# Linear Transformations

Evan Halloran

## Abstract

[1] The theory of linear transformations is the only "solved" field of mathematics—that is to say, every question formulated within the confines of linear algebra has already been solved. Because of this, it is often said that if one can reduce a problem in a different field (e.g. algebra, analysis, partial differential equations) to that of linear algebra, then a solution will emerge with ease. Linearization as a technique is ubiquitous throughout mathematics, and facility with linear algebra will pay dividends. These notes will cover both the basic computations of linear algebra as well as the theory of linear transformations as a whole.

[1] The content of these notes is largely based on *Linear Algebra* by Insel, Spence, and Friedberg.

# Contents

# Chapter 1

# Preliminaries

## 1.1 Sets

**Definition** (Georg Cantor, 1895)**.** A **set** is any collection $M$ of definite, distinguishable objects $m$ of our perception or thought conceived of as a whole. The objects $m$ of $M$ are called the **elements** of $M$.

> "Nowadays it is known to be possible, logically speaking, to derive practically the whole known mathematics from a single source, namely the Theory of Sets."
> — Nicolas Bourbaki, 1950

**Notation.** $M = \{m \mid \text{defining property for } m \text{ to belong to } M\}$. If $m$ is an element of $M$, we write $m \in M$ and say $m$ belongs to $M$. Otherwise, we write $m \notin M$.

**Examples.**

$$
\begin{aligned}
\mathbb{N} = \{1,\ 2,\ 3,\ 4,\ ...\} = &\ \text{positive integers} \\
-1 \notin \mathbb{N}_0 = \{0,\ 1,\ 2,\ 3,\ ...\} = &\ \text{non-negative integers} \\
-1 \in \mathbb{Z} = \{0,\ 1,\ -1,\ 2,\ -2,\ 3,\ -3,\ ...\} = &\ \text{integers} \\
\sqrt{2} \notin \mathbb{Q} = \left\{ \frac{a}{b} \;\middle|\; a\,, b \in \mathbb{Z},\ b \neq 0 \right\} = &\ \text{rational numbers} \\
\sqrt{2} \in \mathbb{R} = &\ \text{real numbers} \\
i \in \mathbb{C} = &\ \text{complex numbers}
\end{aligned}
$$

If a set $M$ has only finitely many elements then it's called *finite*, and $|M|$ represents the number of elements in $M$, i.e. the cardinality (size) of $M$. For example, $|\{1, 2, 3\}| = 3$.

**Caution.** A set is not the same as a list or sequence. For instance, $\{1,2\} = \{2,1\} = \{1,2,1\}$, but $(1,2) \neq (2,1)$. The later are examples of ordered pairs.

**Example.** The **empty set** $\emptyset$ is the only set which has **no** element, i.e. $|\emptyset| = 0$.

**Notation** (operations on sets)**.** Let $M$ and $N$ be sets.
$M \subset N$: $M$ is a **subset** of $N$, i.e. every $m \in M$ also belongs to $N$. $\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$
$M \cap N$: **intersection** of $M$ and $N$, consists of all $m \in M$ which also belong to $N$.
$M \backslash N$: **complement** of $N$ in $M$, consists of all $m \in M$ which do not belong to $N$.


(a) $M \cap N$ (b) $M \setminus N$

Figure 1.1: Venn diagrams illustrating set intersection and set difference.

**Definition.** The **cartesian product** of two sets $M$ and $N$ is the set
$M \times N = \{(a,b) \mid a \in M,\, b \in N\}$ consisting of all ordered pairs with first coordinate in $M$ and second coordinate in $N$. If $a \neq b$, then $(a,b) \neq (b,a)$, even if $N = M$.

**Example.** The cartesian product $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \{(a,b) \mid a,b \in \mathbb{R}\}$ is the set of points (vectors) in the plane.

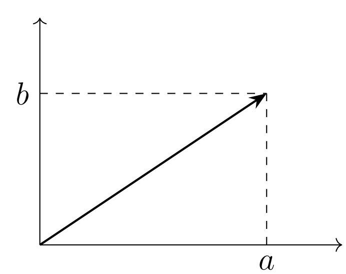



Figure 1.2: The vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ in $\mathbb{R}^2$.

**Example.** If $M = \{1,2\}$ and $N = \{1,2,3\}$, then $|M \times N| = |M| \cdot |N| = 6$. In particular, $(1,1) \in M \times N$.

**Definition.** The **power set** $\mathcal{P}(M)$ of a set $M$ consists of all subsets of $M$, including $\emptyset$ ($\emptyset \subset M$) and $M$ itself ($M \subset M$).

**Example.** If $M = \{1, 2\}$, then $\mathcal{P}(M) = \{\emptyset, \{1, 2\}, \{1\}, \{2\}\}$ and $|\mathcal{P}(M)| = 4$.

**Fact.** $|M| < \infty \implies |\mathcal{P}(M)| = 2^{|M|}$.

## Exercises

**Proof by induction:** Suppose $P(n)$ is a statement which makes sense for every integer $n$ which is greater than or equal to a certain given number $n_0$ (typically $n_0$ is zero or one). Our goal it to prove that $P(n)$ is a true statement for every integer $n$ greater than or equal to $n_0$. The technique "proof by induction" can often be used for this purpose. This kind of proof proceeds in two steps.

Step 1 "Induction anchor": We need to prove that $P(n_0)$ is true. This is usually easy and is often just a verification. (Without having done this, there is no way to proceed).

Step 2 "Induction step": We assume that $P(n)$ is true for some $n$ greater than or equal to $n_0$. This is called the "inductive hypothesis." Then we deduce that $P(n+1)$ is true (**without specifying what number** $n$ **is**). This step usually requires some idea and is not 'automatic.'

If step 2 has been completed, the principle of induction says that $P(n)$ is true for every $n$ greater than or equal to $n_0$.

**Example.** Let $n_0 = 1$ and $P(n)$ be the statement: "The sum $S(n) = 1 + 2 + ... + n$ is equal to $n(n+1)/2$.
Step 1: When $n = n_0 = 1$, the sum $S(1)$ has just one term, namely 1, and $S(1) = 1 = 1 \cdot (1+1)/2$. Hence the statement $P(1)$ is correct.
Step 2: We assume that $S(n) = n(n+1)/2$ (the inductive hypothesis). We observe that $S(n+1) = S(n) + n + 1$. Hence, by the inductive hypothesis, $S(n+1) = n(n+1)/2 + n + 1 = (n+1) \cdot (n/2 + 1) = (n+1) \cdot (n+2)/2$. We have thus shown that $P(n+1)$ is true if $P(n)$ is true, for an arbitrary $n$ greater than or equal to 1. Therefore, $P(n)$ is true for all positive integers.

1. Prove by induction that the sum $1 + 4 + 9 + ... + n^2$ of the first $n$ squares of positive integers is equal to $n(n+1)(2n+1)/6$, for every positive integer $n$.

2. Prove by induction that $1 + 3 + 5 + ... + (2n+1) = n^2$ for every positive integer $n$.

3. Given a set $X$ with $m$ distinct elements $x_1, x_2, ..., x_m$ and a set $Y$ with $n$ distinct elements $y_1, y_2, ..., y_n$, find a formula for the cardinality of the product set $X \times Y$ of all pairs $(x, y)$ with $x$ in $X$ and $y$ in $Y$, and prove it by induction on the cardinality $n$ of $Y$.

4. Let $X$ be a set of cardinality $n$. Show that the power set $\mathcal{P}(X)$ has $2^n$ elements.

## 1.2 Maps

**Definition.** Let $X$ and $Y$ be sets. A **map** (function) from $X$ to $Y$ (written $f : X \to Y$, or $X \xrightarrow{f} Y$) is a rule which assigns to every $x \in X$ a unique element $f(x) \in Y$, written $x \mapsto f(x)$. The set $X$ is called the **domain** of $f$, and $Y$ is called the **target** of $f$.

**Examples.** Some mappings are presented below.

1. $X = \mathbb{R} = Y, \quad f(x) = x^2$.

2. $X = \mathbb{R}$, $Y = \mathbb{Z}$, $\quad f(x) = \lfloor x \rfloor$ = floor function of $x$ = largest integer $\leq x$.

3. $X = \mathbb{R}^2$, $Y = \mathbb{R}$, $\quad f(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]) = 3x - 2y$.

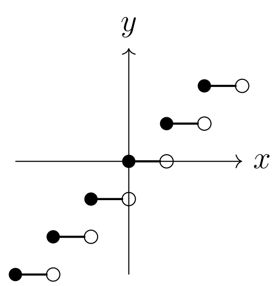



Figure 1.3: Graph of the floor function $y = \lfloor x \rfloor$.

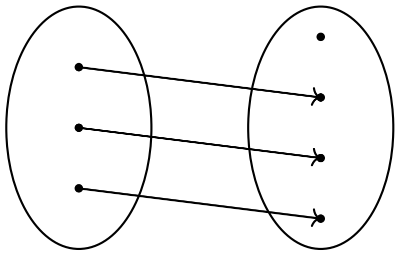
**Definition.** Let $f : X \to Y$ be a map. The map $f$ is called

1. **injective** (or one-to-one) if for every $x_1, x_2 \in X$ the following is true:
   $f(x_1) = f(x_2) \implies x_1 = x_2$;

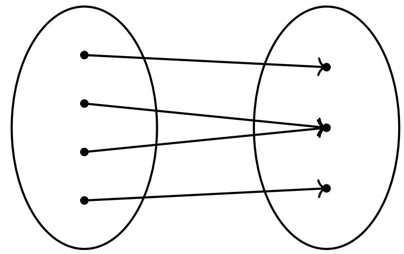
2. **surjective** (or onto) if for every $y \in Y$, there is some $x \in X$ such that $f(x) = y$;

3. **bijective** if it is injective and surjective.

**Fact.** If $f : X \to Y$ and $|X|, |Y| < \infty$, then $f$ injective $\implies |X| \leq |Y|$.



(a) Injective but not surjective



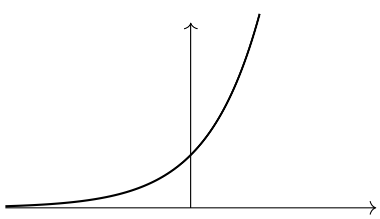
(b) Surjective but not injective

**Examples.**

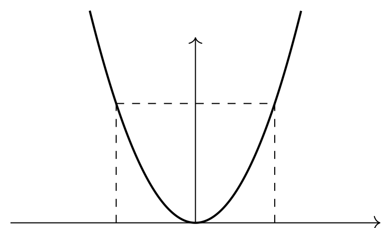
1. $f : \mathbb{R} \to \mathbb{R}, \quad f(x) = e^x$ is injective but not surjective.

2. $f : \mathbb{R} \to \mathbb{R}_{>0} := \{x \in \mathbb{R} \mid x > 0\}, \quad f(x) = e^x$ is bijective. In this case, the inverse of $f$ is $f^{-1} : \mathbb{R}_{>0} \to \mathbb{R}, \quad f^{-1}(x) = \ln(x)$.

3. $f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2$ is neither injective nor surjective. However, $f : [0, \infty) \to \mathbb{R}$, $f(x) = x^2$ is injective.



(a) $y = e^x$ maps onto $\mathbb{R}_{>0}$.



(b) $y = x^2$ maps $x$ and $-x$ to the same value.

**Some notation from logic.**

1. $\forall =$ "for all"

2. $\exists =$ "there exist(s)"

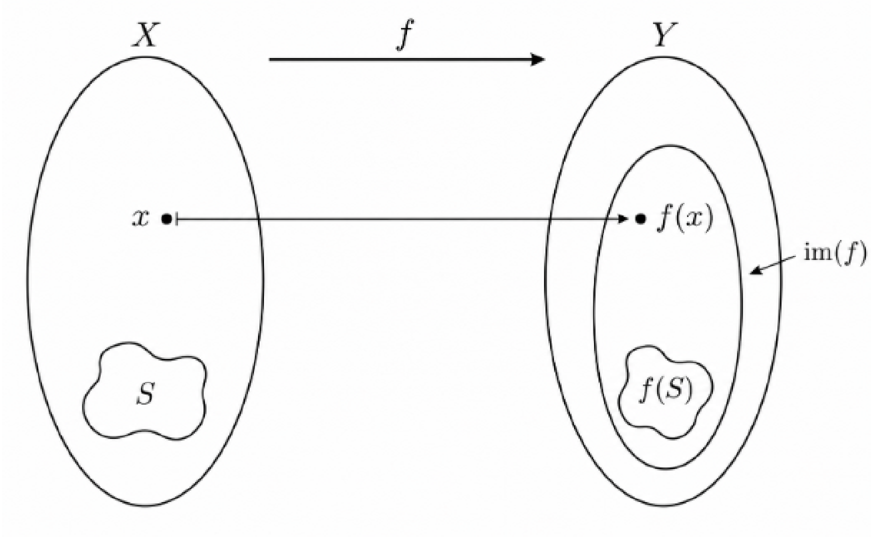
3. $\implies =$ "implication"

4. $A \iff B =$ "$A$ and $B$ are equivalent"

The first two of these symbols are called quantifiers.

**Example.** $f : X \to Y$ is surjective if $\forall y \in Y \ \exists x \in X : f(x) = y$.

**Definition.** Given $X \xrightarrow{f} Y$, we call $f(x)$ the **image of** $x \in X$. If $S \subset X$ is a subset of $X$, then $f(S) = \{f(x) \mid x \in S\} \subset Y$ is called the **image of** $S$. We define the **image of** $f$ to be $\mathrm{im}(f) := f(X)$.




**Note:** $f$ is surjective $\iff$ $\mathrm{im}(f) = Y$.

**Definition.** Given $T \subset Y$, we call $f^{-1}(T) = \{x \in X \mid f(x) \in T\}$ the **preimage of** $T$ under $f$. An element $x \in X$ with $f(x) = y$ is called **a preimage of** $y$.

**Note:**

1. $f : X \to Y$ is **surjective** if $\forall y \in Y : |f^{-1}(\{y\})| \geq 1$;

2. $f : X \to Y$ is **injective** if $\forall y \in Y : |f^{-1}(\{y\})| \leq 1$;

3. $f : X \to Y$ is **bijective** if $\forall y \in Y : |f^{-1}(\{y\})| = 1$.

**Caution.** Do not confuse $\{y\} \subset Y$ with $y \in Y$. There is a difference between $f^{-1}(\{y\})$ and $f^{-1}(y)$ (the later is only defined when $f$ is bijective).

**Definition.** If $f : X \to Y$ is bijective, then the **inverse map** $f^{-1} : Y \to X$ is defined by $f^{-1}(y) = x$ if $f(x) = y$.

**Note:**

1. $\forall x \in X,\ f^{-1}(f(x)) = x$;

2. $\forall y \in Y,\ f(f^{-1}(y)) = y$.

Two maps $f : X \to Y$ and $g : X' \to Y'$ are the same if $X' = X$ and $Y' = Y$ and $\forall x \in X,\ f(x) = g(x)$.

## Exercises

1. Let $X$ and $Y$ be finite sets, and let $f : X \to Y$ be a map.

   (a) If $f$ is injective, show that $|X|$ is less than or equal to $|Y|$.

   (b) If $f$ is surjective, show that $|X|$ is greater than or equal to $|Y|$.

   (c) If $f$ is bijective, show that $|X| = |Y|$.

2. Which of the following maps $f$ are injective, surjective, or bijective? Justify your response.

   (a) $f : \mathbb{R} \to \mathbb{R},\ \ f(x) = \sin(x)$.

   (b) $f : \mathbb{R}^2 \to \mathbb{R},\ \ f(x, y) = 3x + 2y + 1$.

   (c) $f : [-1, 1] \to [-1, 1],\quad f(x) = x^5$.

   (d) $f : \mathbb{R} \to \mathbb{R},\quad f(x) = x^3 + x^2$.

3. A $2 \times 2$ matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

   with real entries $a, b, c, d$, describes a linear map $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ which sends the vector $(x, y)$ to $(ax + by, cx + dy)$. For which of the following matrices $A$ is the corresponding linear map $T_A$ injective, surjective, or bijective? If it is not injective, find two different vectors which are mapped to the same vector. If it is not surjective, find a vector which is not in the image of $T_A$.

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

   (b) $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

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

## 1.3 Fields

**Definition.** A **field** $F$ is a set equipped with two maps

$$\begin{aligned} \text{“addition”} \ + : &\quad F \times F \to F, \ (a,b) \mapsto a+b \\ \text{“multiplication”} \ \cdot : &\quad F \times F \to F, \ (a,b) \mapsto a \cdot b \ \ (= ab) \end{aligned}$$

satisfying the following conditions (called the field axioms) for all $a$, $b$, $c \in F$:

**F1.** $a+b = b+a$ and $a \cdot b = b \cdot a$ (**commutativity** of $+$ and $\cdot$ )

**F2.** $(a+b)+c = a+(b+c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (**associativity** of $+$ and $\cdot$ )
We may write $a+b+c$ and $abc$ for the sum and product, respectively.

**F3.** $\exists\, 0 \in F$, $\exists\, 1 \in F:$ $a+0=a$ and $a \cdot 1 = a$. Moreover, we require that $0 \neq 1$.
The elements 0 and 1 are called the **additive and multiplicative identities**, respectively, and we sometimes write $0_F$ and $1_F$.

**F4.** $\forall\, a \in F$, $\exists\, b \in F:$ $a+b=0$.
The element $b=-a$ is called the **additive inverse** of $a$.
$\forall\, a \in F \setminus \{0\}$, $\exists\, c \in F:$ $a \cdot c = 1$.
The element $c = a^{-1} = \frac{1}{a}$ is called the **multiplicative inverse** of $a$.

**F5.** $a \cdot (b+c) = ab + ac$ (**distributive law**)

**Examples.**

1. $F = \mathbb{R}$ with the usual addition and multiplication is a field.

2. $F = \mathbb{Q}$ with the usual addition and multiplication is a field.

3. $F = (\mathbb{Z}, +, \cdot)$ is <u>not</u> a field.

4. $\mathbb{F}_2 = \{0, 1\}$ is a field under the following addition and multiplication:

| $+$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

| $\cdot$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

It is easy to verify that $\mathbb{F}_2$ is a field. Notice that it cannot be the case that $1+1=1$, for then $1=0$ which is not permitted in any field.

**Cancellation Laws.** If $a$, $b$, $c$ are elements of a field $F$, then $a+b=a+c \implies b=c$. If $a \neq 0$, then $a \cdot b = a \cdot c \implies b = c$.

## Exercises

1. **A field with three elements**. We consider here the numbers $0, 1$, and $2$ just as symbols. We set $\mathbb{F}_3 = \{0, 1, 2\}$ and define an addition and multiplication by the tables

| $+$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0 | 1 | 2 |
| 1 | 1 | 2 | 0 |
| 2 | 2 | 0 | 1 |

| $\cdot$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 |
| 2 | 0 | 2 | 1 |

With these two operators, $\mathbb{F}_3$ becomes a field. The zero element is 0 and the unit element (= identity for multiplication) is 1.

(a) What is the additive inverse of 1?

(b) What is the multiplicative inverse (= reciprocal) of 2?

(c) Show that the distributive law holds in $\mathbb{F}_3$, that is, for any $a, b, c$ in $\mathbb{F}_3$ one has

$$(a+b) \cdot c = a \cdot c + b \cdot c. \tag{D}$$

This formula is immediately seen to be true when $a$ or $b$ are zero, or if $c$ is zero or one. Hence, verify formula (D) only when $a$ and $b$ are both 1 or 2 and when $c = 2$ (this amounts to just checking four cases).

Note that $1 + 1 + 1 = 0$ in $\mathbb{F}_3$.

*Remark.* For every prime number $p$, there is a field with exactly $p$ elements.

# 1.4 Complex Numbers

**Definition.** Set $i = \sqrt{-1}$ and consider it for the moment just as a symbol. A **complex number** is a linear combination of the form $x + iy$ with $x, y \in \mathbb{R}$.

$$\text{addition:} \quad (x+iy) + (x'+iy') = (x+x') + i(y+y')$$
$$\text{multiplication:} \quad (x+iy) \cdot (x'+iy') = (xx'-yy') + i(xy'+yx')$$

We call $x = \mathrm{Re}(x+iy)$ the real part and $y = \mathrm{Im}(x+iy)$ the imaginary part.

**Note.** $i \cdot i = (0 + i \cdot 1)(0 + i \cdot 1) = (0 \cdot 0 - 1 \cdot 1) + i(0 \cdot 1 + 1 \cdot 0) = -1.$

We consider $x + i \cdot 0$ to be equal to $x$. The set of all numbers $x + iy$ we denote by $\mathbb{C}$, called the **the field of complex numbers**. Commutativity and associativity follow from the definitions of complex addition and multiplication.

$$\begin{aligned}
\text{identity for addition:}\quad & 0 = 0 + i \cdot 0 \\
\text{identity for multiplication:}\quad & 1 = 1 + i \cdot 0 \\
\text{inverse for addition:}\quad & -(x+iy) = -x + i(-y) \\
\text{inverse for multiplication:}\quad & (x+iy)^{-1} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2} \\
& \qquad\qquad\;\; = \frac{x}{x^2+y^2} + i\frac{-y}{x^2+y^2}
\end{aligned}$$

If $z = x + iy$, then $|z| = \sqrt{x^2+y^2}$ is called the (complex) absolute value, or **modulus**, of $z$, and $\bar{z} = x - iy = x + i(-y)$ is called the (complex) **conjugate** of $z$. Note that $z \cdot \bar{z} = |z|^2$ for all complex $z$.

**Fact.** (easy) For any two complex $z$ and $w$, we have

1. $|z \cdot w| = |z| \cdot |w|$,
2. $|z + w| \leq |z| + |w|$ (**triangle inequality**).

Every real number $x$ is considered a complex number: $x = x + i \cdot 0$. In this case, the modulus of $x$ (as a complex number) is the same as the absolute value of $x$ (as a real number)—that is, $|x|_{\mathbb{C}} = |x| =$ usual absolute value.

## Geometric Interpretation

We define a map $\phi : \mathbb{C} \to \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ by

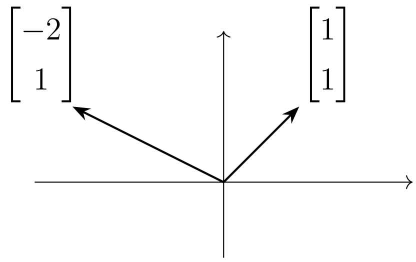
$$\phi(x+iy) = \begin{bmatrix} x \\ y \end{bmatrix}.$$

This map associates complex numbers with their vector representation in the plane.

**Examples.**

1. $\phi(1+i) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$,

2. $\phi(-2+i) = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.



Some easy to facts to verify:

1. $\forall z, w \in \mathbb{C} : \phi(z+w) = \phi(z) + \phi(w)$,

2. $\forall a \in \mathbb{R} : a(x+iy) = (a + i \cdot 0)(x+iy) = ax + i(ay)$ so that $\phi(a \cdot z) = a \cdot \phi(z)$.

Though complex numbers are not themselves vectors, the mapping $\phi$ represents them as such. Thus, we can see how sums and products of complex numbers arise geometrically by examining how $\phi(z+w)$ and $\phi(zw)$ relate to $\phi(z)$ and $\phi(w)$. The mapping $\phi$ relates addition of complex numbers to addiction of vectors, and multiplication of complex numbers to multiplication of lengths and addition of angles of vectors. This first relation is illustrated by fact 1 above. The second relation is not as obvious. Given a non-zero complex number $z = x+iy$, we may factor out the modulus of $z$ from each component and write

$$z = x + iy = |z| \left( \frac{x}{|z|} + i\frac{y}{|z|} \right)$$

Denote the real and imaginary components of $z/|z|$ by $a$ and $b$, respectively. Then, because $|z| = \sqrt{x^2+y^2}$, we have

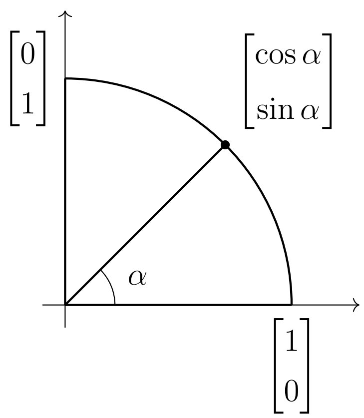
$$a^2 + b^2 = \frac{x^2}{|z|^2} + \frac{y^2}{|z|^2} = \frac{x^2+y^2}{|z|^2} = 1$$

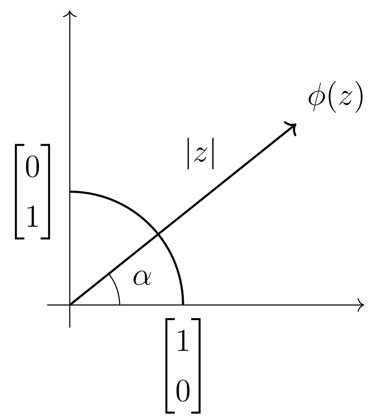
so that the complex number $z/|z| = a + ib$ has unit length. Because $a^2 + b^2 = 1$, there is an angle $\alpha \in [0, 2\pi)$ such that $a = \cos\alpha$ and $b = \sin\alpha$. Applying $\phi$ to $z$, we find

$$\phi(z) = \phi(|z|\,(a+ib)) = |z|\,\phi(a+ib) = |z| \begin{bmatrix} a \\ b \end{bmatrix} = |z| \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}.$$

We are interested in the geometry of the vector $\phi(zw)$ in terms of $\phi(z)$ and $\phi(w)$. By writing $z = |z| \cdot (\cos\alpha + i\sin\alpha)$ and $w = |w| \cdot (\cos\beta + i\sin\beta)$, the product can be written as

$$\begin{aligned} zw &= |z||w| \cdot [(\cos\alpha\cos\beta - \sin\alpha\sin\beta) + i(\cos\alpha\sin\beta + \sin\alpha\cos\beta)] \\ &= |z||w| \cdot [\cos(\alpha+\beta) + i\sin(\alpha+\beta)] \end{aligned}$$




using angle-addition formulas. Finally,

$$\phi(zw) = |z| \cdot |w| \begin{bmatrix} \cos(\alpha + \beta) \\ \sin(\alpha + \beta) \end{bmatrix}$$

so that length of $zw$ is the product of the lengths of $z$ and $w$, and the angle of $zw$ is the sum of the angles of $z$ and $w$.

# Exercises

1. Compute the following complex numbers in the form $x + iy$ with real numbers $x$ and $y$. [Remark: when writing complex numbers, we also use $x + yi$ instead of $x + iy$. This is consistent, as $iy = yi$]

   (a) $(2 + 3i) \cdot (1 - 2i) + 7i - 3$

   (b) $\frac{1-5i}{3+2i} - \frac{3}{2i+1}$

   (c) $\frac{1-i}{1+i} + \frac{1+i}{1-i}$

   (d) $3i \cdot (2 - i) + \frac{5i}{2-i}$

2. **Powers of complex numbers**. The power $z^n$ of a complex number $z$, when $n$ is a positive integer, if the product $z \cdot z \cdot ... \cdot z$ ($n$ copies of $z$). i.e., we multiply $z$ with itself $n$ times. If $n = 0$, then we use the convention that $z^0 = 1$ (even if $z = 0$). If $z$ is non-zero and $n$ is negative, then $z^n$ is defined to be $(\frac{1}{z})^{-n}$. Note that then $-n$ is positive. The following rule is true (and easy to show): $(z^n)^m = z^{n \cdot m}$. For example, $z^8 = ((z^2)^2)^2$. Using these rules, we can simplify computations.
   Compute the following powers of complex numbers:

   (a) $(1 + i)^8$

(b) $(-\frac{1}{2} + i\frac{\sqrt{3}}{2})^3$

(c) $i^{12}$

(d) $(\frac{1}{2} + i\frac{\sqrt{3}}{2})^6$

3. A complex number is called **purely imaginary** if its real part is zero.

   (a) Show that if $z$ is a non-zero complex number, then

   $$\frac{z}{\bar{z}} - \frac{\bar{z}}{z}$$

   is purely imaginary.

   (b) Show that if $z$ is a non-zero complex number, then $z/\bar{z}$ has absolute value 1.

4. **Polar coordinates**. As discussed earlier in the section, one can write every non-zero complex number $z = a + bi$ in the form

   $$r \cdot (\cos\phi + i\sin\phi)$$

   for a positive real number $r$ and a unique $\phi$ in the interval $[0, 2\pi)$. The pair $(r, \phi)$ is called the polar coordinates of $z$. The number $r$ is the complex absolute value of $z$. If we represent $z$ by the vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, then $\phi$ is the angle that this vector forms with the $x$-axis.

   Find the polar coordinates of the following numbers in the form $(r, 2\phi \cdot t)$ for some $t$ in $[0, 1)$:

   (a) $1 + i$

   (b) $1 + i\sqrt{3}$

   (c) $(1 + i)(1 + i\sqrt{3})$

   (d) $(1 + i)^2$

   Draw the Euclidean plane (perpendicular coordinate axes) and depict the vectors corresponding to those four complex numbers. Check that the vector corresponding to the third and fourth numbers—which are products—are related to their factors by the formula "lengths are multiplied and angles are added."

5. **Roots of unity**. A complex number $z$ is called a **root of unity** if there is a positive integer $n$ such that $z^n = 1$. The smallest such positive integer is called the **order** of $z$. For example, $-1$ is a root of unity of order 2.

(a) Draw the Euclidean plane and the roots of unity of order 1 (this is the number one), 2, and 4. Note that they form the vertices of a square.

(b) Draw the Euclidean plane and the roots of unity of order 1,2,3, and 6. Note that these points form the vertices of a regular hexagon.

6. The most important fact about the field of complex numbers $\mathbb{C}$ is that $\mathbb{C}$ is algebraically closed, which means that every non-constant polynomial

$$p(z) = a_n z^n + a_{n_1} z^{n_1} + ... + a_1 z + a_0$$

has a root in $\mathbb{C}$. As a consequence of **long division of polynomials**, every polynomial $p(z)$ as above can be written as

$$p(z) = a_n(z - c_1)(z - c_2)...(z - c_n) \tag{F}$$

where $c_1, ..., c_n$ are the roots of the polynomial (not necessarily distinct, i.e., it may happen that $c_i = c_j$ for some $i$ different from $j$).
Use long division (and the given hints) to find the factorization as in (F) of the following polynomials):

(a) $p(z) = z^4 - 1$
Hint: A root $c$ of this polynomial has the property that $c^4 = 1$.

(b) $p(z) = z^3 + 1$
Hint: A root $c$ of this polynomial has the property that $c^3 = -1$. This implies that $c^6 = 1$. You have determined those numbers before.

(c) $p(z) = z^3 - 2$
Hint: One root is the third root of 2. Let $b$ be this number, i.e. $b = \sqrt[3]{2}$. Use long division to write $p(z) = q(z)(z - b)$ with a polynomial $q(z)$ of degree 2. Find $q(z)$ and its roots.

## 1.5 Polynomials

A **polynomial** in the variable $x$ with coefficients in a field $F$ is an expression of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

where $a_n, ..., a_0 \in F$. If $a_n \neq 0$ then we call $n$ the **degree** of $f(x)$ and write $\deg(f) = n$. If $f(x) = a_0$, we call $f$ a **constant polynomial**. Given two polynomials $f(x) = a_n x^n + ... + a_0$ and $g(x) = b_n x^n + ... + b_0$, we define their sum to be

$$(f + g)(x) = (a_n + b_n)x^n + ... + (a_0 + b_0).$$

The zero polynomial is the polynomial $f = 0 = a_0$ where all coefficients are zero. Given a polynomial $f(x) = a_n x^n + ... + a_0$, we define the evaluation function $\text{ev}(f) : F \to F$ by

$$\text{ev}(f)(c) := a_n c^n + ... + a_1 c + +a_0.$$

If $|F| = \infty$, then $\text{ev}(f)$ determines the polynomial of $f$. But if $|F| < \infty$, then this is not always the case, as the next example illustrates.

**Example.** Consider the field $F = \mathbb{F}_2 = \{0, 1\}$ defined earlier. If we denote the zero polynomial by $f_1$, then $\text{ev}(f_1) : \mathbb{F}_2 \to \mathbb{F}_2$ is the zero map sending $0 \mapsto 0$ and $1 \mapsto 0$. But if $f_2 = x^2 + x = 1 \cdot x^2 + 1 \cdot x + 0$, then $\text{ev}(f_2) : \mathbb{F}_2 \to \mathbb{F}_2$ also sends $0 \mapsto 0$ and $1 \mapsto 0$.

# Chapter 2

# Vector Spaces

## 2.1 Vector Spaces

From now on, $F$ will denote a field (e.g. $F = \mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{F}_2$).

**Definition.** A **vector space** over $F$ (or $F$-vector space) is a set $V$ with two operations

$$\begin{aligned} \text{“addition”} + :\ & V \times V \to V,\ (v, w) \mapsto v + w \\ \text{“scalar multiplication”} \ \cdot\ :\ & F \times V \to V,\ (a, v) \mapsto av \end{aligned}$$

satisfying the following vector space axioms

**VS1.** commutativity of addition,

**VS2.** associativity of addition,

**VS3.** the existence of a zero vector $0 = 0_V = \underline{0} :\ \forall v \in V,\ 0_V + v = v$,

**VS4.** the existence of inverses with respect to addition: $\forall v \in V$, $\exists w \in V :\ v + w = 0_V$ (Notation: $w = -v$),

**VS5.** $\forall v \in V :\ 1 \cdot v = v$

**VS6.** $\forall a, b \in F$, $\forall v \in V :\ (ab) \cdot v = a \cdot (bv)$,

**VS7.** $\forall a \in F$, $\forall v, w \in V :\ a \cdot (v + w) = av + aw$,

**VS8.** $\forall a \in F$, $\forall v \in V :\ (a + b) \cdot v = a \cdot v + b \cdot v$.

The last two of these rules are called the **distributive laws**.

**Examples.**

1. The $n$-**dimensional standard vector space** $F^n$ consisting of all column vectors $[a_1 \ a_2 \ \dots \ a_n]^T$ with entries $a_1, ..., a_n \in F$ is a vector space. Here, $V = F^n = F \times F \times ... \times F$. Addition is **component-wise**:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

2. The set $M_{m\times n}(F)$ of all $m \times n$ rectangular arrays (matrices) with coefficients $a_{ij} \in F$ is a vector space.

$$\left.\overbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}}^{n \text{ columns}}\right\} m \text{ rows}$$

   Special names are given to certain shapes of matrices. A **row vector** is a $1 \times n$ matrix, and a **column vector** is an $m \times 1$ matrix. The set of all column vectors with $m$ entries is denoted by the usual $M_{m\times 1}(F) = F^m$. Addition and scalar multiplication of matrices is component-wise:

$$\begin{bmatrix} a_{ij} \end{bmatrix}_{1\le j\le n}^{1\le i\le m} + \begin{bmatrix} b_{ij} \end{bmatrix} = \begin{bmatrix} a_{ij} + b_{ij} \end{bmatrix}$$
$$c \cdot \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} c \cdot a_{ij} \end{bmatrix}$$

3. Let $X$ be any set. The set $\mathcal{F}(X, F)$ of all maps $f : X \to F$ is an $F$-vector space. Addition and multiplication are defined **pointwise**:

$$(f + g)(x) = f(x) + g(x)$$
$$(c \cdot f)(x) = c \cdot f(x)$$

   The zero vector in $\mathcal{F}(X, F)$ is the zero function $x \mapsto 0_F$. The additive inverse of $f$ is the function $-f$ defined by $(-f)(x) = -f(x)$.

4. The set $C(\mathbb{R}, \mathbb{R})$ of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a vector space.

5. The set $C([0, 1], \mathbb{R})$ of all continuous functions $f : [0, 1] \to \mathbb{R}$ is a vector space.

6. The set

$$F[x] = \{a_n x^n + ... + a_1 x + a_0 \mid n \in \mathbb{N}_0,\ a_0, ..., a_n \in F\}$$

   of all polynomials in $x$ is a vector space. The zero polynomial is the zero vector in the vector space $F[x]$.

7. The set $F[x]_{\leq n} = \{f(x) \in F[x] \mid \deg(f) \leq n\}$ of all polynomials with degree less than or equal to $n$ is a vector space.

Let $V$ be an $F$-vector space. Some general facts follow easily from the definition:

- **Cancellation law:** $\forall v, w, w' \in V\colon\ v + w = v + w' \implies w = w'$.
- $\forall v \in V,\ \forall a \in F$:
    1. $0_F \cdot v = 0_V$,
    2. $(-a) \cdot v = -(a \cdot v)$,
    3. $a \cdot 0_V = 0_V$.

## 2.2 Subspaces

$F$ always denotes a field.

**Definition.** A subset $W$ of a vector space $V$ is called a **subspace** if $W$ equipped with the addition and scalar multiplication of $V$ is a vector space.

**Theorem 1.** *Let $V$ be a vector space and $W \subset V$ a subset. Then $W$ is a subspace iff (= if and only if) the following conditions are fulfilled:*

*(i)* $W \neq \emptyset$,

*(ii)* $\forall v, w \in W:\ v + w \in W$,

*(iii)* $\forall v \in W, \forall a \in F:\ av \in W$.

*Proof.* "$\implies$": If $W$ is a subspace, then $0_V \in W$ so that $W \neq \emptyset$. Moreover, *(ii)* and *(iii)* are true by definition of "vector space."
"$\impliedby$": We need to show that $0_V \in W$ and $-v \in W$ for all $v \in W$. If $w \in W$ is any vector, then $0_F \cdot w = 0_V$ is in W by *(iii)*. If $v \in W$, then $-v = (-1) \cdot v$ is in $W$ by *(iii)*. $\square$

**Examples.**

1. Consider the vector space $V = \mathbb{R}^2$ over the field $F = \mathbb{R}$. The zero-subspace is $\{0\} = \{\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]\}$. All lines of the form $\mathbb{R} \cdot \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ for $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ are subspaces. The vector space $\mathbb{R}^2$ is itself a subspace.

2. If $A \in M_{m\times n}$ is a matrix, then $A^T$ is called the **transpose** of $A$ and has coefficients $a_{ji}$ in position $(i, j)$ [$i$th row, $j$th column]. For instance,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

   A square matrix $A = [a_{ij}]_{1\le i,j\le n}$ is called symmetric if $A = A^T$. Consider the vector space $V = M_n(F) = M_{n\times n}(F)$ of all square $n \times n$ matrices with coefficients in $F$. The set of all symmetric matrices $W = \{A \in V \mid A^T = A\}$ is a subspace of $M_n(F)$.

3. The set

$$D = \left\{ \begin{bmatrix} c_1 & & & \text{\huge 0} \\ & c_2 & & \\ & & \ddots & \\ \text{\huge 0} & & & c_n \end{bmatrix} \;\middle|\; c_1, ..., c_n \in F \right\}$$

   of $n \times n$ diagonal matrices is a subspace of $M_n(F)$.

4. The set $C(\mathbb{R}, \mathbb{R})$ of continuous functions $\mathbb{R} \to \mathbb{R}$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$ of **all** functions $\mathbb{R} \to \mathbb{R}$. The set $C^1(\mathbb{R}, \mathbb{R})$ of differentiable functions $\mathbb{R} \to \mathbb{R}$ is a subspace of $C(\mathbb{R}, \mathbb{R})$.

**Theorem 2.** *If $W_1, W_2$ are subspaces of $V$, then $W_1 \cap W_2$ is a subspace as well. The same is true for any collection of subspaces.*

*Proof.* Exercise 4. □

## Sums and Direct Sums [problem 23 in §1.3]

**Definition.** Let $W_1, W_2$ be subspaces of $V$.

1. $W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1,\ w_2 \in W_2\}$ is called the **sum** of $W_1$ and $W_2$.

2. $V$ is called a **direct sum** of $W_1$ and $W_2$ (written $V = W_1 \oplus W_2$) if

   (i) $\forall v \in V, \exists w_1 \in W_1, w_2 \in W_2 : \ v = w_1 + w_2$,

   (ii) $\forall w_1, w_1' \in W_1, \forall w_2, w_2' \in W_2 : \ w_1 + w_2 = w_1' + w_2' \implies w_1 = w_1'$ **and** $w_2 = w_2'$.

**Example.** Let $V = \mathbb{R}^3$ and consider the subspaces

$$W_1 = \left\{ \begin{bmatrix} 0 \\ b \\ c \end{bmatrix} \middle| \ b, c \in \mathbb{R} \right\}, \quad W_2 = \left\{ \begin{bmatrix} a \\ 0 \\ c \end{bmatrix} \middle| \ a, c \in \mathbb{R} \right\}.$$

$W_1$ is the $(y,z)$-plane and $W_2$ is the $(x,z)$-plane.

We can show that $W_1 + W_2 = \mathbb{R}^3 = V$ because any vector in $\mathbb{R}^3$ can be written as the sum of a vector in $W_1$ and a vector in $W_2$:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{\begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}}_{\in W_2} + \underbrace{\begin{bmatrix} 0 \\ b \\ c \end{bmatrix}}_{\in W_1}.$$

However, it is not the case that $W_1 \oplus W_2 = V$ because we can find a vector in $\mathbb{R}^3$ which can be represented as two distinct sums of vectors in $W_1$ and $W_2$:

$$\begin{aligned} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\in W_2} + \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{\in W_1} \\ &= \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\in W_2} + \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\in W_1}. \end{aligned}$$

**Proposition 2.1.** *Let $W_1, W_2 \subset V$ be subspaces.*

*(i) $W_1 + W_2$ is a subspace of $V$, and it contains $W_1$ and $W_2$.*

*(ii) If $W \subset V$ is a subspace, and $W_1 \subset W$ and $W_2 \subset W$, then $W_1 + W_2 \subset W$.*

*(iii) $V = W_1 \oplus W_2$ if and only if $V = W_1 + W_2$ and $W_1 \cap W_2 = \{0\}$.*

*Proof.* Exercise 6. □

**Definition.** If $W_1, ..., W_n$ are subspaces of $V$, we call $V$ the direct sum of $W_1, ..., W_n$, written $V = W_1 \oplus W_2 \oplus ... \oplus W_n$, if every $v \in V$ can be written in one and only one way as $v = w_1 + w_2 + ... + w_n$ with $w_1 \in W_1, w_2 \in W_2, ..., w_n \in W_n$.

Notice that the uniqueness of the sum is equivalent to

$$\begin{aligned}
&W_1 \cap (W_1 + ... + W_n) = \{0\} \\
&W_2 \cap (W_1 + W_3 + ... + W_n) = \{0\} \\
&\cdots \\
&W_n \cap (W_1 + ... W_{n-1}) = \{0\}.
\end{aligned}$$

## Exercises

1. Let $V$ and $W$ be vector spaces over a field $F$. Let

$$Z = \{(v, w) : v \in V \text{ and } w \in W\}.$$

Prove that $Z$ is a vector space over $F$ with the operations

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2) \text{ and } c(v_1, w_1) = (cv_1, cw_1).$$

*Remark.* The vector space $Z$ in this problem is also called the **direct sum** of $V$ and $W$. This is related to the concept of when a vector space is the direct sum of two subspaces, as discussed in this section. Namely, there is an obvious bijection from $V$ to the set of vectors $V' = \{(v, 0) \mid v \in V\}$ which is a subspace of $Z$, and there is an obvious bijection from $W$ to the set of vectors $W' = \{(0, w) : w \in W\}$ which is also a subspace of $Z$. Then $Z$ is the direct sum of the subspaces $V'$ and $W'$.

2. Determine whether the following sets are subspaces of $\mathbb{R}^3$ under the operations of addition and scalar multiplication defined on $\mathbb{R}^3$. Justify your answers.

   (a) $W_1 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ a_1 = 3a_2 \text{ and } a_3 = -a_2\}$

   (b) $W_2 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ a_1 = a_3 + 2\}$

   (c) $W_3 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ 2a_1 - 7a_2 + a_3 = 0\}$

   (d) $W_4 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ a_1 - 4a_2 - a_3 = 0\}$

   (e) $W_5 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ a_1 + 2a_2 - 3a_3 = 1\}$

   (f) $W_6 = \{(a_1, a_2, a_3) \in \mathbb{R}^3 : \ 5a_1^2 - 3a_2^2 + 6a_3^2 = 0\}$

3. Let $W_1, W_3$ and $W_4$ be as in Exercise 2. Describe $W_1 \cap W_3$, $W_1 \cap W_4$, and $W_3 \cap W_4$, and observe that each is a subspace of $\mathbb{R}^3$.

   *Remark.* When you are asked to "describe" the intersection of subspaces $W_1$ and $W_3$ (for example), this means to find a basis of this subspace. In order to do so, it is helpful to first find a basis for each of the spaces $W_1$, $W_3$, and $W_4$. The definition of "basis" will be recalled later, but in this very concrete situation you should be able to work it out from previous linear algebra classes.

4. Show that the intersection of two subspaces $U$ and $W$ of a vector space $V$ is again a subspace.

5. Prove that $W_1 = \{(a_1, a_2, ..., a_n) \in F^n : \ a_1 + a_2 + ... + a_n = 0\}$ is a subspace of $F^n$, but $W_2 = \{(a_1, a_2, ..., a_n) \in F^n : \ a_1 + a_2 + ... + a_n = 1\}$ is not.

6. Let $W_1$ and $W_2$ be subspaces of a vector space $V$.

   (a) Prove that $W_1 + W_2$ is a subspace of $V$ that contains both $W_1$ and $W_2$.

   (b) Prove that any subspace of $V$ that contains both $W_1$ and $W_2$ must also contain $W_1 + W_2$.

7. Show that $F^n$ is the direct sum of the subspaces

$$W_1 = \{(a_1, a_2, ..., a_n) \in F^n : \ a_n = 0\}$$

   and

$$W_2 = \{(a_1, a_2, ..., a_n) \in F^n : \ a_1 = a_2 = ... = a_{n-1} = 0\}.$$

8. Recall: Let $W_1, ..., W_n$ be subspaces of a vector space $V$. Then $V$ is called the direct sum of $W_1, W_2, ..., W_n$ if every vector $v$ in $W$ can be written in one and only one way as a sum $v = w_1 + w_2 + ... + w_n$ with vectors $w_i$ in $W_i$ (for each $i = 1, ..., n$).
   We define subspaces of $\mathbb{C}^3$ as follows: $W_1$ is the subspace of all vectors $[z\ z\ z]^T$ for $z$ in $\mathbb{C}$; $W_2$ is the subspace of all vectors of the form $[a\ b\ c]^T$ in $\mathbb{C}^3$ with $a + b + c = 0$; $W_3$ is the subspace of all vectors of the form $[z\ 2z\ -z]^T$ with $z$ in $\mathbb{C}$; $W_4$ is the subspace of all vectors of the form $[w\ 0\ 0]^T$ with $w$ in $\mathbb{C}$.

   (a) Is $\mathbb{C}^3$ the direct sum of $W_1$ and $W_2$?

   (b) Is $\mathbb{C}^3$ the direct sum of $W_2$ and $W_3$?

(c) Is $\mathbb{C}^3$ the direct sum of $W_1$ and $W_3$ and $W_4$?

(d) Is $\mathbb{C}^3$ the direct sum of $W_1$ and $W_2$ and $W_4$?

Justify your answers.

*Remark.* That we consider here subspaces of $\mathbb{C}^3$ is immaterial. Arguments which show that $\mathbb{C}^3$ is or is not a direct sum of the given subspaces do not really depend on the fact that we work with complex numbers here.

## 2.3 Linear Combinations and Systems of Linear Equations

**Definition.** Let $V$ be a vector space and $S \subset V$ be a subset of $V$.

1. The vector $v \in V$ is called a **linear combination** of vectors in $S$ if there exist $v_1, ..., v_n \in S$ and $a_1, ..., a_n \in F$ with
$$v = a_1 v_1 + a_2 v_2 + ... + a_n v_n.$$
The elements $a_1, ..., a_n$ are called the **coefficients** of the linear combination.

2. The **span** of $S$ is the set of all linear combinations of elements in $S$, denoted by $\operatorname{span}(S)$.

3. The span of the empty set $\emptyset$ is the zero vector space $\{0_V\}$. (This convention is inspired by the so-called "empty sum" $\sum_{i\in\emptyset} v_i = 0_V$).

**Key Questions:** Given $S = \{v_1, ..., v_n\}$ in $V$, we ask:

1. What is $\operatorname{span}(S)$? In other words, for which vectors $b \in V$ can we find $x_1, ..., x_n \in F$ such that
$$x_1 v_1 + ... + x_n v_n = b? \qquad (\mathbf{LS})$$

2. For which $[x_1 \; ... \; x_n]^T \in F^n$ do we have $x_1 v_1 + ... + x_n v_n = 0_V$?

If $V = F^m$, and if

$$v_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \; \ldots, \; v_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

then the linear system (**LS**) becomes

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

and this equation is equivalent to the $m$ equations

$$\begin{aligned} x_1 a_{11} + x_2 a_{12} + ... + x_n a_{1n} &= b_1 \\ \vdots \qquad \vdots \qquad\qquad & \vdots \\ x_1 a_{m1} + x_2 a_{m2} + ... + x_n a_{mn} &= b_m. \end{aligned}$$

If we set

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in M_{m \times n}(F),$$

then the linear system becomes

$$A \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

1. The first question becomes: for which $b \in F^m$ can we solve $Ax = b$?

2. The second question becomes: what are the solutions $x$ to $Ax = 0$?

**Definition.** We say that the subset $S \subset V$ **generates** $V$ if $\text{span}(S) = V$.

**Theorem 3.** $\text{span}(S)$ *is a subspace of* $V$*, and every subspace* $W \subset V$ *which contains* $S$ *also contains* $\text{span}(S)$.

## 2.4 Linear Dependence and Linear Independence

**Definition.** Let $I$ be any set (an "index set"). A system of vectors in $V$ indexed by $I$ is a map $I \to V$, $i \mapsto v_i$. If $I = \{1, ..., n\}$, then a system indexed by $I$ is a sequence $v_1, v_2, ..., v_n$.

**Definition.** A system of vectors $(v_i)_{i \in I}$ is called **linearly independent** if for all systems of scalars $(a_i)_{i \in I}$ in $F$ with the property that all but finitely many $a_i = 0$,

$$\sum_{i \in I} a_i v_i = 0_V \implies \forall i \in I : \; a_i = 0.$$

Concretely, if $I = \{1, ..., n\}$, then

$$a_1 v_1 + ... + a_n v_n = 0 \implies a_1 = a_2 = ... = a_n = 0.$$

Otherwise, the system $(v_i)_{i \in I}$ is called **linearly dependent**. The linear combination

$$0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n = 0_V$$

is called the **trivial linear combination**, or the **trivial representation** of the zero vector.

**Examples.**

1. Consider

$$v_1 = e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; v_2 = e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; v_3 = e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

   The system $(e_1, e_2, e_3)$ is linearly independent, but $(e_1, e_2, e_3, [1\ 1\ 1]^T)$ is linearly dependent:

$$e_1 + e_2 + e_3 + (-1) \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

2. Consider $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$. Because

$$2 \cdot v_1 + v_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

   the system $(v_1, v_2)$ is linearly dependent.

*Remark.*

1. Any system which contains the same vector twice ($v_i = v_j$ for some $i \neq j$) is linearly dependent: $1 \cdot v_i + (-1) \cdot v_j = 0_V$.

2. Any system which contains the zero vector is linearly dependent: $1 \cdot 0_V = 0_V$.

3. The "empty system" is linearly independent.

**Theorem 4.** *Suppose $S_1 = (v_1, ..., v_n)$ and $S_2 = (v_1, ..., v_n, v_{n+1}, ..., v_{n+m})$ are two systems. If $S_1$ is linearly dependent, then $S_2$ is linearly dependent. Conversely, if $S_2$ is linearly independent, then so is $S_1$.*

**Theorem 5.** *Suppose $S = (v_1, ..., v_n)$ is linearly independent and $w \in V$. Then $S' := (v_1, ..., v_n, w)$ is linearly dependent if and only if $w \in \operatorname{span}(S)$.*

*Proof.* " $\Longrightarrow$ ": There are $a_1, ..., a_n, c \in F$, not all zero, such that $a_1v_1 + ... + a_nv_n + cw = 0_V$. Since $S$ is linearly independent, we must have $c \neq 0$. But then

$$\begin{aligned} w &= -\frac{1}{c}(a_1v_1 + ... + a_nv_n) \\ &= \left(-\frac{a_1}{c}\right) v_1 + \cdots + \left(-\frac{a_n}{c}\right) v_n \in \operatorname{span}(S). \end{aligned}$$

" $\Longleftarrow$ ": We can write $w = c_1v_1 + ... + c_nv_n$ with $c_i \in F$. Then

$$c_1v_1 + ... + c_nv_n + (-1) \cdot w = 0.$$

Since $-1 \neq 0$, $S' = (v_1, ..., v_n, w)$ is linearly dependent. □

**Corollary 5.1.** *If $S$ generates $V$ and no proper subset $S'$ of $S$ generates $V$, then $S$ is linearly independent.*

*Proof.* Suppose $S = (v_1, ..., v_n)$ would be linearly dependent. Then there are scalars $c_1, ..., c_n$ **not** all zero with $c_1v_1 + ... + c_nv_n = 0$. Suppose $c_i \neq 0$ for some $i$. Then

$$v_i = \left(-\frac{c_1}{c_i}\right) v_1 + \left(-\frac{c_2}{c_i}\right) v_2 + \cdots + \left(-\frac{c_{i-1}}{c_i}\right) v_{i-1} + \left(-\frac{c_{i+1}}{c_i}\right) v_{i+1} + \cdots + \left(-\frac{c_n}{c_i}\right) v_n,$$

which shows that $v_i \in \operatorname{span}(v_i, ..., v_{i-1}, v_{i+1}, ..., v_n)$. Then, $\operatorname{span}(v_i, ..., v_{i-1}, v_{i+1}, ..., v_n) = V$, [for if $w \in S$ and $w \in \operatorname{span}(S \setminus \{w\})$, then $\operatorname{Span}(S \setminus \{w\}) = \operatorname{span}(S)$]. This contradicts our assumption. □

## 2.5 Bases and Dimension

**Definition.** A system of vectors $(v_i)_{i \in I}$ in $V$ is called a **basis** if it spans $V$ and is linearly independent.

**Examples.**

1. Recall that the empty set $\emptyset$ is linearly independent and $\operatorname{span}(\emptyset) = \{0\}$. Thus the empty set is a basis of the zero vector space.

2. Set

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad ..., \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ n \end{bmatrix}.$$

Then $(e_1, ..., e_n)$ is a basis of $F^n$ called the **standard basis** of $F^n$.

3. The system $(1, x, x^2, x^3, ...)$ is a basis of the $F$-vector space of polynomials $F[x]$ in $x$.

**Theorem 6.** *A system $\beta = (v_1, ..., v_n)$ in $V$ is a basis of $V$ if and only if there exist unique scalars $a_1, ..., a_n \in F$ such that $v = a_1v_1 + ... + a_nv_n$.*

*Proof.*

$$\begin{aligned} a_1v_1 + ... + a_nv_n = b_1v_1 + ... + b_nv_n \iff \\ (\underbrace{a_1 - b_1}_{=0})v_1 + ... + (\underbrace{a_n - b_n}_{=0})v_n = 0_V. \end{aligned}$$

□

**Theorem 7.** *If $V$ is generated by a finite system $\alpha = (v_1, ..., v_r)$, then there are indices $1 \leq i_1 < i_2 < ... < i_n \leq r$ such that $(v_{i_1}, v_{i_2}, ..., v_{i_n})$ is a basis.*

**Theorem 8** (Replacement Theorem)**.** *Suppose $\alpha = (v_1, ..., v_n)$ spans $V$ and $\gamma = (w_1, ..., w_m)$ is linearly independent. Then $m \leq n$ and there are indices $1 \leq i_1 < i_2 < ... < i_r \leq n$ such that the system $(w_1, ..., w_m, v_{i_1}, v_{i_2}, ..., v_{i_n})$ is a basis of $V$.*

**Corollary 8.1.** *If $V$ has a finite basis $\beta = (v_1, ..., v_n)$ and $\beta' = (w_1, ..., w_m)$ is also a basis of $V$, then $m = n$.*

*Proof.* Apply the Replacement Theorem to $\alpha := \beta$ and $\gamma := \beta'$ to deduce that $m \leq n$. Apply the Replacement Theorem again to $\alpha := \beta'$ and $\gamma := \beta$ to similarly find that $n \leq m$. Therefore $m = n$. □

**Definition.** A vector space $V$ is called **finite dimensional** if it has a finite basis. The cardinality of a basis is called the **dimension** of $V$.

**Examples.**

1. $\dim(\{0\}) = |\emptyset| = 0$.

2. $\dim(F^n) = |(e_1, ..., e_n)| = n$.

**Corollary 8.2.** *Let $V$ be any vector space of dimension $n$.*

1. *Any finite generating set for $V$ contains at least $n$ elements, and if it contains $n$ elements then it is a basis.*

2. *Any linearly independent finite system of vectors in $V$ has at most $n$ vectors, and if it has $n$ vectors then it is a basis.*

3. *Every linearly independent system can be extended to a basis.*

## Exercises

Let

$$v_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, v_2 = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, ..., v_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

be $n$ vectors in $F^m$, and let

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

be another vector in $F^m$. Let $A = [a_{ij}]$ be the $m \times n$ matrix ($m$ rows and $n$ columns) which has the vectors $v_1$,...,$v_n$ as columns (so that the $i$-th column is $v_i$). Problems in linear algebra often amount to solving an equation of the form $Ax = b$ where $x$ is a column vector in $F^n$. We denote by $\mathrm{Sol}(A \mid b)$ the set of all vectors $x$ in $F^n$ for which $Ax = b$.

For clarity, we sometimes denote the zero vector in $F^m$ (respectively $F^n$) by $0_m$ (respectively $0_n$).

For example, saying that the system of vectors $(v_1, ..., v_n)$ is linearly independent is equivalent to saying that $x = 0_n$ is the only solution to $Ax = 0_m$.

Let $[A' \mid b']$ be a row-echelon form (or the reduced row-echelon form, whatever you find more convenient) of $[A \mid b]$. Then, as shown in the next chapter, $\mathrm{Sol}(A \mid b) = \mathrm{Sol}(A' \mid b')$, and the later is usually much easier to determine.

A **pivot** in $[A' \mid b']$ is the leading non-zero entry in a row. If this entry is in position $(i, j)$ (i.e. $i$-th row and $j$-th column), then we call $(i, j)$ a **pivot position**. A **pivot column** of

$[A' \mid b']$ is a column which contains a pivot. By definition, the pivots, pivot positions, and pivot columns of the initial augmented matrix $[A \mid b]$ are those of the REF (= row-echelon form) $[A' \mid b']$.

For example, the matrix

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

has REF

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

and its pivot positions are $(1,1),(2,2),(3,4)$. Hence, the pivot positions of the initial matrix $A$ are also $(1,1),(2,2),(3,4)$, and the pivot columns of $A$ are the first, second, and fourth.

A general fact that you can use in the following exercises is the following: if $A$ has pivot columns $1 \le i_1 < ... < i_r \le n$ (the $i_1$-th column, the $i_2$-th column, and so forth are pivot columns), then those columns are a maximally linearly independent subsystem of the system of all columns of $A$.

This is useful when one wants to find a basis of a subspace $W$ of $F^m$ which is spanned by the $n$ vectors $v_1, ..., v_n$. For example, the subspace $W$ of $\mathbb{C}^3$ spanned by the vectors $v_1, v_2, v_3, v_4$ which are the rows of the matrix $A$ above has basis $(v_1, v_2, v_4)$.

Other facts that you can use without proof are the following:

(i) The vectors $v_1, ..., v_n$ span $F^m$ if and only if $A' = \mathrm{REF}(A)$ has no zero column.

(ii) The system $(v_1, ..., v_n)$ is a basis of $F^m$ if and only if $A' = \mathrm{REF}(A)$ has a pivot in every column and every row (which implies that $m = n$, of course).

In the exercises, use the following three row operations:

(i) $R_i \to R_j + cR_i$ means to add $c$ times the $i$-th row to the $j$-th row

(ii) $R_i \leftrightarrow R_j$ means to interchange the $i$-th row and the $j$-th column

(iii) $R_i \to cR_i$ means to multiply the $i$-th row with the (non-zero) scalar $c$.

1. In each part, determine whether the given vector is in the span of $S$.

(g) $\begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \right\}$

(h) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \right\}$

To tackle these problems, convert any $2 \times 2$ matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

into a column vector $[a\ b\ c\ d]$ in $\mathbb{R}^4$. Then perform the computations with these column vectors.

2. Determine which of the following sets are bases for $\mathbb{R}^3$.

   (a) $\{(1, 0, -1),\ (2, 5, 1),\ (0, -4, 3)\}$

   (c) $\{(1, 2, -1),\ (1, 0, 2),\ (2, 1, 1)\}$

3. Determine which of the following sets are bases for $\mathrm{P}_2(\mathbb{R}) = \mathbb{R}[x]_{\leq 2}$.

   (b) $\{1 + 2x + 2^2,\ 3 + x^2,\ x + x^2\}$

   (c) $\{1 - 2x - 2x^2,\ -2 + 3x - x^2,\ 1 - x + 6x^2\}$

   First, convert any of the polynomials $a + bx + cx^2$ to a column vector $[a\ b\ c]$ and then proceed. Saying that the so-obtained column vectors are a basis of $P_2(\mathbb{R})$ is equivalent to saying that the corresponding polynomials are a basis of $P_2(\mathbb{R})$.

4. Do the polynomials $x^3 - 2x^2 + 1$, $4x^2 - x + 3$, and $3x - 2$ generate $P_2(\mathbb{R})$? Justify your answer. (This can be determined without computation).

5. Is $\{(1, 4, -6),\ (1, 5, 8),\ (2, 1, 1),\ (0, 1, 0)\}$ a linearly independent subset of $\mathbb{R}^3$? Justify your answer. (This can be determined without computation).

6. **Lagrange interpolation**. Given three distinct (real) numbers $t_0 < t_1 < t_2$ and values $b_0, b_1, b_2$ (not necessarily distinct), there is a unique polynomial $f(t) = x_0 + x_1 t + x_2 t^2$ such that

$$f(t_0) = b_0, \quad f(t_1) = b_1, \quad f(t_2) = b_2.$$

The problem is to find the coefficients $x_0, x_1, x_2$. To do so, define the column vectors

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} t_0 \\ t_1 \\ t_2 \end{bmatrix}, v_3 = \begin{bmatrix} t_0^2 \\ t_1^2 \\ t_2^2 \end{bmatrix}$$

and $b = [b_0\ b_1\ b_2]^T$. Then the three equations hold if and only if $Ax = b$ where $A$ is the matrix with columns $v_1, v_2, v_3$ and $x = [x_0\ x_1\ x_2]^T$.

In each part, use the Lagrange interpolation formula to construct the polynomial of smallest degree whose graph contains the following points.

(a) $(-2, -6)$, $(-1, 5)$, $(1, 3)$

(b) $(-4, 24)$, $(1, 9)$, $(3, 3)$

# Chapter 3

# The Gauss-Jordan Elimination Algorithm

Facility with matrix algebra is assumed, however this chapter will briefly review the standard computations relating to systems of linear equations. Let

$$v_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \ v_2 = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, \ \ldots, \ v_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

be vectors in $F^m$. Set

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in M_{m\times n}(F), \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in F^m.$$

**Problem**: We want to find the solutions $x = [x_1 \ \dots \ x_n]^T \in F^n$ of the equation $x_1v_1 + \ldots + x_nv_n = b$ which is equivalent to $Ax = b$.

We denote by $\mathrm{Sol}(A \mid b) = \{x \in F^n \mid Ax = b\}$ the **solutions set** of $Ax = b$.

**Terminology**: We call

$$\left[\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_n \end{array}\right]$$

the **augmented matrix** associated to the linear system $Ax = b$.

**Key observation**: If $A' \mid b'$ is obtained from $A \mid b$ by an elementary row operation, then $\forall x \in F^n: \ Ax = b \iff A'x = b'$. In other words: $\mathrm{Sol}(A \mid b) = \mathrm{Sol}(A' \mid b')$.

*Proof.* $\square$

# Chapter 4

# Linear Transformations and Matrices

The eponymous chapter of these notes.

## 4.1 Linear Transformations, Null Spaces, and Ranges

Let $V$, $W$ be $F$-vector spaces.

**Definition.** A map $V \xrightarrow{T} W$ ($T : V \to W$) is called a **linear transformation** (or $F$-linear, or linear) if:

1. $\forall v_1, v_2 \in V :\ T(v_1 + v_2) = T(v_1) + T(v_2)$,

2. $\forall v \in V\ \forall c \in F :\ T(c \cdot v) = c \cdot T(v)$.

We denote by $\mathrm{Hom}_F(V, W) = \mathrm{Hom}(V, W)$ the set of all linear transformations $V \to W$.

**Example.** The zero map which is defined by $V \to W$, $v \mapsto 0_W$ is a linear map.

**Definition.** Let $T \in \mathrm{Hom}(V, W)$. Then $N(T) = \{v \in V \mid T(v) = 0\}$ is called the **null space** (or kernel) of $T$. The set $R(T) = \{T(v) \mid v \in V\} = \{w \in W \mid \exists v \in V : w = T(v)\}$ is called the **range** (or **image**) of $T$.

**Theorem 9.** *$N(T)$ and $R(T)$ are subspaces of $V$ and $W$, respectively.*

*Proof.* For $N(T)$: $T(0_V) = T(0_F \cdot 0_V) = 0_F \cdot T(0_V) = 0_W$ so that $0_V \in N(T)$. If $v_1, v_2 \in N(T)$, then $T(v_1 + v_2) = T(v_1) + T(v_2) = 0_W + 0_W = 0_W$ so that $v_1 + v_2 \in N(T)$. $\square$

**Theorem 10.** *Let $T \in \mathrm{Hom}(V, W)$ and suppose $(v_1, ..., v_n)$ generate $V$. Then* $\mathrm{R}(T) = \mathrm{span}(T(v_1), ..., T(v_n))$.

*Proof.* If $w \in \mathrm{R}(T)$, then $w = T(v)$ for some $v \in V$, so we can write $v = c_1v_1 + ... + c_nv_n$ by assumption. Then

$$\begin{aligned} T(v) &= T(c_1v_1) + T(c_2v_2 + ... + c_nv_n) \\ &= c_1T(v_1) + T(c_2v_2) + T(c_3v_3 + ... + c_nv_n) \\ &= c_1T(v_1) + c_2T(v_2) + ... + c_nT(v_n). \end{aligned}$$

□

**Definition.** Let $T \in \mathrm{Hom}(V, W)$. The number $\mathrm{nul}(T) = \mathrm{nullity}(T) = \dim_F(\mathrm{N}(T))$ is called the **nullity** of $T$. The number $\mathrm{rk}(T) = \mathrm{rank}(T) = \dim_F(\mathrm{R}(T))$ is called the **rank** of $T$.

**Theorem 11** (Rank-Nullity Theorem or Dimension Theorem)**.** *If $T : V \to W$ is a linear transformation and $n = \dim(V) < \infty$, then $\mathrm{rk}(T) + \mathrm{nul}(T) = n$.*

*Proof.* Let $m = \mathrm{nul}(T) = \dim \mathrm{N}(T)$, and let $(v_1, v_2, ..., v_m)$ be a basis of $\mathrm{N}(T)$. By theorem 8, we can extend this basis of $\mathrm{N}(T)$ to a basis $(v_1, v_2, ..., v_m, v_{m+1}, ..., v_n)$ of $V$. Then, by theorem 10, we have

$$\begin{aligned} \mathrm{R}(T) &= \mathrm{span}(T(v_1), T(v_2), ..., T(v_m), T(v_{m+1}), ..., T(v_n)) \\ &= \mathrm{span}(T(v_{m+1}), ..., T(v_n)). \end{aligned}$$

Set $n = m + r$, so that $T(v_{m+1}), ..., T(v_{m+r}) = T(v_n)$ are $r$ vectors.

**Claim.** The system $(T(v_{m+1}), ..., T(v_{m+r}))$ is linearly independent.

*Proof of claim.* Suppose $c_1T(v_{m+1}) + c_2T(v_{m+2}) + ... + c_rT(v_{m+r}) = 0_W$. Then $T(c_1v_{m+1} + c_2v_{m+2} + ... + c_rv_{m+r}) = 0_W$ so that $\sum_{i=1}^{r} c_iv_{m+i} \in \mathrm{N}(T)$. Then we may write $\sum_{i=1}^{n} c_iv_{m+i} = \sum_{j=1}^{m} a_jv_j$ where $v_{1\le j\le m}$ are the basis vectors of $\mathrm{N}(T)$. Rearranging, we get

$$0 = a_1v_1 + ... + a_mv_m + (-c_1)v_{m+1} + ... + (-c_r)v_{m+r}$$

which implies that $a_1 = a_2 = a_3 = ... = a_m = -c_1 = ... = -c_r = -0$ because $v_1 \le i \le n$ is a basis of $V$. Hence, $T(v_{m+1}), ..., T(v_{m+r})$ are linearly independent. □

Now, $(T(v_{m+1}, ..., T(v_{m+r}))$ is a basis of $\mathrm{R}(T)$, so $\mathrm{rk}(T) = \dim \mathrm{R}(T) = r$. Since $m = \mathrm{nul}(T)$, we find that

$$\mathrm{rk}(T) + \mathrm{nul}(T) = r + m = n = \dim(V).$$

□

**Theorem 12.** *Let $T : V \to W$ be linear and $\dim(V) = \dim(W) = n$. Then the following statements are equivalent:*

1. *$T$ is injective,*
2. *$T$ is surjective,*
3. *$T$ is bijective,*
4. *$\mathrm{N}(T) = \{0\}$,*
5. *$\mathrm{nul}(T) = 0$,*
6. *$\mathrm{R}(T) = W$,*
7. *$\mathrm{rk}(T) = n$.*

*Proof.*

$$\begin{aligned} T \text{ injective} &\iff \mathrm{N}(T) = \{0\} \iff \mathrm{nul}(T) = 0 \\ &\iff \dim \mathrm{R}(T) = n = \dim(W) \iff \mathrm{R}(T) = W \\ &\iff T \text{ surjective} \end{aligned}$$

Since $T$ injective $\implies$ $T$ surjective, then $T$ injective $\implies$ $T$ bijective. □

**Theorem 13.** *Let $V$, $W$ be vector spaces, $\beta = (v_1, ..., v_n)$ a basis of $V$, and $w_1, ..., w_n$ any sequence of vectors in $W$. Then there is one and only one $T \in \mathrm{Hom}(V, W)$ with $T(v_i) = w_i$ for $i = 1, 2, ..., n$.*

*Proof.* Given $v \in V$, write $v = a_1v_1 + a_2v_2 + ... + a_nv_n$ with unique $a_1, ..., a_n \in F$. Set

$$T(v) = a_1w_1 + a_2w_2 + ... + a_nw_n.$$

□

**Corollary 13.1.** *If $S, T \in \mathrm{Hom}(V, W)$ and $\beta = (v_1, ..., v_n)$ is a basis of $V$, then $S(v_i) = T(v_i)$ for all $i = 1, ..., n$ implies that $S = T$.*

## 4.2 The Matrix Representation of a Linear Transformation

**Definition.** Let $\beta = (v_1, ..., v_n)$ be a basis of $V$. Then given any vector $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ with $c_i \in F$, the column vector

$$[v]_\beta = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in F^n$$

is called the **coordinate vector** of $v$ with respect to the basis $\beta$. The scalars $c_1, ..., c_n$ are called the **coordinates of** $v$ **with respect to** $\beta$.

**Example.** The set

$$\beta = \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$$

is a basis of $\mathbb{R}^2$ (or $\mathbb{C}^2$).

**Lemma 13.1.** $[\cdot]_\beta : V \to F^n$, $v \mapsto [v]_\beta$ *is a linear map:*

1. $\forall v, v' \in V: \; [v + v']_\beta = [v]_\beta + [v']_\beta$,
2. $\forall c \in F, \forall v \in V: \; [cv]_\beta = c\,[v]_\beta$.

## 4.3 Composition of Linear Transformations and Matrix Multiplication

**Theorem 14.** *Let $V, W, Z$ be vector spaces, and $T : V \to W$, $S : W \to Z$ linear transformations. Then $S \circ T$ is a linear transformation from $V \to Z$.*

$$\begin{aligned} \operatorname{Hom}(W, Z) \times \operatorname{Hom}(V, W) &\to \operatorname{Hom}(V, Z) \\ (S, T) &\mapsto S \circ T \end{aligned}$$

*Proof.*

$$\begin{aligned} (S \circ T)(v + v') &= S(T(v + v')) = S(T(v) + T(v')) = S(T(v)) + S(T(v')) \\ &= (S \circ T)(v) + (S \circ T)(v') \end{aligned}$$

A similar computation shows that $(S \circ T)(cv) = c(S \circ T)(v)$. $\square$

**Theorem 15.** *Let $\beta = (v_1, ..., v_n)$ be a basis of $V$, $\gamma = (w_1, ..., w_n)$ be a basis of $W$, and $\delta = (z_1, ..., z_n)$ be a basis of $Z$. Then for all $T \in \mathrm{Hom}(V, W)$ and $S \in \mathrm{Hom}(W, Z)$:*

$$[S \circ T]_\beta^\delta = [S]_\gamma^\delta \cdot [T]_\beta^\gamma$$

## 4.4 Invertibility and Isomorphisms

## 4.5 The Change of Coordinate Matrix

**Key goals for the remaining chapters:**

1. Given $T : V \to V$, find a basis $\beta$ of $V$ such that $[T]_\beta$ is "very simple."
2. Given $A \in M_n(F)$, we want to find $Q \in GL_n(F)$ such that $Q^{-1}AQ$ is "very simple."

# Chapter 5

# Determinants

## 5.1 Determinants of Order 2

**Definition.** The **determinant** of a $2 \times 2$ matrix is

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

**Theorem 16.** *(i) The determinant is linear in the columns:* $\det[v_1 \;\; (v_2 + cv_3)] = \det[v_1 \;\; v_2] + c \cdot \det[v_1 \;\; v_3]$

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

*(iii)* $\det$ *is linear in the rows*

*(iv)* $\det[v_1 \;\; v_2] = -\det[v_2 \;\; v_1]$

**Theorem 17.** $\det A \neq 0$ *if and only if* $A$ *is invertible. In this case,*

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

## 5.2 Determinants of Order $n$

## 5.3 Properties of Determinants

# Chapter 6

# Diagonalization

## 6.1 Eigenvalues and Eigenvectors

If not said otherwise, $V$ is a finite-dimensional $F$-vector space.

**Definition.** 1. $T \in \mathrm{End}(V)$ is called **diagonalizable** if there exists a basis $\beta$ of $V$ such that $[T]_\beta$ $(= [T]_\beta^\beta)$ is a diagonal matrix.

2. $A \in M_n(F)$ is called **diagonalizable** if there exists a basis $\beta$ of $F^n$ such that $[L_A]_\beta$ is a diagonal matrix.

*Remark.* If $A$ is diagonalizable and $\beta = (v_1, ..., v_n)$ is a basis of $F^n$ such that $[L_A]_\beta$ is diagonal, then

$$[L_A]_\beta = \overbrace{[\mathrm{id}_{F^n}]_{\mathrm{st}}^\beta}^{Q^{-1}} \cdot \overbrace{[L_A]_{\mathrm{st}}^{\mathrm{st}}}^{A} \cdot \overbrace{[\mathrm{id}_{F^n}]_\beta^{\mathrm{st}}}^{Q} = Q^{-1}AQ$$

is diagonal.

## 6.2 Diagonalizability

## 6.3 Invariant Subspaces and the Cayley-Hamilton Theorem

**Theorem 18** (Cayley-Hamilton)**.** *Let $T \in \mathrm{End}(V)$. Then $\Delta_T(T) = 0$ (= zero map).*

*Proof.* Set $S = \Delta_T(T) \in \mathrm{End}(V)$. Our goal is to show that $S$ is the zero map. Hence $S(v) = 0$ for any $v \in V$. Take any (non-zero) vector $v$ and consider

$$W = \mathrm{span}(v, T(v), ...).$$

Set $d = \dim(W)$. Then $\beta = (v, T(v), ..., T^{d-1}(v))$ is a basis of $W$, by 5.11$(i)$. We've seen $T(W) \subset W$. Since $\Delta_{T|_W} \in \mathrm{End}(V)$, then by 5.21, $\Delta_{T|_W}(t)$ divides $\Delta_T(t)$, equivalently $\Delta_T(t) = g(t) \cdot \Delta_{T|_W}$ for some $g(t) \in F[t]$. Evaluate $t = T$: $S = \Delta_T(T) = g(T) \cdot \Delta_{T|_W}(T)$. Suppose $a_0 v + a_1 T(v) + ... + a_{d-1} T^{d-1}(v) + T^d(v) = 0$. Then 5.22$(ii)$ produces $\Delta_{T|_W}(t) = (-1)^d(a_0 + a_1 t + ... + a_{d-1} t^{d-1} + t^d)$. But then $\Delta_{T|_W}(T) = (-1)^d(a_0 + a_1 T + ... + a_{d-1} T^{d-1} + T^d)$ so that $S(v) = g(T)(\Delta_{T|_W}(T)(v)) \implies S$ is the zero map. End of proof. $\square$

# Chapter 7

# Jordan Canonical Form

The namesake of this chapter is French mathematician Camille Jordan (1838-1922).

## 7.1 The Jordan Canonical Form I

The problem we want to solve:
**Problem I**: Given $T \in \mathrm{End}(V)$, is there a basis $\beta$ of $V$ such that $[T]_\beta$ is particularly easy to understand? If $\beta'$ is another basis of $V$ and $Q = [\mathrm{id}_V]_{\beta'}^{\beta}$ is the change-of-basis matrix, then

$$\begin{aligned} B := [T]_{\beta'} &= [\mathrm{id}]_{\beta}^{\beta'}[T]_{\beta}^{\beta}[\mathrm{id}]_{\beta'}^{\beta} \\ &= Q^{-1}[T]_\beta Q. \end{aligned}$$

Hence $A = [T]_\beta$ and $B = [T]_\beta$ are similar. Call the set $\mathrm{cl}(A) = \{B \in M_n(F) \mid B \text{ is similar to } A\}$ the **similarity class** of $A$. Solving Problem I is equivalent to
**Problem II**: Given $A \in M_n(F)$, what is the "simplest" matrix in the similarity class of $A$?

**Examples.**

1. If $T$ is diagonalizable, then any basis $\beta = (v_1, v_2, ..., v_n)$ consisting of eigenvectors gives

$$[T]_\beta = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

where $\lambda_i$ is the eigenvalue of $v_i$, i.e. $T(v_i) = \lambda_i v_i$. This would be considered a solution to Problem I.

2. Suppose $T$ is not diagonalizable but $V$ is $T$-cyclic, hence $V = \text{span}(v, T(v), ..., T^{n-1}(v))$ where $n = \dim(V)$. Then $\beta = (T^{n-1}(v), T^{n-2}(v), ..., T(v), v)$ is a basis of $V$. Also assume $T^n(v) = 0$. Then the $n \times n$ matrix

$$\mathrm{J}_n(0) := [T]_\beta = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & 0 & \ddots & \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & & 0 \end{bmatrix}$$

with ones on the first super-diagonal is a solution to Problem I.

**Definition.** The $m \times m$ matrix

$$\mathrm{J}_m(\lambda) := \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}$$

is called the **Jordan block** of size $m$ with eigenvalue $\lambda$.

## 7.2 The Jordan Canonical Form II

## 7.3 Complexification

Let $V$ be any $\mathbb{R}$-vector space.

**Definition.** The **complexification** of $V$ is a complex vector space $V_\mathbb{C}$ whose set is $V \oplus V = \{(v_1, v_2) \mid v_1, v_2 \in V\}$ and scalar multiplication by $\mathbb{C}$ on $V_\mathbb{C}$ is defined by

$$(x + iy) \cdot (v_1, v_2) = (xv_1 - yx_2, xv_2 + yv_1) \in V \oplus V = V_\mathbb{C}.$$

**Proposition 18.1.** *This gives $V_\mathbb{C}$ the structure of a vector space over $\mathbb{C}$. Moreover, there is a canonical (natural) $\mathbb{R}$-linear injective map $V \to V_\mathbb{C}$ given by $v \mapsto (v, 0)$. We will henceforth identify $V$ with the image of this map. Then*

$$i \cdot V = \{i(v, 0) \mid v \in V\} = \{(0, v) \mid v \in V\}$$

*and $V_\mathbb{C} = V \oplus iV$. [using that $(v_1, v_2) = (v_1, 0) + (0, v_2)$]*

**Proposition 18.2.** *Let $V$ be an $\mathbb{R}$-vector space and $T : V \to V$ an $\mathbb{R}$-linear map. Then $T_\mathbb{C} : V_\mathbb{C} \to V_\mathbb{C}$, $T_\mathbb{C}((v_1, v_2)) = (T(v_1), T(v_2))$ is a $\mathbb{C}$-linear transformation of $V_\mathbb{C}$ to itself, called the **complexification** or extension of $T$ to $V_\mathbb{C}$.*

# Chapter 8

# Inner Product Spaces

In this chapter, $F = \mathbb{R}$ or $F = \mathbb{C}$.

## 8.1 Inner Products and Norms

**Definition.** Let $V$ be an $F$-vector space. An **inner product** on $V$ is a map $V \times V \to F$, $(v_1, v_2) \mapsto \langle v_1, v_2 \rangle$ such that $\forall v, w, v_1, v_2 \in V$ and $c \in F$:

(a) $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$,

(b) $\langle cv, w \rangle = c\langle v, w \rangle$,

(c) $\overline{\langle w, v \rangle} = \langle v, w \rangle$,

(d) $\langle v, v \rangle = 0$ for all $v \neq 0$.

*Remark.* If $F = \mathbb{R}$, then (c) says $\langle w, v \rangle = \langle v, w \rangle$ and the inner product is called **symmetric**. If $F = \mathbb{C}$, then—in general— $\overline{\langle w, v \rangle} = \langle v, w \rangle \neq \langle w, v \rangle$, and $\langle\ ,\ \rangle$ is called **hermitian** (Charles Hermite, 1822-1901). If $\langle x, x \rangle > 0$ whenever $x \neq 0$, then $\langle\ ,\ \rangle$ is called **positive definite**.

**Examples.**

1. Let $F = \mathbb{R}$, $V = \mathbb{R}^n$, and define the standard inner product space

$$\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle_{\text{st}} = \sum_{i=1}^{n} x_i y_i$$

2. Let $F = \mathbb{C}$, $V = \mathbb{C}^n$, and define the standard inner product space

$$\left\langle \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \right\rangle_{\text{st}} = \sum_{i=1}^{n} z_i \overline{w_i}$$

3. Consider $V = C^0([a,b],\ \mathbb{R})$ or $V = C^0([a,b],\ \mathbb{C})$ and define the inner product

$$\langle f, g \rangle = \frac{1}{b-a} \int_a^b f(t)\overline{g(t)}dt.$$

This inner product is positive definite:

$$\langle f, f \rangle = \frac{1}{b-a} \int_a^b f(t)\overline{f(t)}dt = \frac{1}{b-a} \int_a^b |f(t)|^2 dt.$$

If this integral is equal to 0, then $|f(t)|^2$ must be the zero function (because it's continuous!). Thus $f(t) = 0$ for all $t \in [a,b]$.

**Theorem 19.** *Let $(V, \langle , \rangle)$ be an inner product space over $F$. Then $\forall v, w \in V$ and $\forall c \in F$:*

*(a)* $\|cv\| = |c| \cdot \|v\|$,

*(b)* $\|v\| = 0$ *if and only if* $v = 0_V$,

*(c) (Cauchy-Schwartz inequality)* $|\langle v, w \rangle| \leq \|v\|\|w\|$,

*(d) (triangle inequality)* $\|v + w\| \leq \|v\| + \|w\|$.

*Proof.* (a) easy

(b) easy

(c) $0 \leq \langle v - cw, v - cw \rangle = \langle v, v - cw \rangle + \langle -cw, v - cw \rangle = \langle v, v \rangle - \bar{c}\langle v, w \rangle - c\langle w, v \rangle + c\bar{c}\langle w, w \rangle$. Now set $c = \frac{\langle v,w \rangle}{\langle w,w \rangle}$ for $w \neq 0$. Then

$$\begin{aligned} 0 &\leq \langle v, v, \rangle - \frac{|\langle v, w \rangle|^2}{\langle w, w \rangle} + \frac{|\langle v, w \rangle|^2}{\langle w, w \rangle} + \frac{|\langle v, w \rangle|^2}{\langle w, w \rangle^2}\langle w, w \rangle \\ &= \|v\|^2 - \frac{|\langle v, w \rangle|^2}{\|w\|^2} \end{aligned}$$

so that

$$|\langle v, w \rangle|^2 \leq \|v\|^2 \|w\|^2.$$

Finish by taking square roots on both sides.

□

## 8.2 The Gram-Schmidt Orthogonalization Process and Orthogonal Complements

## 8.3 The Adjoint of a Linear Operator

## 8.4 Normal and Self-Adjoint Operators