $N(A - \lambda I_n) \neq \{\bar{0}\}$) **if and only if**

(III) the matrix $A - \lambda I_n$ is **not** invertible.

Justification of the remark

We show that $(I) \Leftrightarrow (II)$ and $(II) \Leftrightarrow (III)$.

$(I) \Rightarrow (II)$ Suppose that λ is an eigenvalue of A. Then, we can find a **non-zero** vector $\bar{u} \in \mathbb{F}^n$ such that $A\bar{u} = \lambda \cdot \bar{u}$. We can then write

$$A\bar{u} = \lambda \cdot \bar{u} = \lambda \cdot (I_n \bar{u}) = (\lambda I_n)\bar{u} \quad \Rightarrow \quad (A - \lambda I_n)\bar{u} = A\bar{u} - (\lambda I_n)\bar{u} = \bar{0}.$$

Thus $\bar{u} \in N(A - \lambda I_n)$, and hence this nullspace contains non-zero vectors.

$(II) \Rightarrow (I)$ Suppose now that there exists a non-zero vector $\bar{v} \in N(A - \lambda I_n)$. Then, by definition of the nullspace, we have that

$$(A - \lambda I_n)\bar{v} = \bar{0} \quad \Rightarrow \quad A\bar{v} = (\lambda I_n)\bar{v} = \lambda \cdot (I_n \bar{v}) = \lambda \cdot \bar{v}.$$

Given that $\bar{v}$ is non-zero, we conclude that it is an eigenvector of A, and that λ is the corresponding eigenvalue of A (so in particular λ is an eigenvalue of A).

$(II) \Rightarrow (III)$ Suppose again that there exists a non-zero vector $\bar{v} \in N(A - \lambda I_n)$. This is equivalent to saying that the linear system $(A - \lambda I_n)\bar{y} = \bar{0}$ has a solution **different from the trivial solution.** But this will imply that not all variables of the system are pivot variables, and hence, no matter which REF of $A - \lambda I_n$ we consider, it will have $< n$ pivots. We can conclude that $A - \lambda I_n$ is **not** invertible.

$(III) \Rightarrow (II)$ Similarly to above, if we suppose that $A - \lambda I_n$ is **not** invertible, then any REF B_1 of this matrix will have $< n$ pivots, and hence the linear system $B_1\bar{y} = \bar{0}$ corresponding to this REF of $A - \lambda I_n$ will have non-trivial solutions. Since this system is equivalent to the system $(A - \lambda I_n)\bar{y} = \bar{0}$, we can conclude that $N(A - \lambda I_n)$ will contain non-zero vectors.

A practice problem

Past Homework Problem. Consider $A \in \mathbb{R}^{n \times n}$, and suppose that λ is an eigenvalue of A.

(i) Show that 2λ is an eigenvalue of $2A$.

(ii) Show that $\lambda + 1$ is an eignevalue of $A + I_n$.

(iii) Show that λ^2 is an eigenvalue of A^2.

(iv) Given a real polynomial $p(x)$, show that $p(\lambda)$ is an eigenvalue of $p(A)$ *(see the note below about how we define $p(A)$)*.

[*Note.* Recall that, if k is a positive integer, we write A^k for the product $A \cdot A \cdots A$ having k factors, all equal to A. Generalising this, if we have a real polynomial $q(x) = c_k x^k + c_{k-1} x^{k-1} + \cdots + c_1 x + c_0$, we set $q(A)$ to be the matrix $c_k A^k + c_{k-1} A^{k-1} + \cdots + c_1 A + c_0 I_n$.]

An important non-commutative ring

Recall HW5, Problem 6: Part (a) tells us that, if $f, g : V_1 \to V_2$ are linear maps, where V_1, V_2 are two vector spaces over the same scalar field, then the function

$$f + g : V_1 \to V_2$$

is also linear.

Moreover, part (b) tells us that, if $h : V_1 \to V_2$ is a linear function, and $u : V_2 \to V_3$ is another linear function, where V_3 is again a vector space over the same scalar field, then the function

$$u \circ h$$

is linear too.

Suppose now that we focus on one vector space V_0, and on **linear functions $f : V_0 \to V_0$.** Let's denote the set of all such linear functions by $L(V_0, V_0)$. By what we just recalled, applied to this more special case, we have that, for every two functions $f_1, f_2 \in L(V_0, V_0)$,

$$f_1 + f_2 \in L(V_0, V_0) \text{ again,} \qquad \text{and similarly } f_2 \circ f_1 \in L(V_0, V_0).$$

Question / Idea. Could we treat addition of functions as an operation on $L(V_0, V_0)$? Similarly, could we view composition of functions **as a type of a multiplication operation** on $L(V_0, V_0)$?

An important non-commutative ring

Very Important Theorem

The triple $\big(\{\text{functions in } L(V_0, V_0)\}, +, \circ\big)$ is a ring.

In general, it is **not** a commutative ring (except when V_0 is spanned by only one vector, and hence it is either a zero vector space or a 'line'). That is, in almost all cases the 'multiplication' $\circ$ in this ring is not commutative.

Remark. The neutral element of addition will be the zero function f_{zero}, that is, the function

$$f_{\text{zero}} : V_0 \to V_0 \quad \text{satisfying} \quad f_{\text{zero}}(\bar{x}) = \bar{0}_{V_0} \quad \text{for all } \bar{x} \in V_0.$$

On the other hand, the neutral element of 'multiplication' is the identity function $\mathrm{id}_{V_0} : V_0 \to V_0$ (why?).

Proof of this theorem?
Verifying the ring structure of $L(V_0, V_0)$

Recall that we need to check

- for every $f_1, f_2 \in L(V_0, V_0)$, $f_1 + f_2 = f_2 + f_1$.
- for every $f_1, f_2, f_3 \in L(V_0, V_0)$, $(f_1 + f_2) + f_3 = f_1 + (f_2 + f_3)$.
- for every $f \in L(V_0, V_0)$, $f + f_{\text{zero}} = f = f_{\text{zero}} + f$.
- for every $f \in L(V_0, V_0)$ we can find $g \in L(V_0, V_0)$ such that
$$f + g = f_{\text{zero}} = g + f.$$

- for every $f_1, f_2, f_3 \in L(V_0, V_0)$, $(f_1 \circ f_2) \circ f_3 = f_1 \circ (f_2 \circ f_3)$.
- for every $f \in L(V_0, V_0)$, $f \circ \mathrm{id}_{V_0} = f = \mathrm{id}_{V_0} \circ f$.
- *(Right Distributive Property)* for every $f_1, f_2, f_3 \in L(V_0, V_0)$,
$$(f_1 + f_2) \circ f_3 = f_1 \circ f_3 + f_2 \circ f_3.$$
- *(Left Distributive Property)* for every $g_1, g_2, g_3 \in L(V_0, V_0)$,
$$g_3 \circ (g_1 + g_2) = g_3 \circ g_1 + g_3 \circ g_2.$$

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 45

Tuesday December 1

An important non-commutative ring

Recall HW5, Problem 6: Part (a) tells us that, if $f, g : V_1 \to V_2$ are linear maps, where V_1, V_2 are two vector spaces over the same scalar field, then the function

$$f + g : V_1 \to V_2$$

is also linear.

Moreover, part (b) tells us that, if $h : V_1 \to V_2$ is a linear function, and $u : V_2 \to V_3$ is another linear function, where V_3 is again a vector space over the same scalar field, then the function

$$u \circ h$$

is linear too.

Suppose now that we focus on one vector space V_0, and on **linear functions $f : V_0 \to V_0$.** Let's denote the set of all such linear functions by $L(V_0, V_0)$. By what we just recalled, applied to this more special case, we have that, for every two functions $f_1, f_2 \in L(V_0, V_0)$,

$$f_1 + f_2 \in L(V_0, V_0) \text{ again,} \qquad \text{and similarly } f_2 \circ f_1 \in L(V_0, V_0).$$

Question / Idea. Could we treat addition of functions as an operation on $L(V_0, V_0)$? Similarly, could we view composition of functions **as a type of a multiplication operation** on $L(V_0, V_0)$?

An important non-commutative ring

Very Important Theorem

The triple $\big(\{\text{functions in } L(V_0, V_0)\}, +, \circ\big)$ is a ring.

In general, it is **not** a commutative ring (except when V_0 is spanned by only one vector, and hence it is either a zero vector space or a 'line'). That is, in almost all cases the 'multiplication' $\circ$ in this ring is not commutative.

Remark. The neutral element of addition will be the zero function f_{zero}, that is, the function

$$f_{\text{zero}} : V_0 \to V_0 \quad \text{satisfying} \quad f_{\text{zero}}(\bar{x}) = \bar{0}_{V_0} \quad \text{for all } \bar{x} \in V_0.$$

On the other hand, the neutral element of 'multiplication' is the identity function $\mathrm{id}_{V_0} : V_0 \to V_0$ (why?).

Proof of this theorem?
Verifying the ring structure of $L(V_0, V_0)$

Recall that we need to check

- for every $f_1, f_2 \in L(V_0, V_0)$, $f_1 + f_2 = f_2 + f_1$.
- for every $f_1, f_2, f_3 \in L(V_0, V_0)$, $(f_1 + f_2) + f_3 = f_1 + (f_2 + f_3)$.
- for every $f \in L(V_0, V_0)$, $f + f_{\text{zero}} = f = f_{\text{zero}} + f$.
- for every $f \in L(V_0, V_0)$ we can find $g \in L(V_0, V_0)$ such that
$$f + g = f_{\text{zero}} = g + f.$$

- for every $f_1, f_2, f_3 \in L(V_0, V_0)$, $(f_1 \circ f_2) \circ f_3 = f_1 \circ (f_2 \circ f_3)$.
- for every $f \in L(V_0, V_0)$, $f \circ \mathrm{id}_{V_0} = f = \mathrm{id}_{V_0} \circ f$.
- *(Right Distributive Property)* for every $f_1, f_2, f_3 \in L(V_0, V_0)$,
$$(f_1 + f_2) \circ f_3 = f_1 \circ f_3 + f_2 \circ f_3.$$
- *(Left Distributive Property)* for every $g_1, g_2, g_3 \in L(V_0, V_0)$,
$$g_3 \circ (g_1 + g_2) = g_3 \circ g_1 + g_3 \circ g_2.$$

Verifying (some of) these properties

- Let $f_1, f_2, f_3 \in L(V_0, V_0)$. In order to verify the *functional identity*

$$(f_1 \circ f_2) \circ f_3 = f_1 \circ (f_2 \circ f_3),$$

what we do is check that, for every $\bar{x} \in V_0$, we have

$$\big((f_1 \circ f_2) \circ f_3\big)(\bar{x}) = \big(f_1 \circ (f_2 \circ f_3)\big)(\bar{x}).$$

Indeed,

$$\begin{aligned} \big((f_1 \circ f_2) \circ f_3\big)(\bar{x}) &= (f_1 \circ f_2)\big(f_3(\bar{x})\big) \quad \text{(by definition of composition)} \\ &= f_1\Big(f_2\big(f_3(\bar{x})\big)\Big). \end{aligned}$$

Analogously, we have

$$\big(f_1 \circ (f_2 \circ f_3)\big)(\bar{x}) = f_1\big((f_2 \circ f_3)(\bar{x})\big) = f_1\Big(f_2\big(f_3(\bar{x})\big)\Big).$$

Verifying (some of) these properties (cont.)

- Let $f_1, f_2, f_3 \in L(V_0, V_0)$. In order to verify the *functional identity*

$$(f_1 + f_2) \circ f_3 = f_1 \circ f_3 + f_2 \circ f_3,$$

we need to check that, for every $\bar{x} \in V_0$, we have that

$$\big((f_1 + f_2) \circ f_3\big)(\bar{x}) = \big(f_1 \circ f_3 + f_2 \circ f_3\big)(\bar{x}).$$

Indeed,

$$\begin{aligned}
\big((f_1 + f_2) \circ f_3\big)(\bar{x}) &= (f_1 + f_2)\big(f_3(\bar{x})\big) && \text{(by definition of composition)} \\
&= f_1\big(f_3(\bar{x})\big) + f_2\big(f_3(\bar{x})\big) && \text{(by definition of functional addition)} \\
&= (f_1 \circ f_3)(\bar{x}) + (f_2 \circ f_3)(\bar{x}) && \text{(again, by definition of composition)} \\
&= \big(f_1 \circ f_3 + f_2 \circ f_3\big)(\bar{x}). && \text{(again, by definition of addition)}
\end{aligned}$$

Verifying (some of) these properties (cont.)

- Let $g_1, g_2, g_3 \in L(V_0, V_0)$. In order to verify the *functional identity*

$$g_3 \circ (g_1 + g_2) = g_3 \circ g_1 + g_3 \circ g_2,$$

we need to check that, for every $\bar{x} \in V_0$, we have that

$$\big(g_3 \circ (g_1 + g_2)\big)(\bar{x}) = \big(g_3 \circ g_1 + g_3 \circ g_2\big)(\bar{x}).$$

Indeed,

$$\begin{aligned}
\big(g_3 \circ (g_1 + g_2)\big)(\bar{x}) &= g_3\big((g_1 + g_2)(\bar{x})\big) && \text{(by definition of composition)} \\
&= g_3\big(g_1(\bar{x}) + g_2(\bar{x})\big) && \text{(by definition of addition)} \\
&= g_3\big(g_1(\bar{x})\big) + g_3\big(g_2(\bar{x})\big) && \textbf{(because } g_3 \textbf{ is linear)} \\
&= (g_3 \circ g_1)(\bar{x}) + (g_3 \circ g_2)(\bar{x}) \\
&= \big(g_3 \circ g_1 + g_3 \circ g_2\big)(\bar{x}).
\end{aligned}$$

The ring $L(V_0, V_0)$

Recall that all the elements of $L(V_0, V_0)$ are linear maps, or equivalently vector space homomorphisms. Given that each such element is a function g from V_0 to itself, we also call these functions *endomorphisms* of V_0. In other words, the ring $L(V_0, V_0)$ is the ring of endomorphisms of V_0.

Recall that this is the second (type of) example of a non-commutative ring that we have seen this term: the first (type of) example was the family of rings $\mathbb{F}^{n\times n}$ with standard matrix addition and matrix multiplication.

Moreover, observe that, if $V_0 = \mathbb{F}^n$ and we consider a linear function $g : \mathbb{F}^n \to \mathbb{F}^n$, then we have said that we can associate with this function a matrix $A \in \mathbb{F}^{n\times n}$ in a natural way: A will be the matrix representation of g, or in other words the matrix that satisfies $A\bar{u} = g(\bar{u})$ for all $\bar{u} \in \mathbb{F}^n$.

We can now view this type of identification of a linear function in $L(\mathbb{F}^n, \mathbb{F}^n)$ and a matrix in $\mathbb{F}^{n\times n}$: it gives us a bijective function

$$F : L(\mathbb{F}^n, \mathbb{F}^n) \to \mathbb{F}^{n\times n}.$$

This naturally inspires the question: how different are these two rings (which are associated with the vector space $\mathbb{F}^n$) ?

The rings $L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n\times n}$

How different are these two rings? Do we absolutely need to study each of them separately?

The first thing we can observe is that both these rings have the same cardinality: this is because we have already found a bijective function F from one ring to the other.

Theorem

$L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n\times n}$ are isomorphic as rings as well.

Structure preserving functions

when the structures are fields,

or when the structures are rings

Field homomorphisms

Definition

Let $\mathbb{F}_1 = \big(\{\text{elements in } \mathbb{F}_1\}, +_1, \cdot_1\big)$ and $\mathbb{F}_2 = \big(\{\text{elements in } \mathbb{F}_2\}, +_2, \cdot_2\big)$ be two fields.

A function $f : \mathbb{F}_1 \to \mathbb{F}_2$ is called a *field homomorphism* if f satisfies the following three properties:

1. for every $a, b \in \mathbb{F}_1$, we have that
$$f(a +_1 b) = f(a) +_2 f(b).$$
2. for every $a, b \in \mathbb{F}_1$, we have that
$$f(a \cdot_1 b) = f(a) \cdot_2 f(b).$$
3. $f\big(1_{\mathbb{F}_1}\big) = 1_{\mathbb{F}_2}$.

Field homomorphisms

Some examples. (i) The *inclusion function*

$$i_1 : \mathbb{Q} \to \mathbb{R} \qquad \text{where } \; i_1(q) = q \text{ for all } q \in \mathbb{Q}$$

is a field homomorphism (in fact, it is the only field homomorphism between these two fields).

(ii) Similarly, the inclusion function

$$i_2 : \mathbb{R} \to \mathbb{C} \qquad \text{where } \; i_2(r) = r \text{ for all } r \in \mathbb{R}$$

is a field homomorphism.

However here we have other field homomorphisms too (but this requires much more advanced tools to justify).

(iii) Both the identity function $\mathrm{id}_{\mathbb{C}}$ and the function

$$\psi : \mathbb{C} \to \mathbb{C}, \qquad \psi(z) \mapsto \bar{z}$$

are field homomorphisms from $\mathbb{C}$ to itself *(because they are also bijective, we call them automorphisms of* $\mathbb{C}$*)*.

Isomorphic fields

Definition

Two fields $\mathbb{F}_1$ and $\mathbb{F}_2$ are called isomorphic if

- there exists an invertible field homomorphism $f : \mathbb{F}_1 \to \mathbb{F}_2$
- such that its inverse function $f^{-1} : \mathbb{F}_2 \to \mathbb{F}_1$ is also a field homomorphism.

In such a case, we write $\mathbb{F}_1 \cong \mathbb{F}_2$ and we call the function f a *field isomorphism* from $\mathbb{F}_1$ to $\mathbb{F}_2$.

It turns out that we can have an analogous theorem to Theorem 4 of Lecture 43, which essentially implied that *linear isomorphisms coincide with bijective linear maps*.

Theorem 4′

Let $\mathbb{F}_1$ and $\mathbb{F}_2$ be two fields, and let $f : \mathbb{F}_1 \to \mathbb{F}_2$ be **a field homomorphism which is also bijective.**

Then the inverse function $f^{-1} : \mathbb{F}_2 \to \mathbb{F}_1$ of f will be a field homomorphism too.

This shows that the second condition in the above definition is redundant, and thus it can be omitted.

Ring homomorphisms

Definition

Let $\mathcal{R}_1 = \big(\{\text{elements in } \mathcal{R}_1\}, +_1, \cdot_1\big)$ and $\mathcal{R}_2 = \big(\{\text{elements in } \mathcal{R}_2\}, +_2, \cdot_2\big)$ be two rings.

A function $g : \mathcal{R}_1 \to \mathcal{R}_2$ is called a *ring homomorphism* if g satisfies the following three properties:

1. for every $w, z \in \mathcal{R}_1$, we have that
$$g(w +_1 z) = g(w) +_2 g(z).$$
2. for every $w, z \in \mathcal{R}_1$, we have that
$$g(w \cdot_1 z) = g(w) \cdot_2 g(z).$$
3. $g\big(1_{\mathcal{R}_1}\big) = 1_{\mathcal{R}_2}$.

Isomorphic rings

Definition

Two rings $\mathcal{R}_1$ and $\mathcal{R}_2$ are called isomorphic if

- there exists an invertible ring homomorphism $g : \mathcal{R}_1 \to \mathcal{R}_2$
- such that its inverse $g^{-1} : \mathcal{R}_2 \to \mathcal{R}_1$ is also a ring homomorphism.

In such a case, we write $\mathcal{R}_1 \cong \mathcal{R}_2$ and we call g a *ring isomorphism* from $\mathcal{R}_1$ to $\mathcal{R}_2$.

It turns out that we again have an analogous theorem to the previous two theorems.

Theorem 4″

Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two rings, and let $g : \mathcal{R}_1 \to \mathcal{R}_2$ be **a ring homomorphism which is also bijective.**

Then the inverse function $g^{-1} : \mathcal{R}_2 \to \mathcal{R}_1$ of g will be a ring homomorphism too.

This shows that the second condition in the above definition is redundant, and thus it can be omitted.

Back to our primary example

Theorem

$L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n \times n}$ are isomorphic rings.

In fact, the function F which maps each linear function in $L(\mathbb{F}^n, \mathbb{F}^n)$ to its (canonical) matrix representation in $\mathbb{F}^{n \times n}$ is a ring isomorphism.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 46

Wednesday December 2

Recall from last time: Field homomorphisms

Definition

Let $\mathbb{F}_1 = \left(\{\text{elements in } \mathbb{F}_1\}, +_1, \cdot_1\right)$ and $\mathbb{F}_2 = \left(\{\text{elements in } \mathbb{F}_2\}, +_2, \cdot_2\right)$ be two fields.

A function $f : \mathbb{F}_1 \to \mathbb{F}_2$ is called a *field homomorphism* if f satisfies the following three properties:

1. for every $a, b \in \mathbb{F}_1$, we have that
$$f(a +_1 b) = f(a) +_2 f(b).$$
2. for every $a, b \in \mathbb{F}_1$, we have that
$$f(a \cdot_1 b) = f(a) \cdot_2 f(b).$$
3. $f\left(1_{\mathbb{F}_1}\right) = 1_{\mathbb{F}_2}$.

Isomorphic fields

Definition

Two fields $\mathbb{F}_1$ and $\mathbb{F}_2$ are called isomorphic if

- there exists an invertible field homomorphism $f : \mathbb{F}_1 \to \mathbb{F}_2$
- such that its inverse function $f^{-1} : \mathbb{F}_2 \to \mathbb{F}_1$ is also a field homomorphism.

In such a case, we write $\mathbb{F}_1 \cong \mathbb{F}_2$ and we call the function f a *field isomorphism* from $\mathbb{F}_1$ to $\mathbb{F}_2$.

It turns out that we can have an analogous theorem to Theorem 4 of Lecture 43, which essentially implied that *linear isomorphisms coincide with bijective linear maps*.

Theorem 4′

Let $\mathbb{F}_1$ and $\mathbb{F}_2$ be two fields, and let $f : \mathbb{F}_1 \to \mathbb{F}_2$ be **a field homomorphism which is also bijective.**

Then the inverse function $f^{-1} : \mathbb{F}_2 \to \mathbb{F}_1$ of f will be a field homomorphism too.

This shows that the second condition in the above definition is redundant, and thus it can be omitted.

Analogously: Ring homomorphisms

Definition

Let $\mathcal{R}_1 = \big(\{\text{elements in } \mathcal{R}_1\}, +_1, \cdot_1\big)$ and $\mathcal{R}_2 = \big(\{\text{elements in } \mathcal{R}_2\}, +_2, \cdot_2\big)$ be two rings.

A function $g : \mathcal{R}_1 \to \mathcal{R}_2$ is called a *ring homomorphism* if g satisfies the following three properties:

1. for every $w, z \in \mathcal{R}_1$, we have that
$$g(w +_1 z) = g(w) +_2 g(z).$$
2. for every $w, z \in \mathcal{R}_1$, we have that
$$g(w \cdot_1 z) = g(w) \cdot_2 g(z).$$
3. $g\big(1_{\mathcal{R}_1}\big) = 1_{\mathcal{R}_2}$.

Isomorphic rings

Definition

Two rings $\mathcal{R}_1$ and $\mathcal{R}_2$ are called isomorphic if

- there exists an invertible ring homomorphism $g : \mathcal{R}_1 \to \mathcal{R}_2$
- such that its inverse $g^{-1} : \mathcal{R}_2 \to \mathcal{R}_1$ is also a ring homomorphism.

In such a case, we write $\mathcal{R}_1 \cong \mathcal{R}_2$ and we call g a *ring isomorphism* from $\mathcal{R}_1$ to $\mathcal{R}_2$.

It turns out that we again have an analogous theorem to the previous two theorems.

Theorem 4″

Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two rings, and let $g : \mathcal{R}_1 \to \mathcal{R}_2$ be **a ring homomorphism which is also bijective.**

Then the inverse function $g^{-1} : \mathcal{R}_2 \to \mathcal{R}_1$ of g will be a ring homomorphism too.

This shows that the second condition in the above definition is redundant, and thus it can be omitted.

A very important example: the rings $L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n\times n}$

Theorem

$L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n\times n}$ are isomorphic rings.

In fact, the function F which maps each linear function in $L(\mathbb{F}^n, \mathbb{F}^n)$ to its (canonical) matrix representation in $\mathbb{F}^{n\times n}$ is a ring isomorphism.

Proof. We need to verify that

(i) F maps the multiplicative identity of $L(\mathbb{F}^n, \mathbb{F}^n)$ to the multiplicative identity of $\mathbb{F}^{n\times n}$;

(ii) F respects addition;

(iii) F respects 'multiplication'.

Regarding (i), we note that the multiplicative identity of $L(\mathbb{F}^n, \mathbb{F}^n)$ is $\mathrm{id}_{\mathbb{F}^n}$, while the multiplicative identity of $\mathbb{F}^{n\times n}$ is the identity matrix I_n.

Clearly, the matrix representation of $\mathrm{id}_{\mathbb{F}^n}$ is I_n, so (i) holds.

Proof of the theorem (cont.)

Regarding (ii), let us consider two arbitrary linear functions f, g in $L(\mathbb{F}^n, \mathbb{F}^n)$, and let us also consider their respective matrix representations A_f and A_g. In other words, we have

$$f(\bar{u}) = A_f \bar{u} \qquad \text{and} \qquad g(\bar{u}) = A_g \bar{u}$$

for all $\bar{u} \in \mathbb{F}^n$.

But then, for every such vector $\bar{u}$, we can write

$$(f + g)(\bar{u}) = f(\bar{u}) + g(\bar{u}) = A_f \bar{u} + A_g \bar{u} = (A_f + A_g)\bar{u}.$$

This shows that the matrix representation of the function $f + g$ is the matrix $A_f + A_g$, and hence we do have

$$F(f + g) = A_f + A_g = F(f) + F(g).$$

Since $f, g \in L(\mathbb{F}^n, \mathbb{F}^n)$ were arbitrary, we see that (ii) holds true as well.

Proof of the theorem (cont.)

Finally, regarding (iii), let us consider again two arbitrary linear functions f, g in $L(\mathbb{F}^n, \mathbb{F}^n)$, and their respective matrix representations A_f and A_g. In other words, we have

$$f(\bar{u}) = A_f \bar{u} \qquad \text{and} \qquad g(\bar{u}) = A_g \bar{u}$$

for all $\bar{u} \in \mathbb{F}^n$.

But then, for every vector $\bar{w} \in \mathbb{F}^n$, we can write

$$\begin{aligned}
(f \circ g)(\bar{w}) &= f\big(g(\bar{w})\big) \\
&= f\big(A_g \bar{w}\big) && \text{(by the above applied with } \bar{u} = \bar{w}) \\
&= A_f\big(A_g \bar{w}\big) && \text{(by the above applied with } \bar{u} = A_g \bar{w}) \\
&= (A_f \cdot A_g)\bar{w}.
\end{aligned}$$

This shows that the matrix representation of the function $f \circ g$ is the matrix $A_f \cdot A_g$, and hence we do have

$$F(f \circ g) = A_f \cdot A_g = F(f) \cdot F(g).$$

Since $f, g \in L(\mathbb{F}^n, \mathbb{F}^n)$ were arbitrary, we see that (iii) holds true as well.

Very Important Remark

By the theorem we just proved, we know that the rings $L(\mathbb{F}^n, \mathbb{F}^n)$ and $\mathbb{F}^{n\times n}$ behave in the exact same way (they are essentially the same ring, it's just that the elements are represented in a different way, or in a sense have a different 'name' and 'shape', in the second structure compared to the first structure).

On the other hand, if V_0 is not a vector space of the form $\mathbb{F}^n$, or, even more generally, if V_0 is not isomorphic to a vector space of the form $\mathbb{F}^n$, then $L(V_0, V_0)$ can have quite unusual properties compared to the properties we know $\mathbb{F}^{n\times n}$ has.

One primary example here. In HW5, you are asked to prove that, if C, D are matrices in $\mathbb{F}^{n\times n}$, and we know that $CD = I_n$, then we can conclude that C is invertible (and as a consequence D is also invertible).

In other words, if a square matrix C has a right inverse, then it has a two-sided inverse, and hence it is invertible.

We will now give an example of a vector space V_0 and an element $f \in L(V_0, V_0)$ such that f has a right inverse, but nevertheless f is not invertible.

Linear functions on polynomials

Consider the space $\mathcal{P}$ of real polynomials. In HW6, you are asked to verify that this is a vector space over $\mathbb{R}$ (moreover, you are asked to show that its dimension over $\mathbb{R}$ is infinite, and hence $\mathcal{P}$ cannot be isomorphic to any vector space of the form $\mathbb{F}^n$).

Set f_d to be the *derivative operator* on $\mathcal{P}$. That is, if $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ is a polynomial in $\mathcal{P}$, then

$$f_d(p) := a_1 + 2a_2x + 3a_3x^2 + \cdots + ma_mx^{m-1}.$$

Similarly, set f_i to be the *integral operator* on $\mathcal{P}$. That is, if $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ is a polynomial in $\mathcal{P}$, then

$$f_i(p) := a_0x + \frac{a_1}{2}x^2 + \frac{a_2}{3}x^3 + \cdots + \frac{a_m}{m+1}x^{m+1}.$$

Observe that both f_d and f_i are linear functions from $\mathcal{P}$ to $\mathcal{P}$, and hence they are contained in $L(\mathcal{P}, \mathcal{P})$.

Question. Is the derivative operator f_d an injective linear function? Or in other words, what is $\mathrm{Ker}(f_d)$?

The answer to the first question is NO. *(Can you justify this?)*

This shows that the derivative operator f_d cannot be **not** invertible (since it fails to be injective). However, we have that

$$f_d \circ f_i = \mathrm{id}_{\mathcal{P}}$$

(why?), and hence f_d has a right inverse (it also follows that f_i has a left inverse, but this function is **not** invertible either; in this case, check that f_i fails to be surjective).

Field homomorphisms are particularly nice

An example of a ring homomorphism that we have essentially already studied is the function

$$h_m : \mathbb{Z} \to \mathbb{Z}_m, \qquad n \in \mathbb{Z} \ \mapsto \ [n]_m \in \mathbb{Z}_m,$$

where $m \geqslant 2$ is some fixed positive integer, and where $[n]_m$ is the congruence class of the integer n modulo m *(check that this is indeed a ring homomorphism).*
Observe that this ring homomorphism is not injective.

On the other hand, field homomorphisms are always much nicer...

Theorem

Let $\mathbb{F}_1, \mathbb{F}_2$ be two fields, and let $f : \mathbb{F}_1 \to \mathbb{F}_2$ be a field homomorphism.

Then f is necessarily injective.

Justification of the theorem

Let $a, b \in \mathbb{F}_1$ with $a \neq b$. We need to show that $f(a) \neq f(b)$.

Observe that $a \neq b \Leftrightarrow a - b \neq 0_{\mathbb{F}_1}$. Thus $a - b$ has a multiplicative inverse $(a-b)^{-1}$ in $\mathbb{F}_1$. But then, because of the defining properties of a field homomorphism, we can write

$$1_{\mathbb{F}_2} = f(1_{\mathbb{F}_1}) = f\big((a-b)\cdot(a-b)^{-1}\big) = f(a-b)\cdot f\big((a-b)^{-1}\big).$$

This shows that $f(a-b)\cdot f\big((a-b)^{-1}\big) \neq 0_{\mathbb{F}_2}$, and hence $f(a-b)$ must be a non-zero element of $\mathbb{F}_2$.

At the same time, again by the definition of field homomorphism, we have that

$$f(a-b) = f(a+(-b)) = f(a) + f(-b) = f(a) - f(b),$$

where the third equality follows from another basic property of field homomorphisms (that they respect additive inverses), which is a consequence of the definition (see HW6, Problem 2(ii)).
Combining the above, we see that $0_{\mathbb{F}_2} \neq f(a-b) = f(a) - f(b)$, which gives that $f(a) \neq f(b)$.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 47

Friday December 4

Notion of 'Linear Independence'

Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

(I) Let $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k \in V$ be vectors (not necessarily different). We will call $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ *linearly independent* if the following holds true:

whenever we have $\lambda_1 \bar{u}_1 + \lambda \bar{u}_2 + \cdots + \lambda_k \bar{u}_k = \bar{0}_V$ for some scalars $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, then necessarily $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0_{\mathbb{F}}$.

(II) Let $S = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k\}$ be a set of vectors from V. The set S is called *linearly independent* if the vectors $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$ are linearly independent.

Subtle difference between the two parts of the definition: When we consider a set of k vectors, then these k vectors are all different (since we don't consider repetitions of elements in a set, and we ignore these repetitions even if they are included in our notation, e.g. when we write $\mathbb{Q} = \{\frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0\}$).

— However, it makes sense in certain settings to ask whether k vectors, not necessarily different, are linearly independent (see next example), that's why we give both parts of the definition.

Examples and non-examples

Consider the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 4}$.

It makes sense in several cases to ask whether the rows of a matrix are linearly independent or not, and similarly whether the columns of a matrix are linearly independent or not (of course, as in the example here, there is no guarantee that these rows (or these columns) will all be different vectors).

In this specific example, the rows of A are **not** linearly independent: indeed, note that

$$\bar{0}_{\mathbb{R}^4} = 1 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 3 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 2 \\ -1 \\ 1 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 3 \end{pmatrix}$$

but the scalars in this linear combination are **not** all zero.

Terminology. In such a case, we will say that the given vectors are *linearly dependent*.

Back to the definition of 'Linear Independence'

Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

(I) Let $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k \in V$ be vectors (not necessarily different). We will call $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ *linearly independent* if the following holds true:

whenever we have $\lambda_1 \bar{u}_1 + \lambda \bar{u}_2 + \cdots + \lambda_k \bar{u}_k = \bar{0}_V$ for some scalars $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, then necessarily $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0_{\mathbb{F}}$.

(II) Let $S = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k\}$ be a set of vectors from V. The set S is called *linearly independent* if the vectors $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$ are linearly independent.

(III) Finally, if $T \subseteq V$ is a possibly infinite set, then we say that T is *linearly independent* if, for every $k \geqslant 1$, and for every k **different** vectors $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_k$ from T, we have that $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_k$ are linearly independent.

In other words, T is linearly independent if every finite subset S_k of T is linearly independent.

One example where part (III) of the definition is relevant: In HW6, Problem 5, you are essentially encouraged to show that the set

$$\{1, x, x^2, \ldots, x^n, x^{n+1}, \ldots\ldots\}$$

of all monomials is a linearly independent subset of the vector space $\mathcal{P}$ of real polynomials.

Equivalent conditions for linear independence

Proposition 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Consider vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ in V.
Then the vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ are linearly independent **if and only if** none of the vectors $\bar{u}_i$ is a linear combination of the remaining vectors $\bar{u}_j$, $j \neq i$.

Proof. Observe that we have to show the equivalence

$$\text{Statement 1} \Leftrightarrow \text{Statement 2},$$

where

Statement 1: the vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ are linearly independent,

and Statement 2: for every $i \in \{1, 2, \ldots, k\}$, the vector $\bar{u}_i$ is **not** a linear combination of the vectors $\bar{u}_j$, $j \neq i$.

Recall that an equivalence involves two implications: in this case, the implications Statement 1 $\Rightarrow$ Statement 2 and Statement 2 $\Rightarrow$ Statement 1.

For this proof, it will be easier to show the **contrapositives** of these implications, which is equivalent to what we were asked to show; in other words, we will now try to show

$\neg$ Statement 2 $\Rightarrow$ $\neg$ Statement 1 and $\neg$ Statement 1 $\Rightarrow$ $\neg$ Statement 2.

Proof of Proposition 1

- $\neg$ **Statement 1** $\Rightarrow$ $\neg$ **Statement 2**.

First, we need to write down precisely what the negations of the given statements are. We have that

$\neg$ Statement 1: the vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ are linearly dependent,

or in other words, there exist scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$, not all of them zero, such that

$$\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \cdots + \lambda_k \bar{u}_k = \bar{0}_V.$$

We also have that

$\neg$ Statement 2: there exists some $i \in \{1, 2, \ldots, k\}$ such that the vector $\bar{u}_i$ can be written as a linear combination of the remaining vectors $\bar{u}_j$, $j \neq i$,

or in other words, we can find scalars μ_j, $j \neq i$, such that $\bar{u}_i = \sum_{j \neq i} \mu_j \bar{u}_j$.

Assume now that $\neg$ Statement 1 is true. Then, we can indeed find scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$, not all of them zero, such that $\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \cdots + \lambda_k \bar{u}_k = \bar{0}_V$. Since not all λ_i are zero, we can find $i_0 \in \{1, 2, \ldots, k\}$ such that $\lambda_{i_0} \neq 0$. But then

$$\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \cdots + \lambda_{i_0} \bar{u}_{i_0} + \cdots + \lambda_k \bar{u}_k = \bar{0}_V$$

$$\Rightarrow \quad \lambda_{i_0} \bar{u}_{i_0} = -\sum_{j \neq i_0} \lambda_j \bar{u}_j$$

$$\Rightarrow \quad \bar{u}_{i_0} = \lambda_{i_0}^{-1} \cdot (\lambda_{i_0} \bar{u}_{i_0}) = \lambda_{i_0}^{-1} \cdot \left(-\sum_{j \neq i_0} \lambda_j \bar{u}_j \right) = \sum_{j \neq i_0} (-\lambda_{i_0}^{-1} \cdot \lambda_j)\, \bar{u}_j.$$

Thus we managed to write the vector $\bar{u}_{i_0}$ as a linear combination of the remaining vectors $\bar{u}_j$.

Proof of Proposition 1 (cont.)

- $\neg$ **Statement 2** $\Rightarrow$ $\neg$ **Statement 1** .

Assume now that $\neg$ Statement 2 is true. Then we can find $i \in \{1, 2, \ldots, k\}$ such that $\bar{u}_i$ is a linear combination of the remaining $\bar{u}_j$, or in other words, such that

$$\bar{u}_i = \sum_{j \neq i} \mu_j \bar{u}_j$$

for some scalars μ_j, $j \neq i$.

But then we can write

$$\bar{u}_i - \left(\sum_{j \neq i} \mu_j \bar{u}_j \right) = \bar{0}_V ,$$

and hence

$$\bar{0}_V = \lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \cdots + \lambda_{i-1} \bar{u}_{i-1} + \lambda_i \bar{u}_i + \lambda_{i+1} \bar{u}_{i+1} + \cdots + \lambda_k \bar{u}_k$$

where $\lambda_j = -\mu_j$ for $j \neq i$, while $\lambda_i = 1$ (this guarantees that not all scalars in the last linear combination are zero).

In other words, we conclude that $\neg$ Statement 1 is also true.

Back to examples and non-examples

Let V be a vector space over a field $\mathbb{F}$.

• Assume that $S_1 = \{\bar{u}\}$ is a subset of V that contains a single element. That is, assume that S_1 is a *singleton*.

Then: (i) if $\bar{u} = \bar{0}$, S_1 is linearly dependent *(indeed, we have that $1 \cdot \bar{0} = \bar{0}$, with the scalar here being non-zero).*

(ii) if $\bar{u} \neq \bar{0}$, S_1 is linearly independent *(why? try to justify this).*

• The empty set $\emptyset$, viewed as a subset of V, is linearly independent *(why?* note that it does not contain any vectors which are linear combinations of other vectors in the set).

• If $S_2 = \{\bar{u}_1, \bar{u}_2\}$, then S_2 is linearly independent **if and only if** none of the vectors $\bar{u}_1, \bar{u}_2$ is a scalar multiple of the other vector. In other words, if S_2 does not contain parallel vectors.

• More generally, if a subset S of V contains a pair of parallel vectors, then S is linearly dependent.

Attention! The converse is not necessarily true: we can find linearly dependent sets of vectors, with no two vectors in the set being parallel (see some of the following examples).

Back to examples and non-examples

- Recall the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 4}$.

We already saw that the rows of A are linearly dependent (and in this case we could check this easily, because two of the rows of A are parallel).

On the other hand, notice that no two of the columns of A are parallel. However, we still have that the columns of A are linearly dependent. Indeed,

$$\operatorname{span}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix}\right) = \operatorname{span}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix} - \begin{pmatrix}1\\0\\1\end{pmatrix}\right) \quad \text{(similar to HW6, Pb4)}$$

$$= \operatorname{span}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\2\\0\end{pmatrix}\right),$$

and thus

$$\begin{pmatrix}1\\-1\\1\end{pmatrix} \in \operatorname{span}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix}\right), \quad \text{and similarly} \quad \begin{pmatrix}3\\1\\3\end{pmatrix} \in \operatorname{span}\left(\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix}\right).$$

On the other hand, $\mathrm{Col}_1(A)$ and $\mathrm{Col}_2(A)$ are not parallel, thus the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ is linearly independent.

Back to examples and non-examples

In fact, we can now make the following observations about the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ based on what we just showed:

- it is a linearly independent set;
- it is a *maximal* linearly independent subset of the columns of A (that is, any other subset of the columns of A which contains $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ will be linearly dependent).
- Even more interestingly, the cardinality of $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ is the largest possible cardinality of a **linearly independent subset of the columns of** A.

Given a vector space V, we will now see that it is important to know what the largest possible cardinality of a linearly independent subset of V is.

In fact, this is one way of defining the *dimension* of a vector space.

Notion of 'Dimension'

1st Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Then the *dimension* of the vector space V **over** $\mathbb{F}$, which we will denote by $\dim_{\mathbb{F}} V$, is the largest possible cardinality of a linearly independent subset of V.

Important Remark. It is very crucial to specify over which scalar field we consider the dimension of V, when V can be viewed as a vector space over more than one field.

Example. Recall that $\mathbb{C} \equiv \mathbb{C}^1$ is a vector space over itself, and at the same time $\mathbb{C}$, when identified with $\mathbb{R}^2$, can be viewed as a vector space over $\mathbb{R}$.

We will see that $\dim_{\mathbb{C}} \mathbb{C} = 1$, while $\dim_{\mathbb{R}} \mathbb{C} = 2$.

When can we extend linearly independent sets?

In other words, when is a linearly independent set **not maximal**?

We have the following theorem providing a criterion.

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Suppose that $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}$ is a linearly independent subset of V, and suppose that

$$\text{there exists } \bar{w} \in V \text{ such that } \bar{w} \notin \operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big).$$

Then the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ is also linearly independent.

We will give the proof of this theorem, and also see immediate consequences of it, in the next lecture.

MATH 127 – Honours Linear Algebra I
Fall Term 2020

Notes for Lecture 48

Monday December 7

Notion of 'Linear Independence'

Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

(I) Let $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k \in V$ be vectors (not necessarily different). We will call $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ *linearly independent* if the following holds true:

whenever we have $\lambda_1 \bar{u}_1 + \lambda \bar{u}_2 + \cdots + \lambda_k \bar{u}_k = \bar{0}_V$ for some scalars $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, then necessarily $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0_{\mathbb{F}}$.

(II) Let $S = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k\}$ be a set of vectors from V. The set S is called *linearly independent* if the vectors $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$ are linearly independent.

(III) Finally, if $T \subseteq V$ is a possibly infinite set, then we say that T is *linearly independent* if, for every $k \geqslant 1$, and for every k **different** vectors $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_k$ from T, we have that $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_k$ are linearly independent.

In other words, T is linearly independent if every finite subset S_k of T is linearly independent.

Equivalent conditions for linear independence

Proposition 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Consider vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ in V.

Then the vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k$ are linearly independent **if and only if** none of the vectors $\bar{u}_i$ is a linear combination of the remaining vectors $\bar{u}_j$, $j \neq i$.

Equivalent conditions for linear independence (cont.)

Proposition 2

Let $\mathbb{F}$ be a field, and let $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$ be vectors of the vector space $\mathbb{F}^n$ (viewed as a vector space over $\mathbb{F}$).

Consider the matrix $A \in \mathbb{F}^{n \times k}$ whose columns are the vectors $\bar{v}_i$ (more precisely, the 1st column of A is the vector $\bar{v}_1$, the 2nd column of A is the vector $\bar{v}_2$, and so on).

Then the vectors $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$ are linearly independent **if and only if** the homogeneous linear system $A\bar{x} = \bar{0}_{\mathbb{F}^n}$ has a unique solution (or in other words its only solution is the trivial solution).

Proof of Proposition 2

Recall that, if $\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} \in \mathbb{F}^k$, then we can write

$$A \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \lambda_1 \mathrm{Col}_1(A) + \lambda_2 \mathrm{Col}_2(A) + \cdots + \lambda_k \mathrm{Col}_k(A)$$

$$= \lambda_1 \bar{v}_1 + \lambda_2 \bar{v}_2 + \cdots + \lambda_k \bar{v}_k.$$

Thus $\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}$ is a solution to the system $A\bar{x} = \bar{0}_{\mathbb{F}^n}$ **if and only if**

$A \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \bar{0}_{\mathbb{F}^n}$ **if and only if** $\lambda_1 \bar{v}_1 + \lambda_2 \bar{v}_2 + \cdots + \lambda_k \bar{v}_k = \bar{0}_{\mathbb{F}^n}$.

But then, if we assume that the vectors $\bar{v}_i$ are linearly independent, we will have that $\lambda_1 \bar{v}_1 + \lambda_2 \bar{v}_2 + \cdots + \lambda_k \bar{v}_k = \bar{0}_{\mathbb{F}^n}$ is true **only when all the λ_i are equal to 0,** $\Rightarrow$ only the trivial solution solves the system $A\bar{x} = \bar{0}_{\mathbb{F}^n}$.

Conversely, if this system has only the trivial solution, then $\lambda_1 \bar{v}_1 + \lambda_2 \bar{v}_2 + \cdots + \lambda_k \bar{v}_k = \bar{0}_{\mathbb{F}^n}$ is true only if all scalars are zero, and hence the vectors $\bar{v}_i$ are linearly independent.

Maximal linearly independent sets

Definition

Let V be a vector space over a field $\mathbb{F}$. We say that a linearly independent subset S of V is a *maximal* linearly independent set if every larger subset T of V that contains S is **not** linearly independent.

More simply, we say that S is a *maximal* linearly independent subset of V if, for every $\bar{w} \in V \setminus S$, we have that the set $S \cup \{\bar{w}\}$ is **not** linearly independent.

Some examples

- In the vector space $\mathbb{F}^n$, the standard basis vectors form a maximal linearly independent set: indeed, if $\bar{w} \in \mathbb{F}^n \setminus \{\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n\}$, then the set

$$\{\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n, \bar{w}\}$$

  is not linearly independent, because one of its vectors can be written as a linear combination of the remaining vectors (more precisely, $\bar{w}$ can be written as a linear combination of the standard basis vectors).

- In the vector space $\mathcal{P}$ of real polynomials, the set

$$\mathrm{Mon}_{\mathcal{P}} = \{1, x, x^2, \ldots, x^n, x^{n+1}, \ldots\ldots\}$$

  is a maximal linearly independent set: indeed, if $p(x) \in \mathcal{P} \setminus \mathrm{Mon}_{\mathcal{P}}$, then $p(x)$ can be written as a linear combination of (finitely many) monomials. In other words, we will be able to write one vector of the set $\mathrm{Mon}_{\mathcal{P}} \cup \{p(x)\}$ as a linear combination of the remaining vectors, and hence the set $\mathrm{Mon}_{\mathcal{P}} \cup \{p(x)\}$ will not be linearly independent.

- Last time we saw that the first two columns of the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 4}$ span the entire Column Space of A. Thus the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ is a maximal linearly independent subset of the Column Space of A (because for every vector $\bar{w} \in \mathrm{CS}(A) \setminus \{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$, we will have that $\bar{w}$ can be written as a linear combination of $\mathrm{Col}_1(A)$ and $\mathrm{Col}_2(A)$, and hence the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A), \bar{w}\}$ will **not** be linearly independent).

Maximal linearly independent sets and dimension

Recall from last time...

1st Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Then the *dimension* of the vector space V **over** $\mathbb{F}$, which we will denote by $\dim_{\mathbb{F}} V$, is the largest possible cardinality of a linearly independent subset of V.

Remark. Clearly, it suffices to consider maximal linearly independent subsets of V when trying to determine the largest possible cardinality of a linearly independent set.

When can we extend linearly independent sets?

In other words, when is a linearly independent set **not maximal**?

We have the following theorem providing a criterion.

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Suppose that $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}$ is a linearly independent subset of V, and suppose that

$$\text{there exists } \bar{w} \in V \text{ such that } \bar{w} \notin \operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big).$$

Then the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ is also linearly independent.

Corollary of Theorem 1

A linearly independent subset S of V is maximal **if and only if**

$$\operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big) = V.$$

Proof of Theorem 1

We need to show that, since $\bar{w} \notin \text{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big)$, the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ will be linearly independent.

Assume towards a contradiction that we can find scalars $\mu_1, \mu_2, \ldots, \mu_k, \mu_{k+1}$, **not all of them zero**, such that

$$\mu_1 \bar{u}_1 + \mu_2 \bar{u}_2 + \cdots + \mu_k \bar{u}_k + \mu_{k+1} \bar{w} = \bar{0}_V.$$

We distinguish two cases:

Case 1: $\mu_{k+1} = 0$. Then (at least) one of the first k scalars, $\mu_1, \mu_2, \ldots, \mu_k$, must be non-zero. At the same time, we can write

$$\mu_1 \bar{u}_1 + \mu_2 \bar{u}_2 + \cdots + \mu_k \bar{u}_k = \bar{0}_V.$$

But this contradicts the given assumption that the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}$ is linearly independent. Thus Case 1 cannot hold.

Case 2: $\mu_{k+1} \neq 0$. Then we can write

$$\begin{aligned} & \mu_1 \bar{u}_1 + \mu_2 \bar{u}_2 + \cdots + \mu_k \bar{u}_k + \mu_{k+1} \bar{w} = \bar{0}_V \\ \Rightarrow \quad & \mu_{k+1} \bar{w} = -\mu_1 \bar{u}_1 - \mu_2 \bar{u}_2 - \cdots - \mu_k \bar{u}_k \\ \Rightarrow \quad & \bar{w} = (-\mu_{k+1}^{-1} \mu_1) \bar{u}_1 + (-\mu_{k+1}^{-1} \mu_2) \bar{u}_2 + \cdots + (-\mu_{k+1}^{-1} \mu_k) \bar{u}_k. \end{aligned}$$

But the latter contradicts the given assumption that $\bar{w} \notin \text{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big)$. Thus Case 2 cannot hold either.

We just saw that we reach a contradiction in all cases, which shows that the assumption that the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ is **not** linearly independent was incorrect.

Back to Theorem 1 and its Corollary

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Suppose that $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}$ is a linearly independent subset of V, and suppose that

$$\text{there exists } \bar{w} \in V \text{ such that } \bar{w} \notin \operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big).$$

Then the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ is also linearly independent.

Corollary of Theorem 1

A linearly independent subset S of V is maximal **if and only if**

$$\operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big) = V.$$

Spanning sets and bases

Definitions

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

- A subset S of V is called a *spanning set* of V if

$$V = \operatorname{span}(S).$$

- A subset $\mathcal{B}$ of V is called a *basis* of V if $\mathcal{B}$ is both a spanning set of V and linearly independent.

Back to Theorem 1 and its Corollary

Theorem 1

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$. Suppose that $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}$ is a linearly independent subset of V, and suppose that

$$\text{there exists } \bar{w} \in V \text{ such that } \bar{w} \notin \operatorname{span}\big(\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k\}\big).$$

Then the set $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_k, \bar{w}\}$ is also linearly independent.

Restating of the Corollary of Theorem 1

A linearly independent subset S of V is maximal **if and only if** S is a basis of V.

In other words, the maximal linearly independent subsets of a vector space V are precisely the different bases of V.

Most Important Theorem about the notion of 'Dimension'

Theorem A

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Every basis of V (that is, every maximal linearly independent subset of V) has the same cardinality.

Necessary ingredient for the proof of Theorem A

Recall the following theorem about homogeneous linear systems:

Theorem 3 of Lecture 39

Let $\mathbb{F}$ be a field, and let $\mathcal{LS}_0$ be a homogeneous underdetermined system of linear equations with coefficients from $\mathbb{F}$.

Then $\mathcal{LS}_0$ has more than one solution, or in other words it has solutions besides the trivial one. In fact, its solution set is at least as big as the set $\mathbb{F}$.

Proof of Theorem A

We will only consider the case of a vector space V for which we already know there exists a finite basis $\mathcal{B}_0 = \{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m\}$ (for vector spaces whose bases are infinite sets, the proof requires more advanced tools which are beyond the scope of this course).

Let $\mathcal{C}$ be another basis of V. We need to show that $|\mathcal{C}| = |\mathcal{B}_0|$.
Assume towards a contradiction that $|\mathcal{C}| \neq |\mathcal{B}_0|$.

We distinguish two cases.

Case 1: $|\mathcal{C}| > |\mathcal{B}_0|$. Then we can definitely find a subset T of $\mathcal{C}$ which contains $m+1$ vectors, say $T = \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_m, \bar{w}_{m+1}\}$.

Also, since T is a subset of a linearly independent set, then T is also linearly independent. Thus, we have that, for every $\lambda_1, \lambda_2, \ldots, \lambda_m, \lambda_{m+1} \in \mathbb{F}$,

$$\lambda_1 \bar{w}_1 + \lambda_2 \bar{w}_2 + \cdots + \lambda_m \bar{w}_m + \lambda_{m+1} \bar{w}_{m+1} = \bar{0}_V$$

implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_m = \lambda_{m+1} = 0_{\mathbb{F}}$.

Proof of Theorem A

Case 1: $|\mathcal{C}| > |\mathcal{B}_0|$. Then we can definitely find a subset T of $\mathcal{C}$ which contains $m+1$ vectors, say $T = \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_m, \bar{w}_{m+1}\}$.
Also, since T is a subset of a linearly independent set, then T is also linearly independent. Thus, we have that, for every $\lambda_1, \lambda_2, \ldots, \lambda_m, \lambda_{m+1} \in \mathbb{F}$,

$$\lambda_1 \bar{w}_1 + \lambda_2 \bar{w}_2 + \cdots + \lambda_m \bar{w}_m + \lambda_{m+1} \bar{w}_{m+1} = \bar{0}_V$$

implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_m = \lambda_{m+1} = 0_{\mathbb{F}}$.
Now, since $\mathcal{B}_0 = \{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m\}$ is a basis of V, we have that $V = \operatorname{span}(\mathcal{B}_0)$. Therefore, for every $j \in \{1, 2, \ldots, m+1\}$, we have that $\bar{w}_j \in \operatorname{span}(\mathcal{B}_0)$. In other words, we can find scalars $a_{i,j}$ such that

$$a_{1,j} \bar{u}_1 + a_{2,j} \bar{u}_2 + \cdots + a_{m,j} \bar{u}_m = \bar{w}_j.$$

Consider the matrix $A = (a_{i,j})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant m+1}}$ we form using all these scalars (where the scalars we need to write the vector $\bar{w}_j$ as a linear combination of the vectors $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m$ are the entries of the j-th column of the matrix).

Then the system $A\bar{x} = \bar{0}_{\mathbb{F}^m}$ is homogeneous and underdetermined, and therefore it will have a solution different from the trivial solution. That is, we should be able to find $\mu_1, \mu_2, \ldots, \mu_{m+1} \in \mathbb{F}$, not all of them zero, such that

$$A \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} = \bar{0}_V.$$

Finishing the proof of Case 1 of Theorem A

We observe that $A \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} = \bar{0}_V$ is equivalent to having

$$\begin{aligned}
\mu_1 a_{1,1} + \mu_2 a_{1,2} + \cdots + \mu_{m+1} a_{1,m+1} &= 0 \\
\mu_1 a_{2,1} + \mu_2 a_{2,2} + \cdots + \mu_{m+1} a_{2,m+1} &= 0 \\
&\vdots \\
\mu_1 a_{m,1} + \mu_2 a_{m,2} + \cdots + \mu_{m+1} a_{m,m+1} &= 0
\end{aligned}$$

which in turn implies that

$$\begin{array}{ccccccccc}
\mu_1 a_{1,1} \bar{u}_1 & + & \mu_2 a_{1,2} \bar{u}_1 & + & \cdots & + & \mu_{m+1} a_{1,m+1} \bar{u}_1 & = & \bar{0} \\
+ & & + & & + & & + & & \\
\mu_1 a_{2,1} \bar{u}_2 & + & \mu_2 a_{2,2} \bar{u}_2 & + & \cdots & + & \mu_{m+1} a_{2,m+1} \bar{u}_2 & = & \bar{0} \\
+ & & + & & + & & + & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
+ & & + & & + & & + & & \\
\mu_1 a_{m,1} \bar{u}_m & + & \mu_2 a_{m,2} \bar{u}_m & + & \cdots & + & \mu_{m+1} a_{m,m+1} \bar{u}_m & = & \bar{0}
\end{array}$$

Note now that, by first adding the entries / scalar multiples in the j-th column of the above array, we get the expression

$$\mu_j a_{1,j} \bar{u}_1 + \mu_j a_{2,j} \bar{u}_2 + \cdots + \mu_j a_{m,j} \bar{u}_m = \mu_j \bar{w}_j.$$

$$\begin{array}{ccccccccc}
\mu_1 a_{1,1}\bar{u}_1 & + & \mu_2 a_{1,2}\bar{u}_1 & + & \cdots & + & \mu_{m+1} a_{1,m+1}\bar{u}_1 & = & \bar{0} \\
+ & & + & & + & & + & & \\
\mu_1 a_{2,1}\bar{u}_2 & + & \mu_2 a_{2,2}\bar{u}_2 & + & \cdots & + & \mu_{m+1} a_{2,m+1}\bar{u}_2 & = & \bar{0} \\
+ & & + & & + & & + & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
+ & & + & & + & & + & & \\
\mu_1 a_{m,1}\bar{u}_m & + & \mu_2 a_{m,2}\bar{u}_m & + & \cdots & + & \mu_{m+1} a_{m,m+1}\bar{u}_m & = & \bar{0}
\end{array}$$

By first adding the entries / scalar multiples in the j-th column of the above array, we get the expression

$$\mu_j a_{1,j}\bar{u}_1 + \mu_j a_{2,j}\bar{u}_2 + \cdots + \mu_j a_{m,j}\bar{u}_m = \mu_j \bar{w}_j.$$

Thus if we add all the left-hand sides of the above equations, the result will equal the linear combination

$$\mu_1 \bar{w}_1 + \mu_2 \bar{w}_2 + \cdots + \mu_{m+1} \bar{w}_m.$$

At the same time, it will equal the sum of the right-hand sides, so it will be equal to $\bar{0}$.

Since not all scalars μ_j are zero, this contradicts the assumption that $T = \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_{m+1}\}$ is linearly independent, and it shows that Case 1 cannot hold, or in other words we cannot have $|\mathcal{C}| > |\bar{B}_0|$.

Case 2: $|\mathcal{B}_0| > |\mathcal{C}|$. This case is completely analogous, it's just that now we need to interchange the roles of $\mathcal{C}$ and $\mathcal{B}_0$ (that is, in this case the columns of the matrix we will construct will correspond to the vectors in $\mathcal{B}_0$ instead).

Thus, this case also leads to a contradiction.

In the end, we must have $|\mathcal{C}| = |\mathcal{B}_0|$.

Notion of 'Dimension'

Theorem A allows us to restate the definition of 'dimension of a vector space', and give a 2nd (equivalent) definition which is much easier to work with.

2nd Definition

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$.

Then the *dimension* of the vector space V **over** $\mathbb{F}$ coincides with the cardinality of any basis of V.

Using this in standard examples of vector spaces?

- By now, it is clear that a basis of the vector space $\mathbb{F}^n$ is the standard basis $\{\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n\}$ (indeed, we can check that this set is linearly independent, and we mentioned earlier that it is a maximal linearly independent set).

  By the definition of 'dimension', we now see that $\dim_{\mathbb{F}} \mathbb{F}^n = n$.

- Consider the vector space $\mathcal{P}_m$ of real polynomials with degree $\leqslant m$: that is, $\mathcal{P}_m = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m : a_0, a_1, a_2, \ldots, a_m \in \mathbb{R}\}$. We can check that this is a vector space over $\mathbb{R}$ (see also HW6 for a very closely related question), and we can also check that a basis of this vector space is the set

  $$\{1, x, x^2, \ldots, x^m\}.$$

  Since this basis contains $m+1$ vectors, we obtain that $\dim_{\mathbb{R}} \mathcal{P}_m = m+1$.

- We saw that a maximal linearly independent subset of the Column Space of the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 4}$ is the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$. In other words, the set $\{\mathrm{Col}_1(A), \mathrm{Col}_2(A)\}$ is a basis of $\mathrm{CS}(A)$, and hence $\dim_{\mathbb{R}} \mathrm{CS}(A) = 2$.