Linear Algebra (Math 110), Summer 2012

George Melvin
University of California, Berkeley
(July 31, 2012 corrected version)

Abstract

These are notes for the upper division course 'Linear Algebra' (Math 110) taught at the University of California, Berkeley, during the summer session 2012. Students are assumed to have attended a first course in linear algebra (equivalent to UCB Math 54). The aim of this course is to provide an introduction to the study of finite dimensional vector spaces over fields of characteristic zero and linear morphisms between them and to provide an abstract understanding of several key concepts previously encountered. The main topics to be covered are: basics of vector spaces and linear morphisms, the Jordan canonical form and Euclidean/Hermitian spaces.

Contents

0 Preliminaries

In this preliminary section we will introduce some of the fundamental language and notation that will be adopted in this course. It is intended to be an informal introduction to the language of sets and functions and logical quantifiers.

0.1 Basic Set Theory

For most mathematicians the notion of a *set* is fundamental and essential to their understanding of mathematics. In a sense, everything in sight is a set (even functions can be considered as sets![1]) .

A vector space is an example of a *set with structure* so we need to ensure that we know what a set is and understand how to write down and describe sets using set notation.

Definition 0.1.1 (Informal Definition)**.** A *set* S is a collection of objects (or elements). We will denote the size of a set S by $|S|$; this will either be a natural number or infinite (we do discuss questions of cardinality of sets).

For example, we can consider the following sets:

- the set P of people in Etcheverry, room 3109, at 10.10am on 6/18/2012,
- the set B of all people in the city of Berkeley at 10.10am on 6/18/2012,
- the set $\mathbb{R}$ of all real numbers,
- the set A of all real numbers that are greater than or equal to π,
- the set $M_{m\times n}(\mathbb{R})$ of all $m \times n$ matrices with real entries,
- the set $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ of all $\mathbb{R}$-linear morphisms with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$,
- the set $C(0,1)$ of all real valued continuous functions with domain $(0,1)$.

Don't worry if some of these words are new to you, we will define them shortly.

You will observe that there are some relations between these sets: for example,

- every person that is an object in the collection P is also an object in the collection B,
- every number that is an object of A is also an object of $\mathbb{R}$.

We say in this case that P (resp. A) is a *subset* of B (resp. $\mathbb{R}$), and write

$$P \subseteq B \text{ (resp. } A \subseteq \mathbb{R}).$$

[1] A function $f : A \to B\ ;\ x \mapsto f(x)$ is the same data as providing a subset $\Gamma_f \subset A \times B$, where $\Gamma_f = \{(x, f(x)) \mid x \in A\}$, the *graph of* f. Conversely, if $C \subset A \times B$ is a subset such that, $\forall a \in A, \exists b \in B$ such that $(a,b) \in C$, and $(a,b) = (a,b') \in C \implies b = b'$, then C is the graph of some function.

Remark. In this class we will use the notations $\subseteq$ and $\subset$ interchangeably and make no distinction between them. On the blackboard I will write $\subseteq$ as this is a notational habit of mine whereas in these notes I shall usually write $\subset$ as it is a shorter command in LaTeX(the software I use to create these notes).

We can also write the following

$$P = \{x \in B \mid x \text{ is in Etcheverry, room 3109, at 10.10am on } 6/18/2012\},$$

or in words:

P is the set of those objects x in B such that x is in Etcheverry, room 3109, at 10.10am on 6/18/2012.

Here we have used

- the logical symbol '$\in$' which is to be translated as 'is a member of' or 'is an object in the collection',
- the vertical bar '$|$' which is to be translated as 'such that' or 'subject to the condition that'.

In general, we will write (sub)sets in the following way:

$$T = \{x \in S \,|\mathcal{P}\},$$

where $\mathcal{P}$ is some property or condition. In words, the above expression is translated as

T is the set of those objects x in the set S such that x satisfies the condition/property $\mathcal{P}$.

For example, we can write

$$A = \{x \in \mathbb{R} \mid x \geq \pi\}.$$

Definition 0.1.2. We will use the following symbols (or logical quantifiers) frequently:

- $\forall$ - translated as 'for all' or 'for every', (the *universal quantifier*)
- $\exists$ - translated as 'there exists' or 'there is, (the *existential quantifier*).

For example, the statement

'for every positive real number x, there exists some real number y such that $y^2 = x$',

can be written

$$\forall x \in \mathbb{R} \text{ with } x > 0, \exists y \in \mathbb{R} \text{ such that } y^2 = x.$$

Remark. Learning mathematics is difficult and can be made considerably more difficult if the basic language is not understood. If you ever encounter any notation that you do not understand please ask a fellow student or ask me and I will make sure to clear things up. I have spent many hours of my life staring blankly at a page due to misunderstood notation so I understand your pain in trying to get to grips with new notation and reading mathematics.

Notation. In this course we will adopt the following notational conventions:

- $\varnothing$, the *empty set* (ie the empty collection, or the collection of no objects),
- $[n] = \{1, 2, 3, \ldots, n\}$,
- $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$, the set of *natural numbers*,
- $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$, the set of *integers*,
- $\mathbb{Z}_{\geq a} = \{x \in \mathbb{Z} \mid x \geq a\}$, and similarly $\mathbb{Z}_{>a}, \mathbb{Z}_{\leq a}, \mathbb{Z}_{<a}$,
- $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$, the set of *rational numbers*,
- $\mathbb{R}$, the set of *real numbers*,

- $\mathbb{C}$, the set of *complex numbers*.

Remark (Complex Numbers)**.** Complex numbers are poorly taught in most places so that most students have a fear and loathing of them. However, there is no need to be afraid! It really doesn't matter whether you consider imaginary numbers to be 'real' (or to exist in our domain of knowledge in this universe), all that matters is that you know their basic properties: a complex number $z \in \mathbb{C}$ is a'number' that can be expressed in the form

$$z = a + b\Delta,\ a, b \in \mathbb{R},$$

where, for now, Δ is just some symbol.

We can add and multiply the complex numbers $z = a + b\Delta, w = c + d\Delta \in \mathbb{C}$, as follows

$$z + w = (a + c) + (b + d)\Delta, \quad z.w = (ac - bd) + (bc + ad)\Delta.$$

If $z = a + b\Delta \in \mathbb{C}$ then the complex number $\tilde{z} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}\Delta$ satisfies

$$z.\tilde{z} = \tilde{z}.z = 1,$$

so that $\tilde{z}$ is the multiplicative inverse of z and we can therefore write $1/z = z^{-1} = \tilde{z}$. Hence, if $z = a + b\Delta, w = c + d\Delta \in \mathbb{C}$, then

$$z/w = z.w^{-1} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}\Delta.$$

Of course, the number $i = 1.\Delta$ satisfies the property that $i^2 = -1$, so that $1.\Delta$ corresponds to the imaginary number i that you learned about in high school. However, as we will be using the letter i frequently for subscripts, we shall instead just write $\sqrt{-1}$ so that we will consider complex numbers to take the form

$$z = a + b\sqrt{-1}.$$

We have the following inclusions

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C},$$

so that, in particular, *every real number is also a complex number* (if $a \in \mathbb{R}$ then we consider $a = a.1 + 0.\sqrt{-1} \in \mathbb{C}$).

Definition 0.1.3 (Operations on Sets)**.** • Suppose that S is a set and S_1, S_2 are subsets.

- the *union of S_1 and S_2* is the set

$$S_1 \cup S_2 = \{x \in S \mid x \in S_1 \text{ or } x \in S_2\}.$$

- the *intersection of S_1 and S_2* is the set

$$S_1 \cap S_2 = \{x \in S \mid x \in S_1 \text{ and } x \in S_2\}.$$

More generally, if $S_i \subset S, i \in J$, is a family of subsets of S, where J is some indexing set, then we can define

$$\bigcup_{i \in J} S_i = \{s \in S \mid s \in S_k, \text{ for some } k \in J\},$$

and

$$\bigcap_{i \in J} S_i = \{s \in S \mid s \in S_k, \forall k \in J\}.$$

• Let A, B be sets.

- the *Cartesian product of A and B* is the set

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

so that the elements of $A \times B$ are *ordered pairs* (a, b), with $a \in A, b \in B$. In particular, it is not true that $A \times B = B \times A$.

Moreover, if $(a, b), (a', b') \in A \times B$ and $(a, b) = (a', b')$, then we must necessarily have $a = a'$ *and* $b = b'$.

For example, consider the following subsets of $\mathbb{R}$:

$$A = \{x \in \mathbb{R} \mid 0 < x < 2\},\ B = \{x \in \mathbb{R} \mid x > 1\},\ C = \{x \in \mathbb{R} \mid x < 0\}.$$

Then,

$$A \cup B = (0, \infty),\ A \cap B = (1, 2),\ A \cap C = \varnothing,\ A \cup B \cup C = \{x \in \mathbb{R} \mid x \neq 0\}.$$

Also, we have

$$A \times C = \{(x, y) \mid 0 < x < 2,\ y < 0\}.$$

0.2 Functions

Functions allow us to talk about certain relationships that exist between sets and allow us to formulate certain operations we may wish to apply to sets. You should already know what a function is but the notation to be introduced may not have been encountered before.

Definition 0.2.1. Let A, B be sets and suppose we have a function $f : A \to B$. We will write the information of the function f as follows:

$$f : A \to B,\ x \mapsto f(x),$$

where $x \mapsto f(x)$ is to be interpreted as providing the data of the function, ie, x is the input of the function and $f(x)$ is the output of the function. Moreover,

- A is called the *domain of* f,
- B is called the *codomain of* f.

For example, if we consider the function $|.| : \mathbb{R} \to [0, \infty)$, the 'absolute value' function, then we write

$$|.| : \mathbb{R} \to [0, \infty),\ x \mapsto \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

The codomain of $|.|$ is $[0, \infty)$ and the domain of $|.|$ is $\mathbb{R}$.

Definition 0.2.2. Let $f : A \to B$ be a function.

- we say that f is *injective* if the following condition is satisfied:

$$\forall x, y \in A, \text{ if } f(x) = f(y) \text{ then } x = y,$$

- we say that f is *surjective* if the following condition is satisfied:

$$\forall y \in B,\ \exists x \in A \text{ such that } f(x) = y,$$

- we say that f is *bijective* if f is both injective and surjective.

It should be noted that the injectivity of f can also be expressed as the following (logically equivalent) condition:

$$\text{if } x, y \in A,\ x \neq y, \text{ then } f(x) \neq f(y).$$

Also, the notion of bijectivity can be expressed in the following way:

$$\forall y \in B, \text{ there is a unique } x \in A \text{ such that } f(x) = y.$$

Hence, if a function is bijective then there exists an inverse function $g : B \to A$ such that

$$\forall x \in A,\ g(f(x)) = x, \text{ and } \forall y \in B,\ f(g(y)) = y.$$

The goal of the first half of this course is an attempt to try and understand the 'linear functions' whose domain and codomain are vector spaces. We will investigate if there is a way to represent the function in such a way that all desirable information we would like to know about the function is easy to obtain. In particular, we will provide (finite) criteria that allow us to determine if a function is injective/surjective (cf. Theorem 1.7.4).

Remark. These properties of a function can be difficult to grasp at first. Students tend to find that injectivity is the hardest attribute of a function to comprehend. The next example is an attempt at providing a simple introduction to the concept of injectivity/surjectivity of functions.

Example 0.2.3. Consider the set P described above (so an object in P is a person in Etcheverry, room 3109, at 10.10am on 6/18/2012) and let $\mathcal{C}$ denote the set of all possible cookie ice cream sandwiches available at C.R.E.A.M. on Telegraph Avenue (for example, vanilla ice cream on white chocolate chip cookies). Consider the following function

$$f : P \to \mathcal{C} \; ; \; x \mapsto f(x) = x\text{'s favourite cookie ice cream sandwich.}$$

In order for f to define a function we are assuming that nobody who is an element of P is indecisive so that they have precisely one favourite cookie ice cream sandwich.[2]

So, for example,

$$f(\textit{George}) = \text{ banana walnut ice cream on chocolate chip cookies.}$$

What does it mean for f to be

- injective? Let's go back to the definition: we require that for any two people $x, y \in P$, if $f(x) = f(y)$ then $x = y$, ie, if any two people in P have the same favourite cookie ice cream sandwich then those two people must be the same person. Or, what is the same, no two people in P have the same favourite cookie ice cream sandwich.
- surjective? Again, let's go back to the definition: we require that, if $y \in \mathcal{C}$ then there exists some $x \in P$ such that $f(x) = y$, ie, for any possible cookie ice cream sandwich y available at C.R.E.A.M. there must exist some person $x \in P$ for which y is x's favourite cookie ice cream sandwich.

There are a couple of things to notice here:

1. in order for f to be surjective, we must necessarily have at least as many objects in P as there are objects in $\mathcal{C}$. That is
$$f \text{ surjective } \implies |P| \geq |\mathcal{C}|.$$
2. in order for f to be injective, there must necessarily be more objects in $\mathcal{C}$ as there are in P. That is
$$f \text{ injective } \implies |P| \leq |\mathcal{C}|.$$
3. if P and $\mathcal{C}$ have the same number of objects then f is injective if and only if f is surjective.

You should understand and provide a short proof as to why these properties hold true.

The fact that these properties are true is dependent on the fact that both P and $\mathcal{C}$ are *finite sets*. We will see a generalisation of these properties to finite dimensional vector spaces and linear morphisms between them: here we replace the 'size' of a vector space by its dimension (a linear algebra measure of 'size').

We will now include a basic lemma that will be useful throughout these notes. Its proof is left to the reader.

Lemma 0.2.4. *Let $f : R \to S$ and $g : S \to T$ be two functions.*

[2] Why are we making this assumption?

- *If f and g are both injective, then $g \circ f : R \to T$ is injective. Moreover, if $g \circ f$ is injective then f is injective.*

- *If f and g are both surjective, then $g \circ f : R \to T$ is surjective. Moreover, if $g \circ f$ is surjective then g is surjective.*

- *If f and g are bijective, then $g \circ f : R \to T$ is bijective.*

1 Vector Spaces & Linear Morphisms

This chapter is intended as a reintroduction to results that you have probably seen before in your previous linear algebra course. However, we will adopt a slightly more grown-up viewpoint and discuss some subtleties that arise when we are considering infinite dimensional vector spaces. Hopefully most of the following results are familiar to you - don't forget or disregard any previous intuition you have gained and think of the following approach as supplementing those ideas you are already familiar with.

1.1 Fields

[[1] p. 1-3]

In your previous linear algebra course (eg. Math 54) you will have mostly worked with column (or row) vectors with real or complex coefficients. However, most of linear algebra does not require that we work only with real or complex entries, only that the set of '*scalars*' we use satisfy some nice properties.

For those of you who have taken an Abstract Algebra course (eg. Math 113) you may have already been introduced to the notion of a *ring* or a *field*. What follows is a very brief introduction to (number) fields.

Definition 1.1.1 (Number Field)**.** A nonempty set $\mathbb{K}$[3] is called an *number field* if

1. $\mathbb{Z} \subset \mathbb{K}$,
2. there are well-defined notions of addition, subtraction, multiplication and division that obey all the usual laws of arithmetic.

Note that, by 1. we have that $\mathbb{K}$ contains every integer $x \in \mathbb{Z}$. Therefore, by 2., since we must be able to divide through by nonzero x, we necessarily have $\mathbb{Q} \subset \mathbb{K}$.

Example 1.1.2. 1. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all examples of number fields. However, $\mathbb{Z}$ is not a number field since 2 does not have a multiplicative inverse in $\mathbb{Z}$.

2. Consider the set

$$\mathbb{Q}(\sqrt{2}) = \{a + b\Delta \mid a, b \in \mathbb{Q}\},$$

where we consider Δ as some symbol. Define an addition on $\mathbb{Q}(\sqrt{2})$ as follows: for $z = a + b\Delta, w = c + d\Delta \in \mathbb{Q}(\sqrt{2})$, define

$$z + w = (a + c) + (b + d)\Delta \in \mathbb{Q}(\sqrt{2}), \quad -z = -a + (-b)\Delta,$$

$$z.w = (ac + 2bd) + (ad + bc)\Delta, \quad z^{-1} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\Delta.$$

Note that $a^2 - 2b^2 \neq 0$, for any $a, b \in \mathbb{Q}$ such that $(a, b) \neq (0, 0)$, since $\sqrt{2}$ is irrational.

Then, it is an exercise to check that all the usual rules of arithmetic (eg. Rules 1-9 on p.1 of [1]) hold for this addition and multiplication just defined. Moreover, it is easy to check that

$$(1.\Delta)^2 = 2,$$

so that we can justify swapping the symbol $\sqrt{2}$ for Δ.

Be careful though: we do not care about the actual value of $\sqrt{2}$ (=1.414...) as a real number, we are only interested in the algebraic properties of this number, namely that it squares to 2, and as such only consider $\sqrt{2}$ as a symbol such that $\sqrt{2}^2 = 2$ in our number field $\mathbb{Q}(\sqrt{2})$.

You should think of $\mathbb{Q}(\sqrt{2})$ in the same way as $\mathbb{C}$: for our purposes of arithmetic, we do not care whether $\sqrt{2}$ is a 'real' number or not, we only care about its basic algebraic properties as a symbol.

[3] We adopt the letter $\mathbb{K}$ for a (number) field following the Germans. In German the word for a '(number) field' is *'(zahlen) korps'*. The notation $\mathbb{Z}$ for the integers also comes from the German word *zahlen*, meaning 'number'. Most of the basics of modern day algebra was formulated and made precise by German mathematicians, the foremost of whom being C. F. Gauss, D. Hilbert, R. Dedekind, E. Noether, E. Steinitz and many, many others.

Why should we care about $\mathbb{Q}(\sqrt{2})$ when we can just think about $\mathbb{R}$? Most of modern number theory is concerned with the study of number fields and there is some sense in which the deep structure of number fields is related to seemingly unrelated areas of mathematics such as real analysis.

3. Let $p > 0$ be a nonsquare integer, so that there does not exist $x \in \mathbb{Z}$ such that $x^2 = p$. Then, we can form the number field $\mathbb{Q}(\sqrt{p})$ in a similar manner as above.

4. The p-adic numbers $\mathbb{Q}_p$: let p be a prime number. Then, you may have heard of the *p-adic numbers*: this is a number field that is obtained from $\mathbb{Q}$ in a similar way that $\mathbb{R}$ can be obtained from $\mathbb{Q}$ (via Cauchy sequences; this is Math 104 material). Essentially, a p-adic number $a \in \mathbb{Q}_p$ can be considered as a formal power series

$$a = \sum_{i \geq m} a_i p^i = a_m p^m + a_{m+1} p^{m+1} + \cdots, \quad a_i \in \{0, 1, \ldots, p-1\}, m \in \mathbb{Z},$$

where we do not care about the fact that this 'sum' does not converge and add and multiply as you would as if you were in high school (remembering to reduce coefficients modulo p). We will not talk about this number field again and if you are interested in learning more simply Google 'p-adic numbers' and there will be plenty information available online.

5. The field of rational polynomials $\mathbb{Q}(t)$: here we have

$$\mathbb{Q}(t) = \{p/q \mid p, q \in \mathbb{Q}[t],\ q \neq 0\},$$

where $\mathbb{Q}[t]$ is the set of polynomials with rational coefficients. For example,

$$\frac{3 - \frac{5}{3}t^8}{t^2 + \frac{2}{7}t^{167}} \in \mathbb{Q}(t).$$

Again, $\mathbb{Q}(t)$ is a number field and arises in algebraic geometry, that area of mathematics concerned with solving systems of polynomials equations (it's very hard!).

Remark. 1. The definition of 'number field' given above is less general than you might have seen: in general, a *field* $\mathbb{K}$ is a nonempty set for which there are well-defined notions of addition, subtraction, multiplication and division (and obeying all the usual laws of arithmetic) without the extra requirement that $\mathbb{Z} \subset \mathbb{K}$; this is the definition given in [1]. The definition we have given in Definition 1.1.1 defines a *field of characteristic zero*.

2. In this course we will only be concerned with 'linear algebra over number fields', meaning the scalars we consider will have to take values in a number field $\mathbb{K}$ as defined in Definition 1.1.1. Moreover, most of the time we will take $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ and when we come to discuss Euclidean (resp. Hermitian) spaces we must necessarily have $\mathbb{K} = \mathbb{R}$ (resp. $\mathbb{K} = \mathbb{C}$).

In grown-up mathematical language we will be studying '**vector spaces over characteristic zero fields**' (or, in an even more grown-up language, $\mathbb{K}$-modules, where $\mathbb{K}$ is a characteristic zero field).

From now on, $\mathbb{K}$ will always denote a number field.

1.2 Vector Spaces

1.2.1 Basic Definitions

[p.31-33, [1]]

Definition 1.2.1 (Vector Space)**.** A $\mathbb{K}$-*vector space* (or *vector space over* $\mathbb{K}$) is a triple (V, α, σ), where V is a nonempty set and

$$\alpha : V \times V \to V\ ;\ (u, v) \mapsto \alpha(u, v), \quad \sigma : \mathbb{K} \times V \to V\ ;\ (\lambda, v) \mapsto \sigma(\lambda, v),$$

are two functions called *addition* and *scalar multiplication*, and such that the following axioms are imposed:

(VS1) α is associative: for every $u, v, w \in V$ we have

$$\alpha(u, \alpha(v, w)) = \alpha(\alpha(u, v), w);$$

(VS2) α is commutative: for every $u, v \in V$ we have

$$\alpha(u, v) = \alpha(v, u);$$

(VS3) there exists an element $0_V \in V$ such that, for every $v \in V$, we have

$$\alpha(0_V, v) = \alpha(v, 0_V) = v.$$

We call 0_V a (in fact, *the*[4]) *zero element* or *zero vector* of V;

(VS4) for every $v \in V$ there exists an element $\hat{v} \in V$ such that

$$\alpha(v, \hat{v}) = \alpha(\hat{v}, v) = 0_V.$$

We call $\hat{v}$ the[5] *negative of* v and denote it $-v$;

(VS5) for every $\lambda, \mu \in \mathbb{K}$, $v \in V$, we have

$$\sigma(\lambda + \mu, v) = \alpha(\sigma(\lambda, v), \sigma(\mu, v));$$

(VS6) for every $\lambda, \mu \in \mathbb{K}$, $v \in V$, we have

$$\sigma(\lambda\mu, v) = \sigma(\lambda, \sigma(\mu, v));$$

(VS7) for every $\lambda \in \mathbb{K}$, $u, v \in V$, we have

$$\sigma(\lambda, \alpha(u, v)) = \alpha(\sigma(\lambda, u), \sigma(\lambda, v));$$

(VS8) for every $v \in V$ we have

$$\sigma(1, v) = v.$$

In case α and σ satisfy the above axioms so that (V, α, σ) is a vector space (over $\mathbb{K}$) we will usually write

$$\begin{array}{c} \alpha(u, v) = u + v, \quad u, v, \in V, \\ \sigma(\lambda, v) = \lambda \cdot v, \quad \text{or simply} \quad \sigma(\lambda, v) = \lambda v, \quad \lambda \in \mathbb{K}, v \in V. \end{array}$$

If (V, α, σ) is a vector space over $\mathbb{K}$ then we will call an element $x \in V$ a *vector* and an element $\lambda \in \mathbb{K}$ a *scalar*.

If $v \in V$ is such that

$$v = \lambda_1 v_1 + ... + \lambda_n v_n,$$

for some vectors $v_1, \dots, v_n \in V$ and scalars $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, then we say that v is a *linear combination* of the vectors $v_1, \dots, v_n$.

[4] See Proposition 1.2.3.

[5] We shall see (Proposition 1.2.4) that the negative of v is unique, so that this notation is well-defined.

Remark 1.2.2. The given definition of a vector space might look cumbersome given the introduction of the functions α and σ. However, it is important to realise that these defined notions of addition and scalar multiplication tell us how we are to 'add' vectors and how we are to 'scalar multiply' vectors by scalars in $\mathbb{K}$; in particular, a nonempty set V may have many different ways that we can define a vector space structure on it, ie, it may be possible that we can obtain two distinct $\mathbb{K}$-vector spaces (V, α, σ) and (V, α', σ') which have the same underlying set but different notions of addition and scalar multiplication. In this case, it is important to know which 'addition' (ie, which function α or α') we are discussing, or which 'scalar multiplication' (ie, which function σ or σ') we are discussing.

In short, the definition of a vector space is the data of providing a nonempty set together with the rules we are using for 'addition' and 'scalar multiplication'.

Notation. • Given a $\mathbb{K}$-vector space (V, α, σ) we will usually know which notions of addition and scalar multiplication we will be discussing so we will often just write V instead of the triple (V, α, σ), the functions α, σ being understood *a priori*.

• We will also frequently denote an arbitrary vector space (V, α, σ) by V, even when we don't know what α, σ are explicitly. Again, we are assuming that we have been given 'addition' and 'scalar multiplication' functions *a priori*.

Proposition 1.2.3. *Let (V, α, σ) be a $\mathbb{K}$-vector space. Then, a zero vector $0_V \in V$ is unique.*

Proof: Suppose there is some element $z \in V$ such that z satisfies the same properties as 0_V, so that, for every $v \in V$, we have $z + v = v + z = v$. Then, in particular, we have

$$0_V = z + 0_V = z,$$

where the first equality is due to the characterising properties of z as a zero vector (Axiom VS3), and the second equality is due to the characterising property of 0_V as a zero vector (Axiom VS3). Hence, $z = 0_V$ so that a zero vector is unique. □

Proposition 1.2.4 (Uniqueness of negatives)**.** *Let (V, α, σ) be a $\mathbb{K}$-vector space. Then, for every $v \in V$, the element $\hat{v}$ that exists by Axiom VS4 is unique.*

Proof: Let $v \in V$ and suppose that $w \in V$ is such that $w + v = v + w = 0_V$. Then,

$$\begin{aligned} w = w + 0_V &= w + (v + \hat{v}), && \text{by defining property of } \hat{v}, \\ &= (w + v) + \hat{v}, && \text{by Axiom VS1}, \\ &= 0_V + \hat{v}, && \text{by assumed property of } w, \\ &= \hat{v}, && \text{by Axiom VS3}. \end{aligned}$$

Hence, $w = \hat{v}$ and the negative of v is unique. □

Proposition 1.2.5 (Characterising the zero vector)**.** *Let (V, α, σ) be a $\mathbb{K}$-vector space. Then, for every $v \in V$ we have $0 \cdot v = 0_V$. Moreover, $\lambda \cdot 0_V = 0_V$, for every $\lambda \in \mathbb{K}$. Conversely, if $\lambda \cdot v = 0_V$ with $v \neq 0_V$, then $\lambda = 0 \in \mathbb{K}$.*

Proof: Let $v \in V$. Then, noting the trivial fact that $0 = 0 + 0 \in \mathbb{K}$, we have

$$\begin{aligned} & 0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v, && \text{by Axiom VS5}, \\ \implies & 0_V = 0 \cdot v + (-0 \cdot v) = (0 \cdot v + 0 \cdot v) + (-0 \cdot v) = 0 \cdot v + (0 \cdot v + (-0 \cdot v)), && \text{using Axiom VS1}, \\ \implies & 0_V = 0 \cdot v + 0_V = 0 \cdot v, && \text{using Axioms VS3 and VS4}. \end{aligned}$$

Furthermore, in a similar way (using Axioms VS3, VS4 and VS7) we can show that $\lambda \cdot 0_V = 0_V$, for every $\lambda \in \mathbb{K}$.[6]

[6] Do this as an exercise!

Conversely, suppose that $v \neq 0_V$ is a vector in V and $\lambda \in \mathbb{K}$ is such that $\lambda v = 0_V$. Assume that $\lambda \neq 0$; we aim to provide a contradiction. Then, λ^{-1} exists and we have

$$\begin{aligned} v = 1 \cdot v &= (\lambda^{-1}\lambda) \cdot v, \quad \text{using Axiom VS8,} \\ &= \lambda^{-1} \cdot (\lambda \cdot v), \quad \text{by Axiom VS6,} \\ &= \lambda^{-1} \cdot 0_V, \quad \text{using our assumption,} \\ &= 0_V, \quad \text{by the result just proved.} \end{aligned}$$

But this contradicts our assumption that v is nonzero. Hence, our initial assumption that $\lambda \neq 0$ cannot hold so that $\lambda = 0 \in \mathbb{K}$.

□

The following examples will be fundamental for the rest of the course so make sure that you acquaint yourself with them as they will be used frequently throughout class and on homework/exams. As such, if you are having trouble understanding them then please ask a fellow student for help or feel free to send me an email and I will help out as best I can.

I have only provided the triple (V, α, σ) in each example, you should *define the zero vector in each example and the negative of an arbitrary given vector* $v \in V$. Also, you should check that the Axioms VS1-VS8 hold true.

Example 1.2.6. 1. For $n \in \mathbb{N}$, consider the $\mathbb{K}$-vector space $(\mathbb{K}^n, \alpha, \sigma)$, where

$$\mathbb{K}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mid x_1, \dots, x_n \in \mathbb{K} \right\},$$

$$\alpha\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right) = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \text{and} \quad \sigma\left(\lambda, \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}.$$

We will usually denote this vector space simply $\mathbb{K}^n$. It is of fundamental importance in all that follows.

We will denote by $e_i \in \mathbb{K}^n$, the column vector that has a 1 in the i^{th} entry and 0s elsewhere, and call it the i^{th} *standard basis vector*.

2. Let S be an arbitrary nonempty set. Then, define the $\mathbb{K}$-vector space $(\mathbb{K}^S, \alpha, \sigma)$, where

$$\mathbb{K}^S = \{\text{functions } f : S \to \mathbb{K}\},$$

$$\alpha(f, g) : S \to \mathbb{K} \; ; \; s \mapsto f(s) + g(s) \in \mathbb{K}, \quad \text{and} \quad \sigma(\lambda, f) : S \to \mathbb{K} \; ; \; s \mapsto \lambda f(s).$$

That is, we have defined the sum of two functions $f, g \in \mathbb{K}^S$ to be the new function $\alpha(f, g) : S \to \mathbb{K}$ such that $\alpha(f, g)(s) = f(s) + g(s)$ (here the addition is taking place inside the number field $\mathbb{K}$).

Since this example can be confusing at first sight, I will give you the zero vector $0_{\mathbb{K}^S}$ and the negative of a vector $f \in \mathbb{K}^S$: since the elements in $\mathbb{K}^S$ are functions we need to ensure that we define a function

$$0_{\mathbb{K}^S} : S \to \mathbb{K},$$

satisfying the properties required of Axiom VS3. Consider the function

$$0_{\mathbb{K}^S} : S \to \mathbb{K} \; ; \; s \mapsto 0,$$

that is, $0_{\mathbb{K}^S}(s) = 0 \in \mathbb{K}$, for every $s \in S$. Let's show that this function just defined satisfies the properties required of a zero vector in $(\mathbb{K}^S, \alpha, \sigma)$. So, let $f \in \mathbb{K}^S$, ie,

$$f : S \to \mathbb{K} \; ; \; s \mapsto f(s).$$

Then, we have, for every $s \in S$,

$$\alpha(f, 0_{\mathbb{K}^S})(s) = f(s) + 0_{\mathbb{K}^S}(s) = f(s) + 0 = f(s),$$

so that $\alpha(f, 0_{\mathbb{K}^S}) = f$. Similarly, we have $\alpha(0_{\mathbb{K}^S}, f) = f$. Hence, $0_{\mathbb{K}^S}$ satisfies the properties required of the zero vector in $\mathbb{K}^S$ (Axiom VS3).

Now, let $f \in \mathbb{K}^S$. We define a function $-f \in \mathbb{K}^S$ as follows:

$$-f : S \to \mathbb{K} \; ; \; s \mapsto -f(s) \in \mathbb{K}.$$

Then, $-f$ satsfies the properties required of the negative of f (Axiom VS4).

For every $s \in S$, we define the *characteristic function of* s, $e_s \in \mathbb{K}^S$, where

$$e_s(t) = \begin{cases} 0, & \text{if } t \neq s, \\ 1, & \text{if } t = s. \end{cases}$$

For example, if $S = \{1, 2, 3, 4\}$ then

$$\mathbb{K}^S = \{\text{functions } f : \{1, 2, 3, 4\} \to \mathbb{K}\}.$$

What is a function $f : \{1, 2, 3, 4\} \to \mathbb{K}$? To each $i \in \{1, 2, 3, 4\}$ we associate a scalar $f(i) \in \mathbb{K}$, which we can also denote $f_i \stackrel{def}{=} f(i)$. This choice of notation should lead you to think there is some kind of similarity between $\mathbb{K}^4$ and $\mathbb{K}^{\{1,2,3,4\}}$; indeed, these two vector spaces are *isomorphic*, which means they are essentially the same (in the world of linear algebra).

For example, we have the characteristic function $e_2 \in \mathbb{K}^{\{1,2,3,4\}}$, where

$$e_2(1) = e_2(3) = e_2(4) = 0, \quad e_2(2) = 1.$$

3. Let $m, n \in \mathbb{N}$ and consider the sets $[m] = \{1, \dots, m\}, [n] = \{1, \dots, n\}$. Then, we have the set

$$[m] \times [n] = \{(x, y) \mid x \in [m], y \in [n]\}.$$

Then, we define the set of $m \times n$ *matrices with entries in* $\mathbb{K}$ to be

$$Mat_{m,n}(\mathbb{K}) \stackrel{def}{=} \mathbb{K}^{[m]\times[n]}.$$

Hence, an $m \times n$ *matrix* A is a function $A : [m] \times [n] \to \mathbb{K}$, ie, a matrix is completely determined by the values $A(i, j) \in \mathbb{K}$, for $(i, j) \in [m] \times [n]$. We will denote such an $m \times n$ matrix in the usual way:

$$A \equiv \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix},$$

where we have $A_{ij} \stackrel{def}{=} A(i, j)$. Thus, we add and scalar multiply $m \times n$ matrices 'entry-wise'.

We will denote the zero vector $0_{Mat_{m,n}(\mathbb{K})} \in Mat_{m,n}(\mathbb{K})$ by $0_{m,n}$.

4. This is a generalisation of the previous examples. Let (V, β, τ) be a vector space and S any nonempty set. Define the $\mathbb{K}$-vector space (V^S, α, σ) where

$$V^S \stackrel{def}{=} \{\text{functions } f : S \to V\},$$

$$\alpha(f, g) : S \to V \; ; \; s \mapsto \beta(f(s), g(s)), \quad \text{and} \quad \sigma(\lambda, f) : S \to V \; ; \; s \mapsto \tau(\lambda, f(s)).$$

This might look a bit confusing to you so let's try and make things a bit clearer: denote the addition afforded by β as $\hat{+}$, so that if $u, v \in V$ then the addition defined in V is denoted $u\hat{+}v$. If we denote the addition we have defined by α in V^S as $\tilde{+}$, then the previous definition states that

$$\forall f, g \in V^S, \text{ define } f\tilde{+}g(=\alpha(f,g)) \in V^S \text{ by } (f\tilde{+}g)(s) = f(s)\hat{+}g(s)(=\beta(f(s),g(s))) \in V.$$

5. Consider the set

$$\mathbb{K}[t] \stackrel{def}{=} \{a_0 + a_1 t + ... + a_m t^m \mid a_m \in \mathbb{K}, m \in \mathbb{Z}_{\geq 0}\},$$

of polynomials with coefficients in $\mathbb{K}$. Define the $\mathbb{K}$-vector space $(\mathbb{K}[t], \alpha, \sigma)$ where, for

$$f = a_0 + a_1 t + ... + a_m t^m,\ g = b_0 + b_1 t + ... + b_n t^n \in \mathbb{K}[t],$$

with $m \leq n$, say, we have

$$\alpha(f,g) = (a_0 + b_0) + (a_1 + b_1)t + ... + (a_m + b_m)t^m + b_{m+1}t^{m+1} + ... + b_n t^n.$$

Also, we have, for $\lambda \in \mathbb{K}$,

$$\sigma(\lambda, f) = (\lambda a_0) + (\lambda a_1)t + ... + (\lambda a_m)t^m.$$

That is, for any two polynomials $f, g \in \mathbb{K}[t]$ we simply add and scalar multiply them 'coefficient-wise'.

For $n \in \mathbb{Z}_{\geq 0}$ we define the vector spaces $(\mathbb{K}_n[t], \alpha_n, \sigma_n)$ where,

$$\mathbb{K}_n[t] = \{f = a_0 + a_1 t + ... + a_m t^m \in \mathbb{K}[t] \mid m \leq n\},$$

$$\alpha_n(f,g) = \alpha(f,g), \quad \text{and} \quad \sigma_n(\lambda, f) = \sigma(\lambda, f).$$

Here $\mathbb{K}_n[t]$ is the set of all polynomials with coefficients in $\mathbb{K}$ of degree at most n; the fact that we have defined the addition and scalar multiplication in $\mathbb{K}_n[t]$ via the addition and scalar multiplication of $\mathbb{K}[t]$ is encoded in the notion of a *vector subspace*.

$(*)$ There is an important (and subtle) point to make here: a vector $f \in \mathbb{K}[t]$ (or $\mathbb{K}_n[t]$) is just a formal expression

$$f = a_0 + a_1 t + ... + a_n t^n, \quad a_0, ..., a_n \in \mathbb{K}.$$

This means that we are only considering a polynomial as a formal symbol that we add and scalar multiply according to the rules we have given above. In particular, **two polynomials $f, g \in \mathbb{K}[t]$ are equal, so that $f = g \in \mathbb{K}[t]$, if and only if they are equal coefficient-wise**.

This might either seem obvious or bizarre to you. I have included this comment as most students are used to seeing polynomials considered as *functions*

$$f : \mathbb{K} \to \mathbb{K}\ ;\ t \mapsto f(t) = \sum_{i=0}^{n} a_i t^i,$$

and I am saying that we do not care about this interpretation of a polynomial, we are only concerned with its formal (linear) algebraic properties. This is why I will write 'f' instead of '$f(t)$' for a polynomial. You might wonder why we even bother writing a polynomial as

$$f = a_0 + a_1 t + ... + a_n t^n,$$

when we don't care about the powers of t that are appearing, we could just write

$$f = (a_0, ..., a_n),$$

and 'represent' the polynomial f by this row vector (or, even a column vector). Well, the $\mathbb{K}$-vector space of polynomials has a further property, it is an example of a $\mathbb{K}$-algebra: for those of you who will take Math 113, this means $\mathbb{K}[t]$ is not only a $\mathbb{K}$-vector space, it is also has the structure of an algebraic object called a *ring*. This extra structure arises from the fact that we can multiply polynomials together (in the usual way) and in this context we do care about the powers of t that appear.

6. All the previous examples of $\mathbb{K}$-vector spaces have underlying sets that are *infinite*. What happens if we have a $\mathbb{K}$-vector space (V, α, σ) whose underlying set V is *finite*?

First, we give an example of a $\mathbb{K}$-vector space containing one element: we define the[7] *trivial* $\mathbb{K}$*-vector space* to be $(\underline{Z}, \alpha, \sigma)$ where $\underline{Z} = \{0_{\underline{Z}}\}$ and

$$\alpha : \underline{Z} \times \underline{Z} \to \underline{Z} \,;\, (0_{\underline{Z}}, 0_{\underline{Z}}) \mapsto 0_{\underline{Z}}, \quad \text{and} \quad \sigma : \mathbb{K} \times \underline{Z} \to \underline{Z} \,;\, (\lambda, 0_{\underline{Z}}) \mapsto 0_{\underline{Z}}.$$

This defines a structure of a $\mathbb{K}$-vector space on $\underline{Z}$.[8]

Now, recall that we are assuming that $\mathbb{K}$ is a number field so that $\mathbb{Z} \subset \mathbb{K}$ and $\mathbb{K}$ must therefore be infinite. Also, let's write (as we will do for the rest of these notes) λv instead of $\sigma(\lambda, v)$, for $v \in V, \lambda \in \mathbb{K}$.

Since V defines a vector space then we must have the zero element $0_V \in V$ (Axiom VS3). Suppose that there exists a nonzero vector $w \in V$ (ie $w \neq 0_V$); we aim to provide a contradiction, thereby showing that V *must contain exactly one element*. Then, since $\lambda w \in V$, for every $\lambda \in \mathbb{K}$, and $\mathbb{K}$ is infinite we must necessarily have distinct scalars $\mu_1, \mu_2 \in \mathbb{K}$ such that $\mu_1 w = \mu_2 w$ (else, all the λw's are distinct, for all possible $\lambda \in \mathbb{K}$. Since there are an infinite number of these we can't possibly have V finite). Furthermore, we assume that both μ_1 and μ_2 are nonzero scalars[9]. Hence, we have

$$w = 1w = (\mu_1^{-1}\mu_1)w = \mu_1^{-1}(\mu_1 w) = \mu_1^{-1}(\mu_2 w) = (\mu_1^{-1}\mu_2)w,$$

where we have used Axiom VS6 for the third and fifth equalities and our assumption for the fourth equality.

Hence, adding $-(\mu_1^{-1}\mu_2)w$ to both sides of this equation gives

$$w + (-(\mu_1^{-1}\mu_2)w) = 0_V \implies (1 - \mu_1^{-1}\mu_2)w = 0_V, \text{ by Axiom VS5.}$$

Therefore, by Proposition 1.2.5, we must have

$$1 - \mu_1^{-1}\mu_2 = 0 \in \mathbb{K} \implies \mu_1 = \mu_2,$$

contradicting the fact that μ_1 and μ_2 are distinct. Hence, our intial assumption - that there exists a nonzero vector $w \in V$ - cannot hold true, so that $V = \{0_V\}$.

Any $\mathbb{K}$-vector space (V, α, σ) for which V is a finite set must be a trivial $\mathbb{K}$-vector space.

7. The set of complex numbers $\mathbb{C}$ is a $\mathbb{R}$-vector space with the usual addition and scalar multiplication (scalar multiply $z \in \mathbb{C}$ by $x \in \mathbb{R}$ as $xz \in \mathbb{C}$). Moreover, both $\mathbb{R}$ and $\mathbb{C}$ are $\mathbb{Q}$-vector spaces with the usual addition and scalar multiplication.

However, $\mathbb{R}$ is not a $\mathbb{Q}(\sqrt{-1})$-vector space, where $\mathbb{Q}(\sqrt{-1}) = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Q}\}$ with the arithmetic laws defined in a similar way as we have defined for $\mathbb{C}$.

Moreover, $\mathbb{Q}$ is not a $\mathbb{R}$-vector space nor a $\mathbb{C}$-vector space; $\mathbb{R}$ is not a $\mathbb{C}$-vector space.

Remark 1.2.7. From now on we will no longer denote a particular $\mathbb{K}$-vector space that appears in Example 1.2.6 as a triple (V, α, σ) but only denote its underlying set V, the operations of addition and scalar multiplication being understood to be those appearing above. We will also write $u + v$ (resp. λv) instead of $\alpha(u, v)$ (resp. $\sigma(\lambda, v)$) for these examples.

[7] Technically, we should say 'a' instead of 'the' as any one-point set defines a $\mathbb{K}$-vector space and all of them are equally 'trivial'. However, we can show that all of these trivial $\mathbb{K}$-vector spaces are *isomorphic* (a notion to be defined in the next section) so that, for the purposes of linear algebra, they are all (essentially) the same. For our purposes we will not need to care about this (extremely subtle) distinction.

[8] Exercise: check this. Note that this is the *only* possible $\mathbb{K}$-vector space structure we can put on $\underline{Z}$. Moreover, if you think about it, you will see that *any* set with one element defines a $\mathbb{K}$-vector space.

[9] Why? Try and prove this as an exercise.

1.2.2 Subspaces

[p.42, [1]]

Definition 1.2.8 (Subspace)**.** Let (V, α, σ) be a $\mathbb{K}$-vector space and $U \subset V$ a nonempty subset. Then, we say that U is a *vector subspace of* V if the following properties hold:

(SUB1) $0_V \in U$,

(SUB2) for every $u, v \in U$, we have $\alpha(u, v) \in U$, (*closed under addition*)

(SUB3) for every $\lambda \in \mathbb{K}, u \in U$, we have $\sigma(\lambda, u) \in U$. (*closed under scalar multiplication*)

In fact, we can subsume these three properties into the following single property

(SUB) for every $u, v \in U, \mu, \lambda \in \mathbb{K}$, we have $\alpha(\sigma(\mu, u), \sigma(\lambda, v)) \in U$ (ie, $\mu u + \lambda v \in U$).

In this case, U can be considered as a $\mathbb{K}$-vector space in its own right: we have a triple $(U, \alpha_{|U}, \sigma_{|U})$ where $\alpha_{|U}$ (resp. $\sigma_{|U}$) denote the functions α (resp. σ) *restricted to* U.[10] Notice that we need to ensure that U is closed under addition (and scalar multiplication) in order that the functions $\alpha_{|U}$ and $\sigma_{|U}$ are well-defined.

Example 1.2.9. Recall the examples from Example 1.2.6 and our conventions adopted thereafter (Remark 1.2.7).

0. There are always two obvious subspaces of a $\mathbb{K}$-vector space V: namely, V is a subspace of itself, and the subset $\{0_V\} \subset V$ is a subspace of V called the *zero subspace*. We call these subspaces the *trivial subspaces of* V. All other subspaces are called *nontrivial*.

1. Consider the $\mathbb{Q}$-vector space $\mathbb{Q}^3$ and the subset

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{Q}^3 \mid x_1 + x_2 - 2x_3 = 0 \right\}.$$

Then, U is a $\mathbb{Q}$-vector space.

How can we confirm this? We need to show that U satisfies the Axiom SUB from Definition 1.2.8. So, let $u, v \in U$ and $\mu, \lambda \in \mathbb{Q}$. Thus,

$$u = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \text{with} \quad x_1 + x_2 - 2x_3 = 0, \quad \text{and}$$

$$v = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \text{with} \quad y_1 + y_2 - 2y_3 = 0.$$

Then,

$$\mu u + \lambda v = \begin{bmatrix} \mu x_1 \\ \mu x_2 \\ \mu x_3 \end{bmatrix} + \begin{bmatrix} \lambda y_1 \\ \lambda y_2 \\ \lambda y_3 \end{bmatrix} = \begin{bmatrix} \mu x_1 + \lambda y_1 \\ \mu x_2 + \lambda y_2 \\ \mu x_3 + \lambda y_3 \end{bmatrix},$$

and to show that $\mu u + \lambda v = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in U$ we must show that $z_1 + z_2 - 2z_3 = 0$: indeed, we have

$$\begin{aligned} z_1 + z_2 - 2z_3 &= (\mu x_1 + \lambda y_1) + (\mu x_2 + \lambda y_2) - 2(\mu x_3 + \lambda y_3), \\ &= \mu(x_1 + x_2 - 2x_3) + \lambda(y_1 + y_2 - 2y_3), \\ &= 0 + 0 = 0. \end{aligned}$$

[10] Here is an important but subtle distinction: the functions α and $\alpha_{|U}$ are not the same functions. Recall that when we define a function we must also specify its domain and codomain. The functions α and $\alpha_{|U}$ are defined as

$$\alpha : V \times V \to V \; ; \; (u, v) \mapsto \alpha(u, v), \quad \alpha_{|U} : U \times U \to U \; ; \; (u', v') \mapsto \alpha(u', v'),$$

So, technically, even though α and $\alpha_{|U}$ are defined by the same 'rule', they have different (co)domains so should be considered as different functions. The same reasoning holds for σ and $\sigma_{|U}$ (how are these functions defined?)

Hence, U is a vector subspace of $\mathbb{Q}^3$.

2. Consider the subset

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1 = 1 \right\} \subset \mathbb{C}^2,$$

of the $\mathbb{C}$-vector space $\mathbb{C}^2$. Then, U is not a vector subspace of $\mathbb{C}^2$.

How can we show that a given subset E of a $\mathbb{K}$-vector space V is not a vector subspace? We must show that E does not satisfy *all* of the Axioms SUB1-3 from Definition 1.2.8, so we need to show that *at least one* of these axioms fails to hold. If you are given some subset (eg. $U \subset \mathbb{C}^2$ above) and want to determine that it is not a subspace, the first thing to check is whether the zero vector is an element of this subset: for us, this means checking to see if $0_{\mathbb{C}^2} \in U$. This is easy to check: we have

$$0_{\mathbb{C}^2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and since the first entry of $0_{\mathbb{C}^2} \neq 1$ it is not an element in the set U. Hence, Axiom SUB1 does not hold for the subset $U \subset \mathbb{C}^2$ so that U is not a subspace of $\mathbb{C}^2$.

However, it is possible for a subset $E \subset V$ of a vector space to contain the zero vector and still not be a subspace.[11]

3. This is an example that requires some basic Calculus.

Consider the $\mathbb{R}$-vector space $\mathbb{R}^{(0,1)}$ consisting of all $\mathbb{R}$-valued functions

$$f : (0,1) \to \mathbb{R},$$

and the subset

$$C_{\mathbb{R}}(0,1) = \{f \in \mathbb{R}^{(0,1)} \mid f \text{ is continuous}\}.$$

Then, it is a fact proved in Math 1A that $C_{\mathbb{R}}(0,1)$ is a vector subspace of $\mathbb{R}^{(0,1)}$: namely, the (constant) zero function is continuous, the sum of two continuous functions is again a continuous function and a scalar multiple of a continuous function is a continuous function.

4. Consider the $\mathbb{Q}$-vector space $Mat_3(\mathbb{Q})$ of 3×3 matrices with $\mathbb{Q}$-entries. Then, for a 3×3 matrix A define the *trace of A* to be

$$\mathrm{tr}(A) = \mathrm{tr}\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \right) = a_{11} + a_{22} + a_{33} \in \mathbb{Q}.$$

Denote

$$sl_3(\mathbb{Q}) = \{A \in Mat_3(\mathbb{Q}) \mid \mathrm{tr}(A) = 0\},$$

the set of 3×3 matrices with trace zero. Then, $sl_3(\mathbb{Q})$ is a subspace of $Mat_3(\mathbb{Q})$. Let's check the Axioms SUB1-3 from Definition 1.2.8 (or, equivalently, you can just check Axiom SUB):

SUB1: recall that the zero vector in $Mat_3(\mathbb{Q})$ is just the zero matrix

$$0_{Mat_3(\mathbb{Q})} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, it is trivial to see that this matrix has trace zero so that $0_{Mat_3(\mathbb{Q})} \in sl_3(\mathbb{Q})$.

[11] For example, consider the subset $\mathbb{Q} \subset \mathbb{R}$ of the $\mathbb{R}$-vector space $\mathbb{R}$.

SUB2: let $A, B \in sl_3(\mathbb{Q})$ be two matrices with trace zero, so that

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \text{with} \quad a_{11} + a_{22} + a_{33} = 0, \quad \text{and}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}, \quad \text{with} \quad b_{11} + b_{22} + b_{33} = 0.$$

Then,

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \\ a_{31} + b_{31} & a_{32} + b_{32} & a_{33} + b_{33} \end{bmatrix},$$

and

$$(a_{11} + b_{11}) + (a_{22} + b_{22}) + (a_{33} + b_{33}) = (a_{11} + a_{22} + a_{33}) + (b_{11} + b_{22} + b_{33}) = 0 + 0 = 0,$$

so that $A + B \in sl_3(\mathbb{Q})$.

SUB3: let $A, \in sl_3(\mathbb{Q}), \lambda \in \mathbb{Q}$. Then,

$$\lambda A = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \lambda a_{13} \\ \lambda a_{21} & \lambda a_{22} & \lambda a_{23} \\ \lambda a_{31} & \lambda a_{32} & \lambda a_{33} \end{bmatrix},$$

and

$$\lambda a_{11} + \lambda a_{22} + \lambda a_{33} = \lambda(a_{11} + a_{22} + a_{33}) = \lambda.0 = 0.$$

Hence, $sl_3(\mathbb{Q})$ is a subspace of the $\mathbb{Q}$-vector space $Mat_3(\mathbb{Q})$.

5. Consider the subset $\mathrm{GL}_3(\mathbb{Q}) \subset Mat_3(\mathbb{Q})$, where

$$\mathrm{GL}_3(\mathbb{Q}) = \{A \in Mat_3(\mathbb{Q}) \mid \det(A) \neq 0\}.$$

Here, $\det(A)$ denotes the determinant of A that you should have already seen in Math 54 (or an equivalent introductory linear algebra course). Then, $\mathrm{GL}_3(\mathbb{Q})$ is not a vector subspace.

Again, we need to show that at least one of Axioms SUB1-3 does not hold: we will show that Axiom SUB2 does not hold. Consider the 3×3 identity matrix $I_3 \in Mat_3(\mathbb{Q})$. Then, $I_3 \in \mathrm{GL}_3(\mathbb{Q})$ and $-I_3 \in \mathrm{GL}_3(\mathbb{Q})$. However,

$$\det(I_3 + (-I_3)) = \det(0_{Mat_3(\mathbb{Q})}) = 0,$$

so that $\mathrm{GL}_3(\mathbb{Q})$ is not closed under addition, therefore is not a subspace of $Mat_3(\mathbb{Q})$.

Note that we could have also shown that $0_{Mat_3(\mathbb{Q})} \notin \mathrm{GL}_3(\mathbb{Q})$.[12]

6. This example generalises Examples 1 and 4 above.[13] Consider the subset

$$U = \left\{ \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n \mid A\underline{x} = \underline{0} \in \mathbb{K}^m \right\},$$

where A is an $m \times n$ matrix with entries in $\mathbb{K}$. Then, U is a vector subspace of $\mathbb{K}^n$.

In the next section we will see that we can interpret U as the *kernel* of a linear transformation

$$T_A : \mathbb{K}^n \to \mathbb{K}^m \; ; \; \underline{x} \mapsto A\underline{x},$$

[12] Here, the symbol $\notin$ should be translated as 'is not a member of' or 'is not an element of'.

[13] Once we have (re)considered linear transformations in the next section you should explain why we are generalising those particular examples.

defined by the matrix A. As such, we will leave the proof that U is a subspace until then.

Note here an important point: if $\underline{b} \neq \underline{0} \in \mathbb{K}^m$ then the subset

$$W = \left\{ \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n \mid A\underline{x} = \underline{b} \in \mathbb{K}^m \right\} \subset \mathbb{K}^n,$$

is *not* a vector subspace of $\mathbb{K}^n$. For example, $\underline{0}_{\mathbb{K}^n} \notin W$. This is the generalisation of Example 2 above.

So, a slogan might be something like

'**subspaces are kernels of linear maps**'.

In fact, this statement is more than just a slogan[14]:

Theorem. *Let V be a $\mathbb{K}$-vector space and $U \subset V$ a subspace. Then, there exists a $\mathbb{K}$-vector space W and a $\mathbb{K}$-linear morphism $\pi : V \to W$ such that $U = \ker \pi$.*

We will now provide some constructions that allow us to form new subspaces from old subspaces.

Definition 1.2.10 (Operations on Subspaces)**.** Let V be a $\mathbb{K}$-vector space, $U, W \subset V$ two subspaces of V. Then,

- the *sum of U and W* is the subspace[15]

$$U + W = \{u + w \in V \mid u \in U, w \in W\},$$

- the *intersection of U and W* is the subspace[16]

$$U \cap W = \{v \in V \mid v \in U \text{ and } v \in W\},$$

Moreover, these notions can be extended to arbitrary families of subspace $(U_j)_{j \in J}$, with each $U_j \subset V$ a subspace of V.

We say that the sum of U and W is a *direct sum*, if $U \cap W = \{0_V\}$ is the zero subspace of V. In this case we write

$$U \bigoplus W, \quad \text{instead of} \quad U + W. \quad \text{(cf. p.45, [1])}$$

Proposition 1.2.11. *Let V be a $\mathbb{K}$-vector space and $U, W \subset V$ vector subspaces. Then, $V = U \oplus W$ if and only if, for every $v \in V$ there exists unique $u \in U$ and unique $w \in W$ such that $v = u + w$.*

Proof: ($\Rightarrow$) Suppose that $V = U \oplus W$. By definition, this means that $V = U + W$ and $U \cap W = \{0_V\}$. Hence, for every $v \in V$ we have $u \in U, w \in W$ such that $v = u + w$ (dy the definition of the sum $U + W$). We still need to show that this expression is unique: if there are $u' \in U, w' \in W$ such that

$$u' + w' = v = u + w,$$

then we have

$$u' - u = w - w',$$

and the LHS of this equation is a vector in U (since it's a subspace) and the RHS is a vector in W (since it's a subspace). Hence, if we denote this vector y (so $y = u' - u = w - w'$) then we have $y \in U$ and $y \in W$ so that $y \in U \cap W$, by definition. Therefore, as $U \cap W = \{0_V\}$, we have $y = 0_V$ so that

$$u' - u = 0_V, \quad \text{and} \quad w - w' = 0_V,$$

giving $u = u'$ and $w = w'$ and the uniqueness is verified.

[14] The proof of the Theorem requires the notion of a *quotient space*.
[15] You will prove this for homework.
[16] You will prove this for homework.

($\Leftarrow$) Conversely, suppose that every vector $v \in V$ can be expressed uniquely as $v = u + w$ for $u \in U$ and $w \in W$. Then, the existence of this expression for each $v \in V$ is simply the statement that $V = U + W$. Moreover, let $x \in U \cap W$, so that $x \in U$ and $x \in W$. Thus, there are $u \in U$ and $w \in W$ (namely, $u = x$ and $w = x$) such that

$$0_V + w = x = u + 0_V,$$

and since U and W are subspaces (so that $0_V \in U, W$) we find, by the uniqueness of an expression for $x \in U \cap W \subset V$, that $u = 0_V = w$. Hence, $x = 0_V$ and $U \cap W = \{0_V\}$. □

1.3 Linear Dependence & $\text{span}_{\mathbb{K}}$

In this section we will make precise the notion of *linear (in)dependence*. This is a fundamental concept in linear algebra and abstracts our intuitive notion of *(in)dependent directions* when we consider the (Euclidean) plane $\mathbb{R}^2$ or (Euclidean) space $\mathbb{R}^3$.

Definition 1.3.1. Let V be a $\mathbb{K}$-vector space[17], and let $\{v_1, \dots, v_n\} \subset V$ be some subset. A *linear relation (over $\mathbb{K}$) among $v_1, \dots, v_n$* is an equation

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0_V, \tag{1.3.1}$$

where $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ are scalars.

If $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$ then we call (1.3.1) a *trivial linear relation (among $v_1, \dots, v_n$)*.

If at least one of $\lambda_1, \dots, \lambda_n$ is nonzero, so that

$$\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \neq \underline{0}_{\mathbb{K}^n},$$

then we call (1.3.1) a *nontrivial linear relation (among $v_1, \dots, v_n$)*.

Now, let $E \subset V$ be an arbitrary nonempty subset (possibly infinite; NOT necessarily a subspace). We say that E is *linearly dependent (over $\mathbb{K}$)* if there exists $v_1, \dots, v_n \in E$ and a nontrivial linear relation (over $\mathbb{K}$) among $v_1, \dots, v_n$.

If E is not linearly dependent (over $\mathbb{K}$) then we say that E is *linearly independent (over $\mathbb{K}$)*.

Remark 1.3.2. There are some crucial remarks to make:

1. We have defined linear (in)dependence for an arbitrary nonempty subset E of a $\mathbb{K}$-vector space V. In particular, E may be infinite (for example, we could take $E = V$![18]). However, for a subset to be linearly dependent we need only find a linear relation among finitely many vectors in E. Hence, if there is a linear relation (over $\mathbb{K}$) of the form (1.3.1) for some vectors $v_1, \dots, v_n \in V$ and some scalars $\lambda_1, \dots, \lambda_n$ (at least one of which is nonzero), then for *any* subset $S \subset V$ such that $\{v_1, \dots, v_n\} \subset S$, we must have that S is linearly dependent.

2. We will make more precise the notion of linear independence: suppose that $E \subset V$ is a linearly independent set. What does this mean? Definition 1.3.1 defines a subset of V to be linearly independent if it is not linearly dependent. Therefore, a subset E is linearly independent is equivalent to saying that there cannot exist a nontrivial linear relation (over $\mathbb{K}$) among any (finite) subset of vectors in E.

So, in order to show that a subset is linearly independent we need to show that no nontrivial linear relations (over $\mathbb{K}$) can exist among vectors in E. This is equivalent to showing that the only linear relations that exist among vectors in E must necessarily be trivial:

> Suppose that we can write 0_V as a linear combination
> $$\lambda_1 v_1 + \dots + \lambda_n v_n = 0_V,$$
> for some $v_1, \dots, v_n \in E$ and scalars $\lambda_1, \dots, \lambda_n \in \mathbb{K}$. Then, $\lambda_1 = \cdots = \lambda_n = 0$.

[17] Recall our conventions for notation after Remark 1.2.2.

[18] V is always linearly dependent. Why?

Thus, in order to show a given subset E of a vector space E is linearly independent (this could be asked as a homework question, for example) you must show that the above statement is true. This usually requires some thought and ingenuity on your behalf. However, once we have the notion of coordinates (with respect to a basis) we can turn this problem into one involving row-reduction (yay!).

However, to show that a subset $E \subset V$ is linearly dependent you need to find explicit vectors $v_1, \dots, v_n \in E$ and explicit scalars $\lambda_1, \dots, \lambda_n$ (not all of which are zero) so that there is a nontrivial linear relation

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0_V.$$

This can sometimes be quite difficult! However, once we have the notion of coordinates then we need only try to determine solutions to a matrix equation.

3. We have defined the notion of linear (in)dependence (over $\mathbb{K}$). We will usually omit the phrase 'over $\mathbb{K}$' as it will be assumed implicit that we are seeking linear relations over $\mathbb{K}$ when we are considering subsets of $\mathbb{K}$-vector spaces.

Proposition 1.3.3. *Let V be a $\mathbb{K}$-vector space and $E \subset V$ some nonempty subset. Then, if $0_V \in E$ then E is linearly dependent.*

Proof: We must show that there exists a nontrivial linear relation among some collection of vectors $v_1, \dots, v_n \in E$. We know that $0_V \in V$ and that there is the (obvious?) linear relation

$$1 \cdot 0_V = 0_V,$$

where we have used Proposition 1.2.5. Since we have found a nontrivial linear relation we conclude that E must be linearly dependent. □

Lemma 1.3.4. *Let V be a $\mathbb{K}$-vector space and $E \subset V$ some subset. Then, E is linearly dependent if and only if there exists a vector $v \in E$ that can be written as a linear combination of some of the others.*

Proof: ($\Rightarrow$) Suppose that E is linearly dependent. Then, there exists $v_1, \dots, v_n \in E$ and a nontrivial linear relation

$$\lambda_1 v_1 + \dots + \lambda v_n = 0_V.$$

We may assume, without loss of generality, that $\lambda_1 \neq 0$ and λ_1^{-1} therefore exists. Then, let $v = v_1$ so that we have

$$v = -\lambda_1^{-1} (\lambda_2 v_2 + \dots + \lambda_n v_n).$$

Hence, $v = v_1$ is a linear combination of some of the other vectors in E.

The converse is left to the reader. □

Corollary 1.3.5. *Let V be a $\mathbb{K}$-vector space, $E \subset V$ a nonempty subset. If E is linearly independent and $v \notin \operatorname{span}_{\mathbb{K}} E$ then $E \cup \{v\}$ is linearly independent.*

Proof: This follows from Lemma 1.3.4: if $E' = E \cup \{v\}$ were linearly dependent then there would exist some $u \in E'$ such that u can be written as a linear combination of other vectors in E'. WE can't have $u = v$, since $v \notin \operatorname{span}_{\mathbb{K}} E$. Hence, $u \in E$ so that it is possible to write u as a linear combination of vectors in E. In this case, E would be linearly dependent by Lemma 1.3.4 which is absurd, since E is assumed linearly independent. Hence, it is not possible for E' to be linearly dependent so it must be linearly independent. □

Question. Why did we care about finding $\lambda_1 \neq 0$? Why did we not just take the nontrivial relation appearing in the proof of Lemma 1.3.4 and move everything to one side except $\lambda_1 v_1$?

Example 1.3.6. Most of the following examples will only concern the linear (in)dependence of finite subsets E. However, I will include a couple of examples where E is infinite to highlight different methods of proof:

1. Consider the $\mathbb{R}$-vector space $\mathbb{R}^3$ and the subset

$$E = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}.$$

How do we determine linear (in)dependence of E? We must consider the vector equation

$$\lambda_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} = 0_{\mathbb{R}^3}.$$

Then, if there exists a particular $\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \neq 0_{\mathbb{R}^3}$ satisfying this vector equation then E is linearly dependent as we have found a nontrivial linear relation among the vectors in E. Otherwise, E must be linearly independent.

So, determining the linear (in)dependence of $E \subset \mathbb{R}^3$ boils down to **solving the homogeneous matrix equation**

$$A\underline{\lambda} = 0_{\mathbb{R}^3}, \quad \underline{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix},$$

where A is the 3×3 matrix whose columns are the vectors in E. Thus, we must row-reduce A and determine whether there exists a free variable or not: in the language of Math 54, we must determine if there exists a column of A that is not a pivot column.

2. The previous example generalises to any finite subset $E \subset \mathbb{K}^m$, for any $n \in \mathbb{N}$. Let $E = \{v_1, \dots, v_n\} \subset \mathbb{K}^m$ be a subset. Then, determining the linear (in)dependence of E is the same as solving the homogeneous matrix equation

$$A\underline{\lambda} = 0_{\mathbb{K}^m}, \quad \underline{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix},$$

where $A = [v_1 \ v_2 \ \cdots \ v_n]$ is the $m \times n$ matrix whose columns are the vectors in E.

If the only solution to this matrix equation is the zero solution (ie, the only solution is $\underline{\lambda} = 0_{\mathbb{K}^n}$) then E is linearly independent. Otherwise, E is linearly dependent and any nonzero solution you find will determine a nontrivial linear relation among $v_1, v_2, \dots, v_n$.

In general, we want to try and turn a linear (in)dependence problem into one that takes the preceeding form as then we need only row-reduce a matrix and determine pivots.

3. This example is quite subtle and leads to number theoretic considerations: consider the $\mathbb{Q}$-vector space $\mathbb{R}$. Then, the subset $E = \{1, \sqrt{2}\} \subset \mathbb{R}$ is linearly independent (over $\mathbb{Q}$!).

Indeed, consider a linear relation (over $\mathbb{Q}$)

$$a_1.1 + a_2.\sqrt{2} = 0 \in \mathbb{R}, \quad \text{where } a_1, a_2 \in \mathbb{Q}.$$

Assume that E is linearly dependent; we aim to provide a contradiction. Suppose that one of a_1 or a_2 is nonzero (in fact, we must have both of a_1 and a_2 are nonzero. Why?) Then, we have

$$\sqrt{2} = -\frac{a_1}{a_2} \in \mathbb{Q}.$$

However, the Greeks discovered the (heretical[19]) fact that $\sqrt{2}$ is irrational, therefore we can't possibly have that $\sqrt{2} \in \mathbb{Q}$. As such, our intial assumption that E is linearly dependent must be false, so that E is linearly independent (over $\mathbb{Q}$).

If we consider $\mathbb{R}$ as a $\mathbb{R}$-vector space then E is no longer linearly independent: we have

$$-\sqrt{2}.1 + 1.\sqrt{2} = 0 \in \mathbb{R},$$

[19]It is believed that the Pythagorean school in ancient Greece kept the irrationality of $\sqrt{2}$ a secret from the public and that Hippasus was murdered for revealing the secret!

is a nontrivial linear relation (over $\mathbb{R}$) among $1, \sqrt{2}$.

This example highlights the fact that it is important to understand which scalars you are allowed to use in a vector space as properties (for example, linear (in)dependence) can differ when we change scalars.

4. Consider the $\mathbb{R}$-vector space $\mathbb{R}_3[t]$ given in Example 1.2.6. Then, the subset

$$E = \{1, t, t^2, t^3\} \subset \mathbb{R}_3[t],$$

is linearly independent.

We must show that the boxed statement in Remark 1.3.2 holds. So, assume that we have a linear relation

$$\lambda_1.1 + \lambda_2.t + \lambda_3.t^2 + \lambda_4.t^3 = 0_{\mathbb{R}_3[t]},$$

with $\lambda_1, \ldots, \lambda_4 \in \mathbb{R}$. Then, by definition, the zero polynomial $0_{\mathbb{R}_4[t]}$ is the polynomial that has all coefficients equal to zero. Therefore, from our remarks in Example 1.2.6 we must have $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$ (polynomials are equal if and only if they have equal coefficients).

5. This example might appear to be the same as the previous example but it is actually different: consider $C_{\mathbb{R}}(0,1)$, the $\mathbb{R}$-vector space of continuous functions $f : (0,1) \to \mathbb{R}$. Let $E = \{f_0, f_1, f_2, f_3\}$, where

$$f_i : (0,1) \to \mathbb{R} \; ; \; x \mapsto x^i.$$

Then, E is linearly independent.

Indeed, suppose that we have a linear relation

$$\lambda_0 f_0 + \ldots + \lambda_3 f_3 = 0_{C_{\mathbb{R}}(0,1)}, \quad \lambda_1, \ldots, \lambda_3 \in \mathbb{R}.$$

Now, this is a linear relation between functions $(0,1) \to \mathbb{R}$, and any two such functions f, g are equal if and only if we have $f(x) = g(x)$, for every $x \in (0,1)$. Hence, we are supposing that

$$\lambda_0 f_0(x) + \lambda_1 f_1(x) + \lambda_2 f_2(x) + \lambda_3 f_3(x) = 0_{C_{\mathbb{R}}(0,1)}(x) = 0, \quad \text{for every } x \in (0,1),$$

$$\implies \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 = 0, \quad \text{for every } x \in (0,1).$$

There are now several ways to proceed: we can either use some calculus or a fundamental fact from algebra. Using calculus, we can differentiate this equation with respect to x repeatedly to obtain that $\lambda_3 = \lambda_2 = \lambda_1 = \lambda_0 = 0$. Alternatively, we can use the following basic fact from algebra: if we assume that one of the λ_i's is nonzero then the polynomial on the LHS of the above equation (considered as a function of x, not a formal expression) can have *at most* three distinct roots. However, since $(0,1)$ is infinite we can choose four distinct roots (for example, $x = 0.1, 0.2, 0.3, 0.4$ are roots), which is absurd. Hence, our assumption that one of the λ_i is nonzero is false, so that $\lambda_0 = \ldots = \lambda_3 = 0$ and E is linearly independent.

There is also a linear algebra approach to this problem that will appear on a worksheet.

6. Examples 4 and 5 can be generalised to show that, if $I \subset \mathbb{Z}_{\geq 0}$ is some set of non-negative integers, then the subsets

$$E_1 = \{t^i \mid i \in I\} \subset \mathbb{K}[t], \quad E_2 = \{f_i(x) = x^i \mid i \in I\},$$

are linearly independent.

We now introduce the second fundamental notion concerning vector spaces, that of the linear span of a subset (in [1] this is called the *linear manifold* defined by a subset).

Definition 1.3.7. Let V be a $\mathbb{K}$-vector space, $E \subset V$ some nonempty subset. Then, the $\mathbb{K}$-*linear span of* E is the set of all possible linear combinations of vectors in E,

$$\operatorname{span}_{\mathbb{K}} E = \{\lambda_1 v_1 + \ldots + \lambda_n v_n \mid v_1, \ldots, v_n \in E, \; \lambda_1, \ldots, \lambda_n \in \mathbb{K}\}.$$

If $E \subset V$ is a subset such that $\operatorname{span}_{\mathbb{K}} E = V$, then we say that E *spans* V, or that E *is a spanning set of* V.

Proposition 1.3.8. *Let V be a $\mathbb{K}$-vector space, $E \subset V$ some nonempty subset. Then,* $\text{span}_{\mathbb{K}} E$ *is a vector subspace of V.*

Proof: We will show that $\text{span}_{\mathbb{K}} E$ satsfies Axiom SUB from Definition 1.2.8: let

$$v = \lambda_1 v_1 + ... + \lambda_n v_n,\ u = \mu_1 u_1 + ... + \mu_p u_p \in \text{span}_{\mathbb{K}} E,$$

and $\alpha, \beta \in \mathbb{K}$. Then,

$$\begin{aligned} \alpha u + \beta v &= \alpha(\mu_1 u_1 + ... + \mu_p u_p) + \beta(\lambda_1 v_1 + ... \lambda_n v_n) \\ &= \alpha\mu_1 u_1 + ... + \alpha\mu_p u_p + \beta\lambda_1 v_1 + ... + \beta\lambda_n v_n, \end{aligned}$$

is a linear combination of elements of E. Hence, by the definition of $\text{span}_{\mathbb{K}} E$, $\alpha u + \beta v \in \text{span}_{\mathbb{K}} E$. □

Conversely, we have that every subspace $U \subset V$ is the span of some subset: namely, $\text{span}_{\mathbb{K}} U = U$.

Lemma 1.3.9. *Let V be a $\mathbb{K}$-vector space and $E_1 \subset E_2 \subset V$ nonempty subsets of V. Then,*

$$\text{span}_{\mathbb{K}} E_1 \subset \text{span}_{\mathbb{K}} E_2,$$

and $\text{span}_{\mathbb{K}} E_1$ *is a subspace of* $\text{span}_{\mathbb{K}} E_2$.

Proof: Left to the reader. □

Lemma 1.3.10 (Elimination Lemma)**.** *Let V be a $\mathbb{K}$-vector space and $E \subset V$ some nonempty subset. Suppose that E is linearly dependent. Then, there exists a vector $v \in E$ such that, if $E' = E \setminus \{v\}$[20], then*

$$\text{span}_{\mathbb{K}} E = \text{span}_{\mathbb{K}} E'.$$

Hence, we can remove a vector from E without changing the subspace spanned by E.

Proof: Since E is linearly dependent then, by Lemma 1.3.4, there exists a vector $v \in E$ such that v is a linear combination of some other vectors in E, that is

$$v = \lambda_1 v_1 + ... + \lambda_n v_n,$$

with $v_1, ..., v_n \in E$ and $\lambda_1, ..., \lambda_n \in \mathbb{K}$. Moreover, we can assume that $v \neq v_j$, for each j; this is the same as saying that $v \in \text{span}_{\mathbb{K}} E'$. We will show that this v satisfies the conditions of the Lemma.

Now, as $E' \subset E$ then we can use the previous Lemma to conclude that

$$\text{span}_{\mathbb{K}} E' \subset \text{span}_K E.$$

If we can now show that $\text{span}_{\mathbb{K}} E \subset \text{span}_{\mathbb{K}} E'$ then we must have equality

$$\text{span}_{\mathbb{K}} E' = \text{span}_{\mathbb{K}} E.$$

So, let $u \in \text{span}_{\mathbb{K}} E$. Therefore, by the definition of $\text{span}_{\mathbb{K}} E$, we have

$$u = \mu_1 u_1 + ... \mu_k u_k,$$

with $u_1, ..., u_k \in E$ and we can assume that $u_i \neq u_j$ for $i \neq j$. If there is some u_i such that $v = u_i$, then

$$u = \mu_1 u_1 + ... + \mu_{i-1} u_{i-1} + \mu_i v + \mu_{i+1} u_{i+1} + ... + \mu_k u_k.$$

Hence, we have $u_1, ..., u_{i-1}, u_{i+1}, ..., u_k \in E' \subset \text{span}_{\mathbb{K}} E'$ and $v \in \text{span}_{\mathbb{K}} E'$, so that by Proposition 1.3.8, we must have $u \in \text{span}_{\mathbb{K}} E'$.

Now, if $v \neq u_i$, for each i, then each $u_i \in E'$ so that $u \in \text{span}_{\mathbb{K}} E'$. In either case, we have shown that $u \in \text{span}_{\mathbb{K}} E'$ and, since u was arbitrary, we must have $\text{span}_{\mathbb{K}} E \subset \text{span}_{\mathbb{K}} E'$ and the result follows. □

Remark. Lemma 1.3.10 has the following consequence: if E is a finite linearly dependent set that spans the $\mathbb{K}$-vector space V , then there is a subset of E that forms a basis of V. You should already be aware of what a basis is; however, for completeness, we will (re)introduce this notion in an upcoming section in a (perhaps) not so familiar way that fits better with the intuition behind a basis.

[20] If $S \subset T$ are sets, then define

$$T \setminus S = \{t \in T \mid t \notin S\},$$

the collection of all elements of T that are not elements of S.

1.4 Linear Morphisms, Part I

We have given an introduction to vector spaces and we have introduced the fundamental ideas of linear (in)dependence and spans. In this section we will consider the relationships that can exist between distinct vector spaces and which respect the 'linear algebraic' structure of vector spaces: this is thel notion of a *linear morphism* between vector spaces.

Definition 1.4.1. Let V and W be $\mathbb{K}$-vector spaces.

• A function

$$f : V \to W \ ; \ v \mapsto f(v),$$

is called a $\mathbb{K}$-*linear morphism between* V *and* W if the following properties hold:

(LIN1) for every $u, v \in V$, we have

$$f(u+v) = f(u) + f(v);$$

where the '$+$' on the LHS of this equation is addition in V and the '$+$' on the RHS of this equation is addition in W,

(LIN2) for every $u \in V$, $\lambda \in \mathbb{K}$, we have

$$f(\lambda v) = \lambda f(v);$$

where the scalar multiplication on the LHS of this equation is occuring in V and on the RHS of this equation it is occuring in W.

In fact, we can subsume both of these properties into

(LIN) for every $u, v \in V$, $\lambda \in \mathbb{K}$, we have

$$f(u + \lambda v) = f(u) + \lambda f(v),$$

where the scalar multiplication on the LHS of this equation is occuring in V and on the RHS of this equation it is occuring in W.

• For given $\mathbb{K}$-vector spaces V and W we denote the set of all $\mathbb{K}$-linear morphisms by

$$\mathrm{Hom}_{\mathbb{K}}(V, W) = \{f : V \to W \mid \ f \text{ linear}\}.$$

• The set of all $\mathbb{K}$-linear morphisms from a $\mathbb{K}$-vector space V to itself is denoted

$$\mathrm{End}_{\mathbb{K}}(V) \stackrel{def}{=} \mathrm{Hom}_{\mathbb{K}}(V, V).$$

A vector $f \in \mathrm{End}_{\mathbb{K}}(V)$ is called an *endomorphism* of V. For every $\mathbb{K}$-vector space V there exists the *identity morphism of* V, denoted $\mathrm{id}_V \in \mathrm{End}_{\mathbb{K}}(V)$. See the upcoming examples (Example 1.4.8,).

• We will use the adjectives 'injective', 'surjective' and 'bijective' to describe linear morphisms that satisfy the corresponding conditions.

• A bijective linear morphism will be called an *isomorphism*.

The set of all bijective $\mathbb{K}$-linear morphisms from a $\mathbb{K}$-vector space V to itself is denoted

$$\mathrm{GL}_{\mathbb{K}}(V) = \{f \in \mathrm{End}_{\mathbb{K}}(V) \mid \ f \text{ is bijective}\}.$$

We will see that, in the world of linear algebra, $\mathbb{K}$-vector spaces that are isomorphic have the same linear algebraic properties (and, therefore, can be regarded as 'the same').

Notation. You may have seen the phrases '**linear map**', '**linear transformation**' or '**linear function**': these all mean the same thing, namely, a function satisfying (LIN) above. We are using the word 'morphism' to emphasise the fact that a linear morphism is a function that 'changes' one vector space to another. This is also the fancy grown-up word that certain mathematicians use (myself included) in daily parlance.

Remark 1.4.2. We will see in a later section (Theorem 1.7.4) that, for $f \in \mathrm{End}_{\mathbb{K}}(V)$, with V a *finite dimensional* $\mathbb{K}$-vector space

$$\text{'}f \text{ injective'} \implies \text{'}f \text{ surjective'} \implies \text{'}f \text{ bijective'} \implies \text{'}f \text{ injective'},$$

so that all of these notions are equivalent for finite-dimensional $\mathbb{K}$-vector spaces.

Lemma 1.4.3. *Let $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ be a $\mathbb{K}$-linear morphism between the $\mathbb{K}$-vector spaces V and W. Then, $f(0_V) = 0_W$.*

Proof: We have

$$f(0_V) = f(0_V + 0_V) = f(0_V) + f(0_V), \quad \text{by LIN1},$$

and subtracting $f(0_V)$ from both sides of this equation we obtain

$$0_W = f(0_V).$$

□

Definition 1.4.4. Let V, W be $\mathbb{K}$-vector spaces and $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$. Then,

- the *kernel of f* is the subset

$$\ker f = \{v \in V \mid f(v) = 0_W\} \subset V,$$

- the *image of f* is the subset

$$\mathrm{im} f = \{w \in W \mid w = f(v), \text{ for some } v \in V\} \subset W.$$

Proposition 1.4.5. *Let $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$, for $\mathbb{K}$-vector spaces V, W. Then,*

- $\ker f$ *is a subspace of V,*
- $\mathrm{im} f$ *is a subspace of W.*

Proof: Left to the reader. □

Definition 1.4.6. Let V, W be $\mathbb{K}$-vector spaces. Then, we will define the structure of a $\mathbb{K}$-vector space on the set $\mathrm{Hom}_{\mathbb{K}}(V, W)$: define the $\mathbb{K}$-vector space $(\mathrm{Hom}_{\mathbb{K}}(V, W), \alpha, \sigma)$ where

$$\alpha : \mathrm{Hom}_{\mathbb{K}}(V, W) \times \mathrm{Hom}_{\mathbb{K}}(V, W) \to \mathrm{Hom}_{\mathbb{K}}(V, W) \; ; \; (f, g) \mapsto (\alpha(f, g) : V \to W \; ; \; v \mapsto f(v) + g(v)),$$

$$\sigma : \mathbb{K} \times \mathrm{Hom}_{\mathbb{K}}(V, W) \to \mathrm{Hom}_{\mathbb{K}}(V, W) \; ; \; (\lambda, f) \mapsto (\sigma(\lambda, f) : V \to W \; ; \; v \mapsto \lambda f(v)).$$

As usual we will write

$$\alpha(f, g) = f + g, \quad \text{and} \quad \sigma(\lambda, f) = \lambda f.$$

Whenever we discuss $\mathrm{Hom}_{\mathbb{K}}(V, W)$ as a $\mathbb{K}$-vector space, it will be this $\mathbb{K}$-vector space structure that we mean.

There are a couple of things that need to be checked to ensure that the above defintion of $\mathbb{K}$-vector space on $\mathrm{Hom}_{\mathbb{K}}(V, W)$ makes sense:

1. You need to check that the new functions $f + g$ and λf that we have defined, for $f, g \in \mathrm{Hom}_{\mathbb{K}}(V, W), \lambda \in \mathbb{K}$, are actually elements in $\mathrm{Hom}_{\mathbb{K}}(V, W)$, that is, that they are $\mathbb{K}$-linear morphisms.
2. The zero vector $0_{\mathrm{Hom}_{\mathbb{K}}(V,W)} \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ is the $\mathbb{K}$-linear morphism

$$0_{\mathrm{Hom}_{\mathbb{K}}(V,W)} : V \to W \; ; \; v \mapsto 0_W.$$

3. Given $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ we define the negative of f to be the $\mathbb{K}$-linear morphism

$$-f : V \to W \; ; \; v \mapsto -f(v),$$

where $-f(v)$ is the negative (in W) of the vector $f(v)$, for each $v \in V$.

Remark 1.4.7. 1. The fact that $\mathrm{Hom}_{\mathbb{K}}(V, W)$ has the structure of a $\mathbb{K}$-vector space will be important when we come to consider the Jordan canonical form. In that case, we will be considering the $\mathbb{K}$-vector space $\mathrm{End}_{\mathbb{K}}(V)$ and using some of its basic linear algebraic structure to deduce important properties of $\mathbb{K}$-linear morphisms $f : V \to V$.

2. We can consider $\mathrm{Hom}_{\mathbb{K}}(V, W) \subset W^V$ as a subset of the $\mathbb{K}$-vector space of (arbitrary) functions (recall Example 1.2.6)

$$W^V = \{f : V \to W\}.$$

In fact, $\mathrm{Hom}_{\mathbb{K}}(V, W) \subset W^V$ is a vector subspace.

However, the condition of $\mathbb{K}$-linearity that we have imposed on the functions is very strong and there are far 'fewer' $\mathbb{K}$-linear functions than there are arbitrary functions. For example, we will see in a proceeding section that $\mathrm{Hom}_{\mathbb{K}}(V, W)$ is finite-dimensional, whereas W^V is infinite-dimensional (assuming $W \neq Z$, the trivial vector space introduced in Example 1.2.6, 6[21]).

3. It is **not** true that $\mathrm{GL}_{\mathbb{K}}(V)$ is a vector subspace of $\mathrm{End}_{\mathbb{K}}(V)$, for any $\mathbb{K}$-vector space V that is not the trivial $\mathbb{K}$-vector space $\underline{Z}$ with one element (cf. Example 1.2.6). For example, the zero vector $0_{\mathrm{End}_{\mathbb{K}}(V)} \notin \mathrm{GL}_{\mathbb{K}}(V)$ since $0_{\mathrm{End}_{\mathbb{K}}(V)} : V \to V$ is not an injective function: if $v \in V$ is nonzero in V then

$$0_{\mathrm{End}_{\mathbb{K}}(V)}(v) = 0_V = 0_{\mathrm{End}_{\mathbb{K}}(V)}(0_V),$$

where we have used Lemma 1.4.3 for the RHS equality.

We will now give some basic examples of $\mathbb{K}$-linear morphisms. Most of these should be familiar from your first linear algebra class and, as such, you should feel pretty at ease with showing that the given functions are linear.

Example 1.4.8. 1. Consider the function

$$f : \mathbb{Q}^4 \to \mathbb{Q}^2 \; ; \; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} x_1 - x_4 \\ x_3 + \frac{2}{7}x_1 \end{bmatrix}.$$

Then, f is $\mathbb{Q}$-linear.

2. The function

$$f : \mathbb{R}^2 \mapsto \mathbb{R}^3 \; ; \; \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} x_1^3 + 2x_2^2 \\ -x_1 + \sqrt{2}x_2 \\ x_1 \end{bmatrix},$$

is not $\mathbb{R}$-linear. For example, if it were, then we must have (recall the definition of e_i from Example 1 1.2.6)

$$f(-e_2) = \begin{bmatrix} 2 \\ -\sqrt{2} \\ 0 \end{bmatrix},$$

whereas

$$-f(e_2) = -\begin{bmatrix} 2 \\ \sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -\sqrt{2} \\ 0 \end{bmatrix}.$$

Hence,

$$f(-e_2) \neq -f(e_2),$$

[21] What happens when $W = Z$?

so that Axiom LIN2 does not hold.

The problem we have here is the appearance of 'nonlinear' terms x_2^2 etc. In general, we must only have single powers of x_i appearing as in Example 1.

3. In general, a function

$$f : \mathbb{K}^n \to \mathbb{K}^m \; ; \; \underline{x} \mapsto f(\underline{x}),$$

is a $\mathbb{K}$-linear morphism if and only there exists an $m \times n$ matrix $A \in Mat_{m\times n}(\mathbb{K})$ with entries in $\mathbb{K}$ such that

$$f(\underline{x}) = A\underline{x}, \quad \text{for every } \underline{x} \in \mathbb{K}^n.$$

You should have already seen this result from your first linear algebra class.

Conversely, given $A \in Mat_{m,n}(\mathbb{K})$ we define the $\mathbb{K}$-linear morphism

$$T_A : \mathbb{K}^n \to \mathbb{K}^m \; ; \; \underline{x} \mapsto A\underline{x}.$$

This notation will reappear through these notes.

4. Let V be a $\mathbb{K}$-vector space. Then, the *identity morphism of* V is the $\mathbb{K}$-linear morphism

$$\mathrm{id}_V : V \to V \; ; \; v \mapsto v.$$

It is easy to see that $\mathrm{id}_V \in GL_{\mathbb{K}}(V)$, ie, that id_V is an isomorphism.

5. Let V be a $\mathbb{K}$-vector space and $U \subset V$ be a vector subspace. Then, there is a $\mathbb{K}$-linear morphism

$$i_U : U \to V \; ; \; u \mapsto u,$$

called the *inclusion morphism of* U. It is trivial to verify that this is $\mathbb{K}$-linear. Moreover, i_U is an *injective* morphism, for any subspace $U \subset V$.

6. Let V be a $\mathbb{K}$-vector space and suppose that there are subspaces $U, W \subset V$ such that $V = U \oplus W$. Then, define the *projection morphisms onto* U *and* W as follows:

$$p_U : V \to U \; ; \; v = u + w \mapsto u,$$

$$p_W : V \to W \; ; \; v = u + w \mapsto w.$$

These morphisms are *surjective*.

Note that these functions are well-defined because $V = U \oplus W$ and so every $v \in V$ can be uniquely written as $v = u + w$ (by Proposition 1.2.11). Therefore, we need not worry about whether p_U, p_W are functions.[22]

7. The following are examples from calculus: consider the $\mathbb{R}$-vector space $C_{\mathbb{R}}[0,1]$ of continuous functions $f : [0,1] \to \mathbb{R}$. Then, the function

$$\int_0^1 : C_{\mathbb{R}}[0,1] \to \mathbb{R} \; ; \; f \mapsto \int_0^1 f(x)dx,$$

is $\mathbb{R}$-linear. This should be well-known to all.

If we denote by $C^1(0,1) \subset C_{\mathbb{R}}(0,1)$ the set of all continuous functions $f : (0,1) \to \mathbb{R}$ that are differentiable, then we have an $\mathbb{R}$-linear map[23]

$$\frac{d}{dx} : C^1(0,1) \to C_{\mathbb{R}}(0,1) \; ; \; f \mapsto \frac{df}{dx},$$

[22] If we did not have the uniqueness property, and only knew that $V = U = W$, then it could be possible that $v = u + w = u' + w$ with $u \neq u' \in U$. Then, $p_U(v)$ could equal either u or u', so that p_U can't be a function (recall that a function $f : S \to W$ must assign a unique value $f(s)$, to every $s \in S$).

[23] It is not necessarily true that a function that can be differentiated once can be differentiated *twice*. It is actually surprisingly hard to find such an example but if you take Math 104 you should see the following example of such a function

$$f : \mathbb{R} \to \mathbb{R} \; ; \; x \mapsto \begin{cases} x^2 \sin(x^{-1}), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

which is just the '*derivative with respect to x*' morphism. It is $\mathbb{R}$-linear.

8. This example exhibits a subtlety that we shall come back to in later sections: recall the set of natural numbers $\mathbb{N}$. Define a function

$$T : \mathbb{K}^{\mathbb{N}} \to \mathbb{K}^{\mathbb{N}} \; ; \; (i \mapsto f(i)) \mapsto \left(i \mapsto \begin{cases} 0, & \text{if } i = 1, \\ f(i-1), & \text{if } i \neq 1. \end{cases} \right).$$

That is, if we represent a function $(f : \mathbb{N} \to \mathbb{K} \; ; \; i \mapsto f(i)) \in \mathbb{K}^{\mathbb{N}}$ by an infinite sequence

$$(f_i) = (f_1, f_2, f_3, ...),$$

where $f_i \overset{def}{=} f(i)$, then

$$T((f_i)) = (0, f_1, f_2, f_3, ...).$$

So, T is the '*shift to the right by one place*' function defined on infinite sequences of numbers in $\mathbb{K}$.

Then, it is relatively straightforward to see that T is $\mathbb{K}$-linear and is injective. However, T is **not surjective**: thus, we have an example of an injective linear endomorphism of a $\mathbb{K}$-vector space that is not surjective. As we will see in an upcoming section, this is impossible if $\mathbb{K}^{\mathbb{N}}$ were finite-dimensional (cf. Theorem 1.7.4). Hence, this implies that $\mathbb{K}^{\mathbb{N}}$ is an infinite dimensional $\mathbb{K}$-vector space.

We now recall an important result that allows us to characterise when $\mathbb{K}$-linear morphisms are injective. In practice, whenever you want to show that a morphism is injective you should use the following

Lemma 1.4.9 (Characterising injective linear morphisms)**.** *Let V, W be $\mathbb{K}$-vector spaces, $f : V \to W$ a $\mathbb{K}$-linear morphism. Then, f is injective if and only if* $\ker f = \{0_V\}$.

Proof: ($\Rightarrow$) Suppose that f is injective. Let $v \in \ker f$; we want to show that $v = 0_V$. Now, since $v \in \ker f$, then $f(v) = 0_W$, by the definition of $\ker f$. Furthermore, by Lemma 1.4.3, we know that $f(0_V) = 0_W$. Hence, as f is injective then

$$f(v) = f(0_V) \implies v = 0_V,$$

so that $\ker f = \{0_V\}$.

($\Leftarrow$) Conversely, suppose that $\ker f = \{0_V\}$. We must show that f is injective: therefore, we need to show that, whenever $f(v) = f(w)$, for some $v, w \in V$, then we necessarily have $v = w$. So suppose that there are $v, w \in V$ with $f(v) = f(w)$. Then

$$f(v) = f(w) \implies f(v) - f(w) = 0_W \implies f(v - w) = 0_W, \text{ since } f \text{ linear},$$

so that $v - w \in \ker f = \{0_V\}$. Hence, $v = w$. Therefore, f must be an injective function. □

Remark 1.4.10. In this section, we have given a (re)introduction to linear morphisms (or linear maps, transformations, whatever) and stated some basic properties and examples. However, in practice it is usually pretty difficult to prove certain things about linear morphisms (for example, injectivity, surjectivity etc.) in a direct manner.

In order to make questions easier to understand and solve we will most often *represent* a linear morphism using a *matrix representation*. This will be done in the proceeding sections. However, it should be noted that this approach to attacking problems only works for finite-dimensional vector spaces and the morphisms between them (infinite matrices are difficult to manipulate!).

We finish this section with some important facts that we will use throughout the remainder of these notes.

Theorem 1.4.11 (Invariance of Domain)**.** *Suppose that there exists an isomorphism*

$$f : \mathbb{K}^n \to \mathbb{K}^m.$$

Then, $n = m$.

Proof: This is an exercise in row-reduction and one which you should already be familiar with.

Recall that for any linear morphism $f : \mathbb{K}^n \to \mathbb{K}^m$, there is a matrix A_f called the *standard matrix associated to* f such that

$$\text{for every } \underline{x} \in \mathbb{K}^n,\ f(\underline{x}) = A_f \underline{x}.$$

A_f is defined to be the $m \times n$ matrix whose i^{th} column is the column vector $f(e_i)$, where e_i is the i^{th} standard basis vector of $\mathbb{K}^n$ (Example 1.2.6).
Then, it will be an exercise to show the following:

- f is injective if and only if A_f has a pivot in every column, and
- f is surjective if and only if A_f has a pivot in every row.

Therefore, since we are assuming that f is an isomorphism it must, by definition, be a bijective morphism. Hence, it is both injective and surjective. By the preceding comments we must therefore have a pivot in every column and every row. The only way that this can happen is if $n = m$. □

We will see later, after the introduction of bases for vector spaces, that the converse if also true: namely, **if $n = m$ then $\mathbb{K}^n$ and $\mathbb{K}^m$ are isomorphic.**

Proposition 1.4.12. *Let V, W be $\mathbb{K}$-vector spaces, $E \subset V$ a subset of V. Let $f : V \to W$ be an isomorphism from V to W and denote $f(E) = \{f(e) \mid e \in E\}$, the* image set of E.[24] *Then,*

- *E is linearly independent if and only if $f(E)$ is linearly independent.*
- *E spans V if and only if $f(E)$ spans W.*

Proof: Left to the reader. □

1.5 Bases, Dimension

In this section we will introduce the notion a *basis* of a $\mathbb{K}$-vector space. We will provide several equivalent approaches to the definition of a basis and see that the size of a basis is an invariant[25] of a $\mathbb{K}$-vector space which we will call its *dimension*. You should have already seen the words *basis* and *dimension* in your previous linear algebra course so do not abandon what you already know! We are just simply going to provide some interesting(?) ways we can think about a basis; in particular, these new formulations will allow us to extend our results to infinite dimensional vector spaces.

First, we must introduce a (somewhat annoying) idea to keep us on the straight and narrow when we are considering bases, that of an *ordered set*.

Definition 1.5.1 (Ordered Set)**.** An *ordered set* is a nonempty set S for which we have provided a 'predetermined ordering' on S.

Remark 1.5.2. 1. This definition might seem slightly confusing (and absurd); indeed, it is both of these things as I have not rigorously defined what a 'predetermined ordering' is. Please don't dwell too much on this as we will only concern ourselves with orderings of finite sets (for which it is easy to provide an ordering) or the standard ordering of $\mathbb{N}$. An ordered set is literally just a nonempty set S whose elements have been (strictly) ordered in some way.

For example, suppose that $S = [3] = \{1, 2, 3\}$. We usually think of S has having its natural ordering $(1, 2, 3)$. However, when we consider this ordering we are actually considering the ordered set $(1, 2, 3)$ and not the set S... Confused? I thought so. We could also give the objects in S the ordering $(2, 1, 3)$ and when we do this we have a defined a *different* ordered set to $(1, 2, 3)$.

[24] This is not necessarily the same as the *image of* f, $\mathrm{im} f$, introduced before.

[25] In mathematics, when we talk of an *invariant* we usually mean an atrribute or property of an object that remains unchanged whenever that object is transformed to another via an isomorphism (in an appropriate sense). For example, you may have heard of the *genus* of a (closed) geometric surface: this is an invariant of a surface that counts the number of 'holes' that exist within a (closed) surface. Perhaps you have heard or read the phrase that a mathematician thinks a coffee mug and a donut are indistuingishable. This is because we can continuously deform a donut into a coffee mug, and vice versa. This continuous deformation can be regarded as an 'isomorphism' in the world of (closed) geometric surfaces.

If you are still confused, do not worry. Here is another example: consider the set

$$S = \{\text{Evans Hall, Doe Library, Etcheverry Hall}\}.$$

Now, there is no predetermined way that we can order this set: I might choose the ordering

(Evans Hall, Etcheverry Hall, Doe Library),

whereas you might think it better to choose an ordering

(Doe Library, Etcheverry Hall, Evans Hall).

Of course, neither of these choices of orderings is 'right' and we are both entitled to our different choices. However, these ordered sets **are different**.

The reason we require this silly idea is when we come to consider coordinates (with respect to a given ordered basis). Then, it will be extremely important that we declare an ordering of a basis and that we are consistent with this choice.

2. Other examples of ordered sets include $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ with their usual orderings. We can also order $\mathbb{C}$ in an ordering called a *lexicographic ordering*: here we say that $z = a_1 + b_1\sqrt{-1} < w = a_2 + b_2\sqrt{-1}$ if and only if either, $a_1 < a_2$, or, $a_1 = a_2$ *and* $b_1 < b_2$. Think of this as being similar to the way that words are ordered in the dictionary, except now we consider only 'words' consisting of two 'letters', each of which is a real number.

3. What about some really bizarre set that might be infinite; for example, $\mathbb{R}^{\mathbb{R}}$, the set of all functions $\mathbb{R} \to \mathbb{R}$. How can we order this set? In short, I have no idea! However, there are some very deep results from mathematical logic that say that, if we assume a certain axiom of mathematics (the so-called Axiom of Choice), then every set can be ordered in some manner. In fact, it has been shown that the Axiom of Choice of logically equivalent to this ordering property of sets! If you want to learn more then you should consult Wikipedia and take Math 125A in the Fall Semester.[26]

Therefore, no matter how weird or massively infinite a set is, if you are assuming the Axiom of Choice (which we are) then you can put an ordering on that set, even though you will (a priori) have no idea what that ordering is! All that matters is that such an ordering **exists**.

Definition 1.5.3 (Basis; Ordered Basis)**.** Let V be a $\mathbb{K}$-vector space. A nonempty subset $\mathcal{B} \subset V$ is called a $(\mathbb{K})$-*basis of* V if

- $\mathcal{B}$ is linearly independent (over $\mathbb{K}$), and
- if $\mathcal{B} \subset \mathcal{B}'$ and $\mathcal{B}'$ is linearly independent (over $\mathbb{K}$), then $\mathcal{B}' = \mathcal{B}$.

In this case, we say that $\mathcal{B}$ is *maximal linearly independent*.

An *ordered* $(\mathbb{K})$-*basis of* V is a $(\mathbb{K})$-basis of V that is an ordered set.

Remark 1.5.4. 1. You may have seen a basis of a $\mathbb{K}$-vector space V defined as a subset $\mathcal{B} \subset V$ such that $\mathcal{B}$ is linearly independent (over $\mathbb{K}$) and such that $\text{span}_{\mathbb{K}} \mathcal{B} = V$. The definition given above is equivalent to this and it has been used as the definition of a basis to encapsulate the intuition behind a basis: namely, if $\mathbb{K} = \mathbb{R}$, we can think of a basis of an $\mathbb{R}$-vector space as a choice of 'independent directions' that allows us to consider well-defined coordinates. This idea of 'independent directions' is embodied in the fact that a basis must be a linearly independent set; and the assumption of maximal linear independence is what allows us to obtain well-defined coordinates.

However, just to keep our minds at ease our next result will show the equivalence between Definition 1.5.3 and the definition you have probably seen before.

[26] I have to admit that I do not know any mathematical logic but have come across these ideas during my own excursions in mathematics. There are lots of many interesting results that can be obtained if one assumes the Axiom of Choice: one is called the Banach-Tarski Paradox; another, which is directly related to our studies, is the existence of a basis for *any* $\mathbb{K}$-vector space. In fact, the Axiom of Choice is logically equivalent to the existence of a basis for any $\mathbb{K}$-vector space.

2. We will also see in the homework that we can consider a basis to be a *minimal spanning set* (in an appropriate sense to be defined later); this is recorded in Proposition 1.5.9.

3. It is important to remember that a basis is a subset of V and **not** a subspace of V.

4. We will usually not call a basis of a $\mathbb{K}$-vector space a '$\mathbb{K}$-basis', it being implicitly assumed that we are considering only $\mathbb{K}$-bases when we are talking about $\mathbb{K}$-vector spaces. As such, we will only use the terminology 'basis' from now on.

Proposition 1.5.5. *Let V be a $\mathbb{K}$-vector space and $\mathcal{B} \subset V$ a basis of V. Then, $\operatorname{span}_{\mathbb{K}} \mathcal{B} = V$. Conversely, if $\mathcal{B} \subset V$ is a linearly independent spanning set of V, then $\mathcal{B}$ is a basis of V*

Proof: Let us denote $W = \operatorname{span}_{\mathbb{K}} \mathcal{B}$. Then, because $\mathcal{B} \subset V$ we have $W \subset V$. To show that $W = V$ we are going to assume otherwise and obtain a contradiction. So, suppose that $W \neq V$. This means that there exists $v_0 \in V$ such that $v_0 \notin W$. In particular, $v_0 \notin \mathcal{B} \subset W$. Now, consider the subset $\mathcal{B}' = \mathcal{B} \cup \{v_0\} \subset V$.

Then, by Corollary 1.3.5, $\mathcal{B}'$ is linearly independent.

Now, we use the maximal linear independence property of $\mathcal{B}$: since $\mathcal{B} \subset \mathcal{B}'$ and $\mathcal{B}'$ is linearly independent we must have $\mathcal{B}' = \mathcal{B}$, because $\mathcal{B}$ is a basis. Hence, $v_0 \in \mathcal{B}$. But this contradicts that fact that $v_0 \notin \mathcal{B}$. Therefore, our intial assumption, that $W \neq V$, must be false and we must necessarily have $W = V$.

Conversely, suppose that $\mathcal{B}$ is a linearly independent subset of V such that $\operatorname{span}_{\mathbb{K}} \mathcal{B} = V$. We want to show that $\mathcal{B}$ is a basis, so we must show that $\mathcal{B}$ satisfies the maximal linearly independent property of Definition 1.5.3.

Therefore, suppose that $\mathcal{B} \subset \mathcal{B}'$ and that $\mathcal{B}'$ is linearly independent; we must show that $\mathcal{B}' = \mathcal{B}$. Now, since $\mathcal{B} \subset \mathcal{B}'$ we have $V = \operatorname{span}_{\mathbb{K}} \mathcal{B} \subset \operatorname{span}_{\mathbb{K}} \mathcal{B}' \subset V$, using Lemma 1.3.9. Hence, $\operatorname{span}_{\mathbb{K}} \mathcal{B}' = V = \operatorname{span}_{\mathbb{K}} \mathcal{B}$. Assume that $\mathcal{B} \neq \mathcal{B}'$; we aim to provide a contradiction. Then, for each $w \in \mathcal{B}' \setminus \mathcal{B}$ we have $w \in \operatorname{span}_{\mathbb{K}} \mathcal{B}' = \operatorname{span}_{\mathbb{K}} \mathcal{B}$, so that there exists an expression

$$w = \lambda_1 b_1 + \ldots + \lambda_n b_n,$$

where $b_1, \ldots, b_n \in \mathcal{B}$. But this means that we have a nontrivial[27] linear relation among vectors in $\mathcal{B}'$ (recall that, as $\mathcal{B} \subset \mathcal{B}'$, we have $b_1, \ldots, b_n \in \mathcal{B}'$). However, $\mathcal{B}'$ is linearly independent so that no such nontrivial linear relation can exist. Hence, our initial assumption of the existence of $w \in \mathcal{B}' \setminus \mathcal{B}$ is false, so that $\mathcal{B}' = \mathcal{B}$. The result follows. □

Corollary 1.5.6. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B} \subset V$ a basis of V. Then, for every $v \in V$ there exists a unique expression*

$$v = \lambda_1 b_1 + \ldots + \lambda_n b_n,$$

where $b_1, \ldots, b_n \in \mathcal{B}$, $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$, $n \in \mathbb{N}$.

Proof: By Proposition 1.5.5, we have that $\operatorname{span}_{\mathbb{K}} \mathcal{B} = V$ so that, for every $v \in V$, we can write v as a linear combination of vectors in $\mathcal{B}$

$$v = \lambda_1 b_1 + \ldots + \lambda_n b_n, \quad b_1, \ldots, b_n \in \mathcal{B},$$

where we can further assume that none of $\lambda_1, \ldots, \lambda_n$ is equal to zero.

We need to show that this expression is *unique*: so, suppose that we can write v as a different linear combination

$$v = \mu_1 b'_1 + \ldots + \mu_k b'_k, \quad b'_1, \ldots, b'_k \in \mathcal{B},$$

again assuming that none of the $\mu_1, \ldots, \mu_k$ are equal to zero.

Therefore, we have

$$\lambda_1 b_1 + \ldots + \lambda_n b_n = v = \mu_1 b'_1 + \ldots + \mu_k b'_k,$$

giving a linear relation

$$\lambda_1 b_1 + \ldots + \lambda_n b_n - (\mu_1 b'_1 + \ldots + \mu_k b'_k) = 0_V.$$

[27] Why is this linear relation nontrivial?

Thus, since $\mathcal{B}$ is linearly independent this linear relation must be trivial and, furthermore, since we have assumed that none of the λ's or μ's are zero, the only way that this can happen is if $n = k$ and, without loss of generality, $b_i = b'_i$ and $\lambda_i = \mu_i$. Hence, the linear combination given above is unique. □

Corollary 1.5.7. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = (b_1, \dots, b_n) \subset V$ an ordered basis containing finitely many vectors. Then, V is isomorphic to $\mathbb{K}^n$.*

Proof: This is just a simple restatement of Corollary 1.5.6: we define a function

$$[-]_{\mathcal{B}} : V \to \mathbb{K}^n \; ; \; v \mapsto [v]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix},$$

where

$$v = \lambda_1 b_1 + \dots + \lambda_n b_n,$$

is the unique expression for v coming from Corollary 1.5.6. Uniqueness shows that $[-]_{\mathcal{B}}$ is indeed a well-defined function.

It will be left to the reader to show that $[-]_{\mathcal{B}}$ is a bijective $\mathbb{K}$-linear morphism, thereby showing that it is an isomorphism. □

Definition 1.5.8. Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = \{b_1, \dots, b_n\} \subset V$ an ordered basis containing finitely many vectors. Then, the linear morphism

$$[-]_{\mathcal{B}} : V \to \mathbb{K}^n \; ; \; v \mapsto [v]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix},$$

introduced above is called the *$\mathcal{B}$-coordinate map* or *$\mathcal{B}$-coordinate morphism*.

The following Proposition provides yet another viewpoint of the idea of a basis: it says that a basis is a spanning set that satisfies a certain minimality condition.

Proposition 1.5.9. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B} \subset V$ a basis of V. Then, $\mathcal{B}$ is a minimal spanning set - namely,*

- $\text{span}_{\mathbb{K}} \mathcal{B} = V$, *and*
- *if $\mathcal{B}' \subset \mathcal{B}$ is such that $\text{span}_{\mathbb{K}} \mathcal{B}' = V$ then $\mathcal{B}' = \mathcal{B}$.*

A proof of this Proposition will appear as a homework exercise.

Despite all of these results on bases of vector spaces we have still yet to give the most important fact concerning a basis: that a basis **exists** in an arbitrary $\mathbb{K}$-vector space.

The proof of the general case requires the use of a particularly subtle lemma, called Zorn's Lemma. You can read about Zorn's Lemma on Wikipedia and there you will see that Zorn's Lemma is equivalent to the Axiom of Choice (although the proof of this fact is quite difficult). You will also read on Wikipedia that Zorn's Lemma is logically equivalent to the existence of a basis for an arbitrary $\mathbb{K}$-vector space.

Theorem 1.5.10. *Let V be a $\mathbb{K}$-vector space. Then, there exists a basis $\mathcal{B} \subset V$ of V.*

Proof: Case 1: There exists a finite subset $E \subset V$ such that $\text{span}_{\mathbb{K}} E = V$.

In this case we will use the Elimination Lemma (Lemma 1.3.10) to remove vectors from E until we obtain a linearly independent set. Now, if E is linearly independent then E is a linearly independent spanning set of V and so, by Proposition 1.5.5, E is a basis of V. Therefore, assume that E is linearly dependent. Then, if we write E as an ordered set $E = \{e_1, \dots, e_n\}$, we can use Lemma 1.3.10 to remove a vector from E so that the resulting set is also a spanning set of V; WLOG, we can assume that the vector we remove is e_n. Then, define $E^{(n-1)} = E \setminus \{e_n\}$ so that we have $\text{span}_{\mathbb{K}} E^{(n-1)} = V$. If $E^{(n-1)}$ is linearly independent then it must be a basis (as it is also a spanning set). If $E^{(n-1)}$ is linearly dependent then

we can again use Lemma 1.3.10 to remove a vector from $E^{(n-1)}$ so that the resulting set is a spanning set of V; WLOG, we can assume that the vector we remove is e_{n-1}. Then, define $E^{(n-2)} = E \setminus \{e_{n-2}\}$ so that we have $\text{span}_{\mathbb{K}} E^{(n-2)} = V$. Proceeding in a similar fashion as before we will either have that $E^{(n-2)}$ is linearly independent (in which case it is a basis) or it will be linearly dependent and we can proceed as before, removing a vector to obtain a new set $E^{(n-3)}$ etc.

Since E is a finite set this procedure must terminate after finitely many steps. The stage at which it terminates will have a produced a linearly independent spanning set of V, that is, a basis of V (by Proposition 1.5.5).

Case 2: There does not exist a finite spanning set of V.

In this case we must appeal to Zorn's Lemma: basically, the idea is that we will find a basis by considering a maximal linearly independent subset of V. Zorn's Lemma is a technical result that allows us to show that such a subset always exists and therefore, by definition, must be a basis of V. □

Theorem 1.5.11 (Basis Theorem)**.** *Let V be a $\mathbb{K}$-vector space and $\mathcal{B} \subset V$ a basis such that $\mathcal{B}$ has only finitely many vectors. Then, if $\mathcal{B}'$ is another basis of V then $\mathcal{B}'$ has the same number of vectors as $\mathcal{B}$.*

Proof: Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{B}' = \{b'_1, \dots, b'_m\}$ be two distinct bases of V. Then, by Corollary 1.5.7, we have isomorphisms

$$[-]_{\mathcal{B}} : V \to \mathbb{K}^n, \quad \text{and} \quad [-]_{\mathcal{B}'} : V \to \mathbb{K}^m.$$

Hence, we obtain an isomorphism (since the composition of two isomorphisms is again an isomorphism, by Lemma 0.2.4)

$$[-]_{\mathcal{B}'} \circ [-]_{\mathcal{B}}^{-1} : \mathbb{K}^n \to \mathbb{K}^m,$$

where $[-]_{\mathcal{B}'}^{-1} : \mathbb{K}^m \to V$ is the inverse morphism of $[-]_{\mathcal{B}'}$. Thus, using Theorem 1.4.11, we must have $n = m$, so that $\mathcal{B}$ and $\mathcal{B}'$ have the same size. □

Theorem 1.5.11 states that if V is a $\mathbb{K}$-vector space admitting a finite basis $\mathcal{B}$, then every other basis of V must have the same size as the set $\mathcal{B}$.

Definition 1.5.12. Let V be a $\mathbb{K}$-vector space, $\mathcal{B} \subset V$ a basis of V containing finitely many vectors. Then, the size of $\mathcal{B}$, $|\mathcal{B}|$, is called the *dimension of V (over $\mathbb{K}$)* and is denoted $\dim_{\mathbb{K}} V$, or simply $\dim V$ when no confusion can arise. In this case we will also say that V is *finite dimensional*. If V is a $\mathbb{K}$-vector space that does not admit a finite basis then we will say that V is *infinite dimensional*.

The Basis Theorem (Theorem 1.5.11) ensures that the dimension of a $\mathbb{K}$-vector space is a well-defined number (ie, it doesn't change when we choose a different basis of V).

Now that we have introduced the notion of dimension of a $\mathbb{K}$-vector space we can give one of the fundamental results of finite dimensional linear algebra.

Theorem 1.5.13. *Let V and W be $\mathbb{K}$-vector spaces such that $\dim_{\mathbb{K}} V = \dim_{\mathbb{K}} W < \infty$ is finite. Then, V is isomorphic to W.*

This result, in essence, classifies all finite dimensional $\mathbb{K}$-vector spaces by their dimension. It tells us that any linear algebra question we can ask in a $\mathbb{K}$-vector space V (for example, a question concerning linear independence or spans) can be translated to another $\mathbb{K}$-vector space W which we know has the same dimension as V. This follows from Proposition 1.4.12.

This principle underlies our entire approach to finite dimensional linear algebra: given a $\mathbb{K}$-vector space V such that $\dim_{\mathbb{K}} V = n$, Theorem 1.5.13 states that V is isomorphic to $\mathbb{K}^n$ and Corollary 1.5.7 states that, once we have a basis $\mathcal{B}$ of V, we can use the $\mathcal{B}$-coordinate morphism as an isomorphism from V to $\mathbb{K}^n$. Of course, we still need to find a basis! We will provide an approach to this problem after we have provided the (simple) proof of Theorem 1.5.13.

Proof: The statement that V and W have the same dimension is just saying that any basis of these vector spaces have the same number of elements. Let $\mathcal{B} \subset V$ be a basis of V, $\mathcal{C} \subset W$ a basis of W. Then, we have the coordinate morphisms

$$[-]_{\mathcal{B}} : V \to \mathbb{K}^n \quad \text{and} \quad [-]_{\mathcal{C}} : W \to \mathbb{K}^n,$$

both of which are isomorphisms. Then, the morphism

$$[-]_{\mathcal{C}}^{-1} \circ [-]_{\mathcal{B}} : V \to W,$$

is an isomorphism between V and W. □

Example 1.5.14. 1. The ordered set $\mathcal{B} = (e_1, \dots, e_n) \subset \mathbb{K}^n$ is an ordered basis of $\mathbb{K}^n$, where e_i is the column vector with a 1 in the i^{th} entry and 0 elsewhere.

We will denote this basis $\mathcal{S}^{(n)}$.

It is easy to show that $\mathcal{S}^{(n)}$ is linearly independent and that $\operatorname{span}_{\mathbb{K}} \mathcal{S}^{(n)} = \mathbb{K}^n$. Hence, we have that $\dim_{\mathbb{K}} \mathbb{K}^n = n$.

2. Let S be a finite set and denote $S = \{s_1, \dots, s_k\}$. Then, $\mathcal{B} = (e_{s_1}, \dots, e_{s_k})$ is an ordered basis of $\mathbb{K}^S$, where e_{s_i} is the elementary functions defined in Example 1.2.6.

   We have that $\mathcal{B}$ is linearly independent: for, if there is a linear relation

   $$c_1 e_{s_1} + \dots + c_k e_{s_k} = 0_{\mathbb{K}^S},$$

   then, in particular, evaluating both sides of this equation (of functions) at s_i gives

   $$c_i = c_1 e_{s_1}(s_i) + \dots + c_k e_{s_k}(s_i) = (c_1 e_{s_1} + \dots + c_k e_{s_k})(s_i) = 0_{\mathbb{K}^S}(s_i) = 0.$$

   Hence, $c_i = 0$, for every i, and $\mathcal{B}$ is linearly independent.

   Furthermore, $\mathcal{B}$ is a spanning set of $\mathbb{K}^S$: let $f \in \mathbb{K}^S$. Then, we have an equality of functions

   $$f = f(s_1)e_{s_1} + f(s_2)e_{s_2} + \dots + f(s_n)e_{s_n},$$

   which can be easily checked by showing that

   $$f(t) = (f(s_1)e_{s_1} + f(s_2)e_{s_2} + \dots + f(s_n)e_{s_n})(t), \ \forall t \in S.$$

   Hence, $f \in \operatorname{span}_{\mathbb{K}} \mathcal{B}$ so that, since f was arbitrary, we find $\operatorname{span}_{\mathbb{K}} \mathcal{B} = \mathbb{K}^S$.

   Hence, we see that $\dim_{\mathbb{K}} \mathbb{K}^S = |S|$.

3. It is not true that if S is an infinite set then $\mathcal{B} = \{e_s \mid s \in S\}$ is a basis of $\mathbb{K}^S$, even though $\mathcal{B}$ is a linearly independent set. This is discussed in a worksheet.

4. As a particular example of 2 above, we see that $Mat_{m,n}(\mathbb{K})$ has as a basis the elementary matrices $\mathcal{B} = \{e_{ij} \mid (i,j) \in \{1, \dots, m\} \times \{1, \dots, n\}\}$. These are those matrices that have 0s for all entries except for a 1 in the ij-entry.

   Hence, we see that $\dim_{\mathbb{K}} Mat_{m,n}(\mathbb{K}) = mn$.

1.5.1 Finding a basis

In this section we will provide criteria for determining when a subset E of a finite dimensional $\mathbb{K}$-vector space V is a basis. Hopefully, this is just a recollection of results that you have seen before in your first linear algebra course.

Throughout this section we will fix a finite dimensional $\mathbb{K}$-vector space V such that $\dim_{\mathbb{K}} V = n$ and an ordered basis $\mathcal{B} = (b_1, \dots, b_n)$ (which we know exists by Theorem 1.5.10).

Proposition 1.5.15. *Let $E \subset V$ be a nonempty subset of V.*

a) If E is linearly independent, then $|E| \leq n$.

b) If $\text{span}_{\mathbb{K}} E = V$*, then* $|E| \geq n$*,*

c) If $E \subset V$ *is linearly independent and* $F \subset V$ *is a spanning set, so that* $\text{span}_{\mathbb{K}} F = V$*, then either* $k = n$ *and* E *is a basis of* V*; or,* E *can be extended to a basis of* V *by adding to* E *vectors from* F*. This means, if* $E = \{e_1, \dots, e_k\}$ *then we can find* $f_{k+1}, \dots, f_n \in F$ *such that* $\{e_1, \dots, e_k, f_{k+1}, \dots, f_n\}$ *is a basis of* V*.*

Proof: a) Suppose that E is linearly independent, finite and nonempty and that $|E| > n$, say $|E| = k > n$ and denote $E = \{e_1, \dots, e_k\}$; we aim to provide a contradiction.

In this case, E can't be a basis of V, for otherwise we would contradict the Basis Theorem (Theorem 1.5.11), as E does not have n vectors. Hence, since E is linearly independent we must have that $\text{span}_{\mathbb{K}} E \neq V$ (otherwise E would be a basis, by Proposition 1.5.5). Moreover, we can't have $\mathcal{B} \subset \text{span}_{\mathbb{K}} E$ as then we would have $V = \text{span}_{\mathbb{K}} \mathcal{B} \subset \text{span}_{\mathbb{K}} E$ implying that $V = \text{span}_{\mathbb{K}} E$ (because we would have $\text{span}_{\mathbb{K}} E \subset V$ and $V \subset \text{span}_{\mathbb{K}} E$). Therefore, we can assume, without loss of generality, that $b_1 \notin \text{span}_{\mathbb{K}} E$ so that, by the Elimination Lemma (Lemma 1.3.10), we have that $E_1 = E \cup \{b_1\}$ is a linearly independent set. Then, we can't have that $\text{span}_{\mathbb{K}} E_1 = V$, else we would contradict the Basis Theorem. Thus, $\text{span}_{\mathbb{K}} E_1 \neq V$. Now, without loss of generality, we can assume that $b_2 \notin \text{span}_{\mathbb{K}} E_1$; otherwise, $b_2, \dots, b_n \in \text{span}_{\mathbb{K}} E_1$ and $b_1 \in \text{span}_{\mathbb{K}} E_1$, so that $\mathcal{B} \subset \text{span}_{\mathbb{K}} E_1$ giving $V = \text{span}_{\mathbb{K}} E_1$. Denote $E_2 = E_1 \cup \{b_2\}$. Then, again by the Elimination Lemma, we have that E_2 is a linearly independent set such that $\text{span}_{\mathbb{K}} E_2 \neq V$ (else we would contradict the Basis Theorem). Proceeding in this way we obtain subsets

$$E_i = E_{i-1} \cup \{b_i\},\ i = 1, \dots, n, \quad \text{with } E_0 \stackrel{def}{=} E,$$

that are linearly independent. In particular, we obtain the subset $E_n = E \cup \mathcal{B}$ that is linearly independent and strictly contains $\mathcal{B}$, contradicting the maximal linearly independent property of a basis. Therefore, our initial assumption that $|E| > n$ must be false, so that $|E| \leq n$.

If E is infinite, then every subset of E is linearly independent. Hence, we can find arbitrarily large linearly independent finite subsets of E. Choose a subset E' such that $|E'| > n$. Then we are back in the previous situation, which we have just cannot hold. Hence, we can't have that E is infinite.

b) This is consequence of the method of proof for Case 1 of Theorem 1.5.10. Indeed, either E is an infinite set and there is nothing to prove, or E is a finite set. Then, as in the proof of Theorem 1.5.10, we can find a basis $E' \subset E$ contained in E. Hence, by the Basis Theorem, we see that $n = |E'| \leq |E|$.

c) Let $E \subset V$ be a linearly independent subset of V. Then, by *a*) we know that $|E| \leq n$. Let us write $E = \{e_1, \dots, e_k\}$, so that $k \leq n$.

Case 1: $k = n$: In this case we have that E is a basis itself. This follows by the maximal linear independence property defining a basis as follows: by *a*) we know that every linearly independent set must have at most n vectors in it. Thus, if $E \subset E'$ and E' is linearly independent, then we must necessarily have $E' = E$, since E' cannot have any more than n vectors. This is just the maximal linear independence property defining a basis. Hence, E is a basis of V.

Case 2: $k < n$: Now, by *b*), we know that any spanning set of V must have at least n vectors in it. Hence, since $k < n$ we have $\text{span}_{\mathbb{K}} E \subset V$ while $\text{span}_{\mathbb{K}} E \neq V$. We claim that there exists $f_{k+1} \in F$ such that $f_{k+1} \notin \text{span}_{\mathbb{K}} E$. For, if not, then we would have $F \subset \text{span}_{\mathbb{K}} E$, so that $V = \text{span}_{\mathbb{K}} F \subset \text{span}_{\mathbb{K}} E \subset V$, which is absurd as $\text{span}_{\mathbb{K}} E \neq V$. Then, $F_1 = E \cup \{f_{k+1}\}$ is a linearly independent set, by the Elimination Lemma. If $\text{span}_{\mathbb{K}} F_1 = V$ then we have that F_1 is a basis and we are done. Otherwise, $\text{span}_{\mathbb{K}} F_1 \neq V$. As before, we can find $f_{k+2} \in F$ such that $f_{k+2} \notin \text{span}_{\mathbb{K}} F_1$ and obtain linearly independent set $F_2 = F_1 \cup \{f_{k+2}\}$. Then, either $\text{span}_{\mathbb{K}} F_2 = V$ and we are done, or $\text{span}_{\mathbb{K}} F_2 \neq V$ and we can define a linearly independent set F_3. Proceeding in this manner we either obtain a basis F_i, for some $i < n - k$, or we obtain a linearly independent set F_{n-k} and we are back in Case 1, so that F_{n-k} must be a basis. In either case, we find a basis of the required form. □

Corollary 1.5.16. *Let* V *be a* $\mathbb{K}$*-vector space such that* $\dim_{\mathbb{K}} V = n$ *and* $E \subset V$*.*

- *If* E *is linearly independent and* $|E| = n$*, then* E *is a basis of* V*.*

- *If* $\text{span}_{\mathbb{K}} E = V$ *and* $|E| = n$*, then* E *is a basis of* V*.*

Proof: The first statement was shown in c). The second statement is left to the reader. □

Corollary 1.5.17. *Let V be a $\mathbb{K}$-vector space such that $\dim_{\mathbb{K}} V = n$, $U \subset V$ a subspace. Then, $\dim_{\mathbb{K}} U \leq n$. Moreover, if $\dim_{\mathbb{K}} U = n$, then $U = V$.*

Proof: Let $\mathcal{B}' \subset V$ be a basis of U. Then, $\mathcal{B}'$ is a linearly independent subset of U, therefore a linearly independent subset of V. Hence, by Proposition 1.5.15, we have that $\mathcal{B}'$ contains no more than n vectors. By the definition of dimension the result follows.

Moreover, suppose that $\dim_{\mathbb{K}} U = n$. Then, there is a subset $\mathcal{B}'$ of U that is linearly independent and contains exactly n vectors. Hence, by the previous Corollary, $\mathcal{B}'$ is a basis of V. So, since $\operatorname{span}_{\mathbb{K}} \mathcal{B}' = U$ and $\operatorname{span}_{\mathbb{K}} \mathcal{B}' = V$ we have $U = V$. □

Corollary 1.5.18. *Let V be a $\mathbb{K}$-vector space, $U \subset V$ a subspace. Then, any basis of U can be extended to a basis of V.*

Proof: Let $\mathcal{B}' = \{b'_1, \dots, b'_r\}$ be a basis of U and $\mathcal{B} = \{b_1, \dots, b_n\}$ a basis of V; in particular, $\operatorname{span}_{\mathbb{K}} \mathcal{B} = V$. Then, by Proposition 1.5.15, part c), we can extend $\mathcal{B}'$ to a basis of V using vectors from $\mathcal{B}$. □

Corollary 1.5.19. *Let V be a $\mathbb{K}$-vector space, $U \subset V$ a subspace. Then, there exists a subspace $W \subset V$ such that $V = U \oplus W$. Moreover, in this case we have*

$$\dim V = \dim U + \dim W,$$

and if $\mathcal{B}'$ is any basis of U and $\mathcal{B}''$ is any basis of W then $\mathcal{B} = \mathcal{B}' \cup \mathcal{B}''$ is a basis of V.

Proof: Let $\mathcal{B}' = \{b'_1, \dots, b'_r\}$ be a basis of U and extend to a basis $\mathcal{B} = \{b'_1, \dots, b'_r, b_{r+1}, \dots, b_n\}$ of V, using the previous Corollary. Then, let $W = \operatorname{span}_{\mathbb{K}}\{b_{r+1}, \dots, b_n\}$. Then, since $\mathcal{B}$ is a basis we have that $V = U + W$ (as every vector in v can be expressed as a linear combination of vectors from $\mathcal{B}$). We need to show that $U \cap W = \{0_V\}$. So, let $x \in U \cap W$. Then, we have

$$x = \lambda_1 b'_1 + \dots + \lambda_r b'_r \in U,$$

and

$$x = \mu_1 b_{r+1} + \dots + \mu_{n-r} b_n \in W.$$

Hence,

$$\mu_1 b_{r+1} + \dots + \mu_{n-r} b_n = x = \lambda_1 b'_1 + \dots + \lambda_r b'_r,$$

giving a linear relation

$$\lambda_1 b'_1 + \dots + \lambda_r b'_r - (\mu_1 b_{r+1} + \dots + \mu_{n-r} b_n) = 0_V.$$

Thus, as $\mathcal{B}$ is linearly independent then this linear relation must be trivial so that

$$\mu_1 = \dots = \mu_{n-r} = \lambda_1 = \dots = \lambda_r = 0;$$

hence, $x = 0_V$ so that $U \cap W = \{0_V\}$.

The statement concerning the dimension of V follows from the above proof.

The final statement follows from a dimension count and a simple argument showing that $\mathcal{B} = \mathcal{B}' \cup \mathcal{B}''$ is a linearly independent set. Now we can use Corollary 1.5.16 to deduce that $\mathcal{B}$ is a basis of V. The details are left to the reader. □

We end this section with an important formula relating the dimension of subspaces, the so-called *dimension formula*.

Proposition 1.5.20 (Dimension formula)**.** *Let V be a $\mathbb{K}$-vector space, $U, W \subset V$ two subspaces of V. Then,*

$$\dim(U + W) = \dim U + \dim W - \dim U \cap W.$$

Note that here we are not assuming that $V = U + W$.

Proof: Let $X = U + W$ so that $X \subset V$ is a subspace of V and can be considered as a $\mathbb{K}$-vector space in its own right. Moreover, we have that $U, W, U \cap W \subset X$ are all subspaces of X and $U \cap W$ is a subspace of both U and W.

Now, if $U \subset W$ (resp. $W \subset U$) then we have $U + W = W$ (resp. $U + W = U$) and $U \cap W = U$ (resp. $U \cap W = W$). So, in this case the result follows easily.

Therefore, we will assume that $U \nsubseteq W$ and $W \nsubseteq U$ so that $U \cap W \subset U$ and $U \cap W \subset W$ while $U \cap W \neq U, W$. Using the previous Corollary we have that there are subspaces $U' \subset U$ and $W' \subset W$ such that

$$U = (U \cap W) \oplus U', \quad \text{and} \quad W = (U \cap W) \oplus W'.$$

Let $\mathcal{B}_1$ be a basis of $U \cap W$, $\mathcal{B}_2$ a basis of U' and $\mathcal{B}_3$ a basis of W'. We claim that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ is a basis of $U + W$. Indeed, since $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis of U and $\mathcal{B}_1 \cup \mathcal{B}_3$ is a basis of W (by the previous Corollary), we certainly have that $\operatorname{span}_{\mathbb{K}} \mathcal{B} = U + W$[28]. Furthermore, it is straightforward to show that $\mathcal{B}$ is linearly independent[29] thereby giving that $\mathcal{B}$ is a basis of $U + W$. Thus,

$$\dim(U + W) = \dim U' + \dim U \cap W + \dim W',$$

and

$$\dim U = \dim U' + \dim U \cap W, \quad \text{and} \quad \dim W = \dim W' + \dim U \cap W.$$

Comparing these equations gives the result. □

Example 1.5.21. 1. The subset

$$E = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \right\} \subset \mathbb{Q}^3,$$

defines a basis of $\mathbb{Q}^3$. Since E has consists of 3 vectors and $\dim_{\mathbb{Q}} \mathbb{Q}^3 = 3$, we need only show that E is linearly independent (Corollary 1.5.16). So, by Example 1.3.6, this amounts to showing that the homogeneous matrix equation

$$A\underline{x} = \underline{0},$$

has only one solution, namely the zero solution, where A is the matrix whose columns are the vectors in E. Now, since we can row-reduce

$$\begin{bmatrix} 1 & -1 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \sim I_3,$$

we find that E is indeed linearly independent, so that it must be a basis, by Corollary 1.5.16.

2. Consider the subset

$$E = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\} \subset Mat_2(\mathbb{R}).$$

Then, E is a basis of $Mat_2(\mathbb{Q})$. Again we use Corollary 1.5.16: since E has 4 vectors and $\dim_{\mathbb{R}} Mat_2(\mathbb{R}) = 2.2 = 4$ we need only show that E is linearly independent or that it spans $Mat_2(\mathbb{R})$. We will show that $\operatorname{span}_{\mathbb{R}} E = Mat_2(\mathbb{R})$. So, let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

[28] The reader should check this.

[29] Again, this is an exercise left to the reader.

Then, we have

$$A = a_{11}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + a_{22}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} + \frac{(a_{12}+a_{21})}{2}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} + \frac{(a_{12}-a_{21})}{2}\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix},$$

so that $A \in \text{span}_{\mathbb{R}}\, E$. Since A was arbitrary we must have $\text{span}_{\mathbb{R}}\, E = Mat_2(\mathbb{R})$.

Furthermore, if we consider the ordered basis

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

then the $\mathcal{B}$-coordinate morphism is the linear morphism

$$[-]_{\mathcal{B}} : Mat_2(\mathbb{R}) \to \mathbb{R}^4 \;;\; A = \begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix} \mapsto \begin{bmatrix}a_{11}\\ (a_{12}+a_{21})/2\\ a_{22}\\ (a_{12}-a_{21})/2\end{bmatrix}$$

1.6 Coordinates

([1], Ch. 5)

Throughout this section we assume that all $\mathbb{K}$-vector spaces are finite dimensional.

1.6.1 Solving problems

The results of the previous section form the theoretical underpinning of how we hope to solve linear algebra problems in practice. The existence of an ordered basis $\mathcal{B} = (b_1, \dots, b_n)$ of a $\mathbb{K}$-vector space V from Theorem 1.5.10, such that $n = \dim V$, along with Corollary 1.5.6 and Corollary 1.5.7 allow us to introduce the notion of $\mathcal{B}$-coordinates on V: we have an isomorphism

$$[-]_{\mathcal{B}} : V \to \mathbb{K}^n \;;\; v \mapsto [v]_{\mathcal{B}} = \begin{bmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{bmatrix},$$

where $v = \lambda_1 b_1 + \dots + \lambda_n b_n$ is the unique expression determined in Corollary 1.5.6. Then, using Proposition 1.4.12, we know that questions concerning linear independence and spans of subsets in V have the same answers if we translate them to questions in $\mathbb{K}^n$ via the $\mathcal{B}$-coordinate map. Since we are then talking about sets of column vectors we can use row-reduction methods to answer the question that was originally posed concerning vectors in V.

So, we have the following approach to solving questions about linear independence/spans of subsets $E \subset V$ in finite dimensional $\mathbb{K}$-vector spaces V (we suppose that $n = \dim V$):

0. If $|E| > n$ then E is linearly dependent; if $|E| < n$ then it is not possible that E spans V. This follows from Proposition 1.5.15.
1. Determine an ordered basis $\mathcal{B}$ of V using, for example, Corollary 1.5.16.
2. Using the $\mathcal{B}$-coordinate morphism $[-]_{\mathcal{B}} : V \to \mathbb{K}^n$, determine the set $[E]_{\mathcal{B}} = \{[e]_{\mathcal{B}} \mid e \in E\}$.
3. Using row-reduction determine the linear independence/spanning properties of the set $[E]_{\mathcal{B}}$.
4. By Proposition 1.4.12, linear independence/spanning properties of $[E]_{\mathcal{B}}$ are the same as those of $E \subset V$.

1.6.2 Change of basis/change of coordinates

We have just seen an approach to solving linear independence/spanning property problems for a (finite dimensional) $\mathbb{K}$-vector space V. However, it is not necessarily true that everyone will choose the same ordered basis $\mathcal{B}$ of V: for example, we could choose a different ordering on the same set $\mathcal{B}$, leading to a different ordered basis; or, you could choose an ordered basis that is a completely distinct set from an ordered basis I may choose.

Of course, this should not be a problem when we solve problems as the linear independence/spanning properties of a subset E should not depend on how we want to 'view' that subset, ie, what coordinates we choose. However, given two distinct ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$ of V, it will be the case in general that $[E]_{\mathcal{B}_1}$ and $[E]_{\mathcal{B}_2}$ are different sets so that if we wanted to compare our work with another mathematician we we would need to know how to translate between our two different 'viewpoints' we've adopted, ie, we need to know how to change coordinates.

Proposition 1.6.1 (Change of coordinates)**.** *Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{C} = \{c_1, \dots, c_n\}$ be two ordered bases of V. Let $P_{\mathcal{C}\leftarrow\mathcal{B}}$ be the $n \times n$ matrix*

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[b_1]_{\mathcal{C}}[b_2]_{\mathcal{C}} \cdots [b_n]_{\mathcal{C}}],$$

so that the i^{th} column is $[b_i]_{\mathcal{C}}$, the $\mathcal{C}$-coordinates of b_i. Then, for every $v \in V$, we have

$$[v]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[v]_{\mathcal{B}}.$$

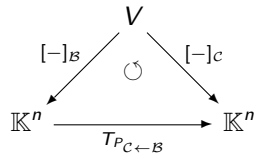
Moreover, if $A \in Mat_n(\mathbb{K})$ is such that

$$[v]_{\mathcal{C}} = A[v]_{\mathcal{B}},\ \forall v \in V,$$

then $A = P_{\mathcal{C}\leftarrow\mathcal{B}}$.

We call $P_{\mathcal{C}\leftarrow\mathcal{B}}$ the *change of coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$*. The formula just given tells us that, given the $\mathcal{B}$-coordinates of a vector $v \in V$, to obtain the $\mathcal{C}$-coordinates of v we must multiply the $\mathcal{B}$-coordinate vector of v on the left by $P_{\mathcal{C}\leftarrow\mathcal{B}}$. Moreover, we see that the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{C}$ is uniquely characterised by this property.

Remark 1.6.2. 1. We can organise this data into a diagram



where

$$T_{P_{\mathcal{C}\leftarrow\mathcal{B}}} : \mathbb{K}^n \to \mathbb{K}^n\ ;\ \underline{x} \mapsto P_{\mathcal{C}\leftarrow\mathcal{B}}\underline{x}.$$

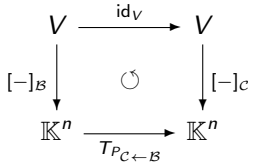
is the linear morphism defined by the matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$.

The symbol '↺' that appears is to be translated as

'the composite morphism $T_{P_{\mathcal{C}\leftarrow\mathcal{B}}} \circ [-]_{\mathcal{B}} : V \to \mathbb{K}^n$ equals the morphism $[-]_{\mathcal{C}} : V \to \mathbb{K}^n$.'

That is, if we start at (the domain) V and follow the arrows either to the left or right then we get the same answer in (the codomain) $\mathbb{K}^n$ (at the bottom right of the diagram). In this case, we say that the **diagram commutes**.

We could also write this diagram as



where $\mathrm{id}_V : V \to V$ is the identity morphism from Example 1.4.8. The reason we are also considering this diagram will become apparent in the following sections.

2. Suppose that $P_{\mathcal{B}\leftarrow\mathcal{C}}$ is the change of coordinate matrix from $\mathcal{C}$ to $\mathcal{B}$. This means that for every $v \in V$ we have

$$[v]_{\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{C}}[v]_{\mathcal{C}}.$$

Then, if we want to change back to $\mathcal{C}$-coordinates, we simply multply on the left by $P_{\mathcal{C}\leftarrow\mathcal{B}}$ so that

$$[v]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[v]_{\mathcal{B}} = P_{\mathcal{C}\leftarrow\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}}[v]_{\mathcal{C}},\ \forall v \in V.$$

This means that the morphism

$$T_{P_{\mathcal{C}\leftarrow\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}}} : \mathbb{K}^n \to \mathbb{K}^n\ ;\ \underline{x} \mapsto P_{\mathcal{C}\leftarrow\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}}\underline{x},$$

is the identity morphism $\mathrm{id}_{\mathbb{K}^n}$ of $\mathbb{K}^n$; this uses the fact that V is isomorphic to $\mathbb{K}^n$.[30]

We will see later on that this implies that $P_{\mathcal{C}\leftarrow\mathcal{B}}$ and $P_{\mathcal{B}\leftarrow\mathcal{C}}$ are invertible matrices and are inverse to each other:

$$P_{\mathcal{C}\leftarrow\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}} = I_n = P_{\mathcal{B}\leftarrow\mathcal{C}}P_{\mathcal{C}\leftarrow\mathcal{B}},$$

where I_n is the $n \times n$ identity matrix.

This should not be surprising: all we have shown here is that the operations 'change coordinates from $\mathcal{B}$ to $\mathcal{C}$' and 'change coordinates from $\mathcal{C}$ to $\mathcal{B}$' are inverse to each other.

Of course, you can also obtain this result knowing that a matrix with linearly independent columns is invertible; this should be familiar to you from your first linear algebra course. However, we have just stated a stronger result: not only have we determined that a change of coordinate matrix is invertible, we have provided what the inverse actually is.

Example 1.6.3. 1. Consider the two ordered bases $\mathcal{S}^{(3)} = (e_1, e_2, e_3)$ and

$$\mathcal{B} = \left(\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

of $\mathbb{Q}^3$. Then, what is the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{S}^{(3)}$? We use the formula given above: we have

$$P_{\mathcal{S}^{(3)}\leftarrow\mathcal{B}} = [[b_1]_{\mathcal{S}^{(3)}}[b_2]_{\mathcal{S}^{(3)}}[b_3]_{\mathcal{S}^{(3)}}] = \begin{bmatrix} 0 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Therefore, the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{S}^{(3)}$ is simply the matrix whose i^{th} column is the i^{th} basis vector of the ordered basis $\mathcal{B}$.

Moreover, if we want to determine the change of coordinate matrix from $\mathcal{S}^{(3)}$ to $\mathcal{B}$ we need to determine the inverse matrix of $P_{\mathcal{S}^{(3)}\leftarrow\mathcal{B}}$, using row-reduction methods, for example.

2. In general, if $\mathcal{S}^{(n)} = (e_1, \dots, e_n)$ is the standard ordered basis of $\mathbb{K}^n$ and $\mathcal{B} = (b_1, \dots, b_n)$ is any other ordered basis of $\mathbb{K}^n$, then the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{S}^{(n)}$ is

$$P_{\mathcal{S}^{(n)}\leftarrow\mathcal{B}} = [b_1\ b_2\ \cdots\ b_n].$$

Again, if we wish to determine the change of coordinate matrix from $\mathcal{S}^{(n)}$ to $\mathcal{B}$ we need to determine the inverse matrix of $P_{\mathcal{S}^{(n)}\leftarrow\mathcal{B}}$. This may not be so easy for large matrices.

[30] Why is this true?

1.7 Linear morphisms II

In this section we will discuss the relationship between linear morphisms (of finite dimensional $\mathbb{K}$-vector spaces) and matrices. This material should be familiar to you from your first linear algebra course.

Throughout this section all $\mathbb{K}$-vector spaces will be assumed finite dimensional.

Definition 1.7.1. Let $f : V \to W$ be a linear morphism of $\mathbb{K}$-vector spaces, $\mathcal{B} = (b_1, \dots, b_n) \subset V, \mathcal{C} = (c_1, \dots, c_m) \subset W$ ordered bases of V, W. Then, the *matrix of f with respect to $\mathcal{B}$ and $\mathcal{C}$* is the $m \times n$ matrix

$$[f]^{\mathcal{C}}_{\mathcal{B}} = [[f(b_1)]_{\mathcal{C}}\ [f(b_2)]_{\mathcal{C}} \cdots [f(b_n)]_{\mathcal{C}}]\,,$$

so that the i^{th} column of $[f]^{\mathcal{C}}_{\mathcal{B}}$ is the $\mathcal{C}$-coordinate vector of $f(b_i) \in W$.

If $V = W$ and $\mathcal{B} = \mathcal{C}$ then we write $[f]_{\mathcal{B}} \stackrel{def}{=} [f]^{\mathcal{B}}_{\mathcal{B}}$.

Lemma 1.7.2. *Let $f : V \to W$ be a linear morphism of $\mathbb{K}$-vector spaces, $\mathcal{B} \subset V, \mathcal{C} \subset W$ ordered bases of V, W. Then, for every $v \in V$, we have*

$$[f(v)]_{\mathcal{C}} = [f]^{\mathcal{C}}_{\mathcal{B}}[v]_{\mathcal{B}}.$$

Moreover, if A is an $m \times n$ matrix such that

$$[f(v)]_{\mathcal{C}} = A[v]_{\mathcal{B}}, \quad \text{for every } v \in V,$$

then $A = [f]^{\mathcal{C}}_{\mathcal{B}}$.

This result should be familiar to you. Note that the standard matrix A_f we defined previously for a linear morphism $f \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$ is just

$$A_f = [f]^{\mathcal{S}^{(m)}}_{\mathcal{S}^{(n)}}.$$

We can record the conclusion of the Lemma 1.7.2 in a diagram in a similar fashion as we did in the previous section for $P_{\mathcal{C}\leftarrow\mathcal{B}}$. We have the commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle [-]_{\mathcal{B}}}\downarrow & \circlearrowleft & \downarrow{\scriptstyle [-]_{\mathcal{C}}} \\ \mathbb{K}^n & \xrightarrow[T_{[f]^{\mathcal{C}}_{\mathcal{B}}}]{} & \mathbb{K}^m \end{array}$$

where $T_{[f]^{\mathcal{C}}_{\mathcal{B}}} : \mathbb{K}^n \to \mathbb{K}^m$ is the 'multiplication by $[f]^{\mathcal{C}}_{\mathcal{B}}$' morphism and the symbol '$\circlearrowleft$' is translated to mean

'the composite morphism $[-]_{\mathcal{C}} \circ f : V \to \mathbb{K}^m$ equals the composite morphism $T_{[f]^{\mathcal{C}}_{\mathcal{B}}} \circ [-]_{\mathcal{B}} : V \to \mathbb{K}^m$';

this is precisely the statement of Lemma 1.7.2.

So, given ordered bases $\mathcal{B} = (b_1, \dots, b_n) \subset V$ and $\mathcal{C} = (c_1, \dots, c_m) \subset W$ of $\mathbb{K}$-vector space V and W, we have just defined a function

$$[-]^{\mathcal{C}}_{\mathcal{B}} : \mathrm{Hom}_{\mathbb{K}}(V, W) \to Mat_{m,n}(\mathbb{K})\ ;\ f \mapsto [f]^{\mathcal{C}}_{\mathcal{B}}.$$

In fact, this correspondence obeys some particularly nice properties:

Theorem 1.7.3. *The function*

$$[-]^{\mathcal{C}}_{\mathcal{B}} : \mathrm{Hom}_{\mathbb{K}}(V, W) \to Mat_{m,n}(\mathbb{K})\ ;\ f \mapsto [f]^{\mathcal{C}}_{\mathcal{B}},$$

satisfies the following properties:

a) $[-]_{\mathcal{B}}^{\mathcal{C}}$ *is an isomorphism of* $\mathbb{K}$*-vector spaces,*

b) if $f \in \mathrm{Hom}_{\mathbb{K}}(U,V), g \in \mathrm{Hom}_{\mathbb{K}}(V,W)$ *and* $\mathcal{A} \subset U, \mathcal{B} \subset V, \mathcal{C} \subset W$ *are bases of* U, V, W*, then*

$$[g \circ f]_{\mathcal{A}}^{\mathcal{C}} = [g]_{\mathcal{B}}^{\mathcal{C}} [f]_{\mathcal{A}}^{\mathcal{B}}.$$

Here

$$g \circ f : U \to V \to W \in \mathrm{Hom}_{\mathbb{K}}(U,W),$$

is the composite morphism and on the RHS of the equation we are considering multiplication of matrices.

c) for the identity morphism $\mathrm{id}_V \in \mathrm{Hom}_{\mathbb{K}}(V,V)$ *we have*

$$[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{C}} = I_n,$$

where I_n *is the* $n \times n$ *identity matrix.*

d) if $V = W$ *then*

$$[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}}.$$

e) If $A \in Mat_{m,n}(\mathbb{K})$ *and*

$$T_A : \mathbb{K}^n \to \mathbb{K}^m \; ; \; \underline{x} \mapsto A\underline{x},$$

so that $T_A \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$*. Then, if* $\mathcal{S}^{(i)} = (e_1, \ldots, e_i)$ *is the standard basis of* $\mathbb{K}^i$*, then*

$$[T_A]_{\mathcal{S}^{(n)}}^{\mathcal{S}^{(m)}} = A.$$

We will now show how we can translate properties of morphisms into properties of matrices:

Theorem 1.7.4. *Let* $f \in \mathrm{Hom}_{\mathbb{K}}(V,W)$ *be a linear morphism of* $\mathbb{K}$*-vector spaces* V, W *and let* $\mathcal{B} \subset V, \mathcal{C} \subset W$ *be ordered bases of* V, W*. Then,*

a) f *is injective if and only if* $[f]_{\mathcal{B}}^{\mathcal{C}}$ *has a pivot in every column,*

b) f *is surjective if and only if* $[f]_{\mathcal{B}}^{\mathcal{C}}$ *has a pivot in every row,*

c) f *is an isomorphism if and only if* $[f]_{\mathcal{B}}^{\mathcal{C}}$ *is a square matrix and has a pivot in every row/column,*

d) Suppose $\dim V = \dim W$*. Then,* f *is injective if and only if* f *is surjective. In particular,*

$$\text{'}f \text{ injective} \implies f \text{ surjective} \implies f \text{ bijective} \implies f \text{ injective'}.$$

Remark 1.7.5. 1. Theorem 1.7.3 states various properties that imply that the association of a linear morphism to its matrix (with respect to some ordered bases) behaves well and obeys certain desirable properties.

- *a*) implies that any question concerning the linear algebra properties of the set of $\mathbb{K}$-linear morphisms can be translated into a question concerning matrices. In particular, since it can be easily seen that $\dim_{\mathbb{K}} Mat_{m,n}(\mathbb{K}) = mn$ and isomorphic $\mathbb{K}$-vector spaces have the same dimension, we must therefore have that

  $$\dim_{\mathbb{K}} \mathrm{Hom}_{\mathbb{K}}(V,W) = mn,$$

  so that $\mathrm{Hom}_{\mathbb{K}}(V,W)$ is finite dimensional.

- *b*) can be summarised by the slogan

  '**matrix multiplication is composition of morphisms**'.

  Together with *e*) this implies that, for an $m \times n$ matrix A and an $n \times p$ matrix B, we have $T_A \circ T_B = T_{AB}$.

2. Theorem 1.7.4 provides a way to show that a linear morphism satisfies certain properties, assuming we have found bases of the domain and codomain.

Conversely, Theorem 1.7.4 is also useful in determining properties of matrices by translating to a property of morphisms. For example, suppose that A, B are $n \times n$ matrices such that $AB = I_n$. By definition, a square matrix P is invertible if and only if there is a square matrix Q such that $PQ = QP = I_n$. Thus, even though we know that $AB = I_n$, in order to show that A (or B) is invertible, we would need to show also that $BA = I_n$. This is difficult to show directly (ie, only using matrices) if you only know that $AB = I_n$. However, if we consider the linear maps T_A and T_B then

$$AB = I_n \implies T_A \circ T_B = T_{AB} = id_{\mathbb{K}^n}. \quad \text{(Theorem 1.7.3, b), c), e))}$$

Now, by Lemma 0.2.4, since $\mathrm{id}_{\mathbb{K}^n}$ is injective then T_B is injective. Thus, T_B is an isomorphism by Theorem 1.7.4 so there exists a morphism $g \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^n)$ such that $g \circ T_B = T_B \circ g = \mathrm{id}_{\mathbb{K}^n}$. Since $T_A \circ T_B = \mathrm{id}_{\mathbb{K}^n}$, then

$$g = \mathrm{id}_{\mathbb{K}^n} \circ g = (T_A \circ T_B) \circ g = T_A \circ (T_B \circ g) = T_A \circ \mathrm{id}_{\mathbb{K}^n} = T_A,$$

because $f \circ \mathrm{id}_{\mathbb{K}^n} = f$, for any function f with domain $\mathbb{K}^n$, and $\mathrm{id}_{\mathbb{K}^n} \circ f = f$, for any function f with codomain $\mathbb{K}^n$. Hence, we have shown that $g = T_A$ so that $\mathrm{id}_{\mathbb{K}^n} = T_B \circ g = T_B \circ T_A = T_{BA}$. Therefore, $I_n = BA$. Note that we have repeatedly used (various parts of) Theorem 1.7.3 in this last collection of justifications.

Suppose that we have distinct ordered bases $\mathcal{B}_1, \mathcal{B}_2 \subset V$ and $\mathcal{C}_1, \mathcal{C}_2 \subset W$ of the $\mathbb{K}$-vector spaces V, W and $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$. How are the matrices $[f]_{\mathcal{B}_1}^{\mathcal{C}_1}$ and $[f]_{\mathcal{B}_2}^{\mathcal{C}_2}$ related?

Proposition 1.7.6. *The matrices $[f]_{\mathcal{B}_1}^{\mathcal{C}_1}$ and $[f]_{\mathcal{B}_2}^{\mathcal{C}_2}$ satsify the following relation*

$$[f]_{\mathcal{B}_2}^{\mathcal{C}_2} = P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1} [f]_{\mathcal{B}_1}^{\mathcal{C}_1} P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2},$$

where we are considering multiplication of matrices on the RHS of this equation. Moreover, if there exists a matrix B such that

$$B = P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1} [f]_{\mathcal{B}_1}^{\mathcal{C}_1} P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2},$$

then $B = [f]_{\mathcal{B}_2}^{\mathcal{C}_2}$.

Proof: The proof is trivial once we have Theorem 1.7.3. If we consider that $P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2} = [\mathrm{id}_V]_{\mathcal{B}_2}^{\mathcal{B}_1}$ and $P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1} = [\mathrm{id}_W]_{\mathcal{C}_1}^{\mathcal{C}_2}$ then the RHS of the desired relation is

$$P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1} [f]_{\mathcal{B}_1}^{\mathcal{C}_1} P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2} = [id_W]_{\mathcal{C}_1}^{\mathcal{C}_2} [f]_{\mathcal{B}_1}^{\mathcal{C}_1} [\mathrm{id}_V]_{\mathcal{B}_2}^{\mathcal{B}_1} = [\mathrm{id}_V \circ f \circ \mathrm{id}_V]_{\mathcal{B}_2}^{\mathcal{C}_2} = [f]_{\mathcal{B}_2}^{\mathcal{C}_2}.$$

□

Corollary 1.7.7. *Let $f \in \mathrm{End}_{\mathbb{K}}(V)$ be an endomorphism of a $\mathbb{K}$-vector space V (recall Definition 1.4.1), $\mathcal{B}, \mathcal{C} \subset V$ ordered bases of V. Then, if we denote $P = P_{\mathcal{B} \leftarrow \mathcal{C}}$, we have*

$$(*) \qquad [f]_{\mathcal{C}} = P^{-1} [f]_{\mathcal{B}} P.$$

Example 1.7.8. 1. Consider the linear morphism

$$f : \mathbb{Q}^3 \to \mathbb{Q}^2 \; ; \; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + 2x_2 \\ -\frac{1}{2}x_2 + 3x_3 \end{bmatrix}.$$

Then, f is linear since we can write

$$f(\underline{x}) = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -\frac{1}{2} & 3 \end{bmatrix} \underline{x}, \quad \text{for every } \underline{x} \in \mathbb{Q}^3.$$

Here, we have

$$A_f = [f]^{\mathcal{S}^{(2)}}_{\mathcal{S}^{(3)}} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -\frac{1}{2} & 3 \end{bmatrix}.$$

Consider the ordered bases

$$\mathcal{B} = \left(\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) \subset \mathbb{Q}^3, \quad \mathcal{C} = \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \subset \mathbb{Q}^2.$$

Then, we can use Proposition 1.7.6 to determine $[f]^{\mathcal{C}}_{\mathcal{B}}$.

We have seen in Example 1.6.3 that

$$P_{\mathcal{S}^{(3)} \leftarrow \mathcal{B}} = \begin{bmatrix} 0 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \quad P_{\mathcal{S}^{(2)} \leftarrow \mathcal{C}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

so that

$$P_{\mathcal{C} \leftarrow \mathcal{S}^{(2)}} = P^{-1}_{\mathcal{S}^{(2)} \leftarrow \mathcal{C}} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Hence,

$$[f]^{\mathcal{C}}_{\mathcal{B}} = P_{\mathcal{C} \leftarrow \mathcal{S}^{(2)}} [f]^{\mathcal{S}^{(2)}}_{\mathcal{S}^{(3)}} P_{\mathcal{S}^{(3)} \leftarrow \mathcal{B}} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & -\frac{1}{2} & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{11}{4} & \frac{1}{4} \\ 3 & -\frac{3}{4} & 4 \end{bmatrix}.$$

2. Consider the linear morphism

$$g : Mat_2(\mathbb{R}) \to Mat_2(\mathbb{R}) \; ; \; A \mapsto A - A^t,$$

where A^t is the transpose of A. It is an exercise to check that g is linear.

We have the standard ordered basis of $Mat_2(\mathbb{R})$, $\mathcal{S} = (e_{11}, e_{12}, e_{21}, e_{22})$, where e_{ij} is the 2×2 matrix with 0 everywhere except a 1 in the ij-entry. Also, we have the ordered bases[31]

$$\mathcal{B} = (e_{12}, e_{21}, e_{11} - e_{22}, e_{11} + e_{22}), \; \mathcal{C} = (e_{11}, e_{22}, e_{12} + e_{21}, e_{12} - e_{21}) \subset Mat_2(\mathbb{R}).$$

Now, we see that

$$g(e_{11}) = 0, \; g(e_{12}) = e_{12} - e_{21}, \; g(e_{21}) = -e_{12} + e_{21}, \; g(e_{22}) = 0,$$

so that

$$[g]_{\mathcal{S}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We use Proposition 1.7.6 to determine $[g]^{\mathcal{C}}_{\mathcal{B}}$. We have

$$P_{\mathcal{S} \leftarrow \mathcal{B}} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \; P_{\mathcal{S} \leftarrow \mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

[31] Check that these are bases of $Mat_2(\mathbb{R})$.

Then,

$$P_{\mathcal{C}\leftarrow\mathcal{S}} = P^{-1}_{\mathcal{S}\leftarrow\mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix},$$

and we have

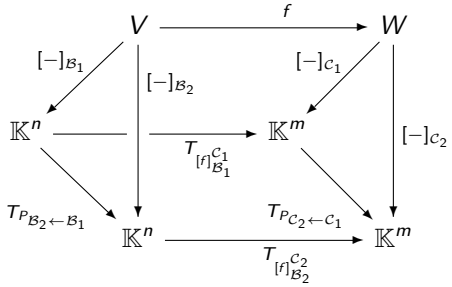
$$[g]^{\mathcal{C}}_{\mathcal{B}} = P_{\mathcal{C}\leftarrow\mathcal{S}}[g]_{\mathcal{S}}P_{\mathcal{S}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}$$

Indeed, we have that $g(e_{12}) = e_{12} - e_{21}$, which gives

$$[g(e_{12})]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Using the definition of $[g]^{\mathcal{C}}_{\mathcal{B}}$ this should be the first column, so that the matrix we have obtained above corroborates this.

Remark 1.7.9. The relationship established in Proposition 1.7.6 can be indicated in the following 'rooftop' or 'prism' diagram (think of the arrow $V \to W$ as the top of the rooftop)



Here we are assuming that all squares that appear are the commutative squares appearing after Lemma 1.7.2, and that the triangles that appear at the end of the prism are the commutative triangles that appeared in Remark 1.6.2. So, Proposition 1.7.6 corresponds to the 'bottom square' being a commutative diagram.

This diagram can be confusing at first but the more you try and understand it the better you will understand the relationship between linear morphisms, matrices and change of coordinates.

Note that in the rooftop diagram all arrows which have some vertical component are isomorphisms; this means that we can go forward and backwards along these arrows.

For example, suppose we start at V and go along the sequence of arrows $(\downarrow, \to)$. Then, the commutativity of the bottom square and the fact that that the arrows $\searrow$ are isomorphisms means we have

$$\to = (\nwarrow, \to, \searrow),$$

where $\nwarrow$ denotes the inverse morphism to $\searrow$.

Then, because we write composition of functions in the reverse order ($g \circ f$ means 'do f first, then g') we have

$$T_{[f]_{\mathcal{B}_2}^{\mathcal{C}_2}} \circ [-]_{\mathcal{B}_2} = T_{P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1}} \circ T_{[f]_{\mathcal{B}_1}^{\mathcal{C}_1}} \circ T_{P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2}} \circ [-]_{\mathcal{B}_2};$$

that is, for every $v \in V$, we have

$$[f]_{\mathcal{B}_2}^{\mathcal{C}_2}[v]_{\mathcal{B}_2} = P_{\mathcal{C}_2 \leftarrow \mathcal{C}_1}[f]_{\mathcal{B}_1}^{\mathcal{C}_1} P_{\mathcal{B}_1 \leftarrow \mathcal{B}_2}[v]_{\mathcal{B}_2},$$

and this is Proposition 1.7.6.

Definition 1.7.10 (Similar matrices)**.** Let $A, B \in Mat_n(\mathbb{K})$. We say that A *is similar to* B if and only if there exists an invertible matrix Q such that

$$A = Q^{-1}BQ.$$

This definition is symmetric with respect to A and B: namely, A is similar to B if and only if B is similar to A, since

$$A = Q^{-1}BQ \implies QAQ^{-1} = B,$$

so that if we let $P = Q^{-1}$ then we have a relation

$$B = P^{-1}AP.$$

Here we have used the (assumed known) fact that $(P^{-1})^{-1} = P$, for any invertible matrix P.

Moreover, if A is similar to B and B is similar to C, so that

$$A = Q^{-1}BQ, \quad \text{and} \quad B = P^{-1}CP,$$

then

$$A = Q^{-1}BQ = Q^{-1}P^{-1}CPQ = (PQ)^{-1}C(PQ),$$

so that A is similar to C.[32]

Corollary 1.7.7 states that matrices of linear endomorphisms with respect to different bases are similar. There is a converse to this result.

Proposition 1.7.11. *Let $A, B \in Mat_n(\mathbb{K})$ be similar matrices, so that $A = P^{-1}BP$, where $P \in \mathrm{GL}_n(\mathbb{K})$ is an invertible $n \times n$ matrix. Then, there exists a linear endomorphism $f \in \mathrm{End}_{\mathbb{K}}(\mathbb{K}^n)$ and ordered bases $\mathcal{B}, \mathcal{C} \subset \mathbb{K}^n$ such that*

$$[f]_{\mathcal{B}} = A, \quad \textit{and} \quad [f]_{\mathcal{C}} = B.$$

Proof: We take $\mathcal{C} = \mathcal{S}^{(n)} = (e_1, \dots, e_n)$, $\mathcal{B} = (b_1, \dots, b_n)$, where b_i is the i^{th} column of P, and $f = T_B \in \mathrm{End}_{\mathbb{K}}(\mathbb{K}^n)$. The details are left to the reader. □

Hence, Proposition 1.7.11 tells us that we can think of similar matrices A and B as being the matrices of the same linear morphism with respect to different ordered bases. As such, we expect that similar matrices should have certain equivalent properties; namely, those properties that can arise by considering the linear morphism T_A (or, equivalently, T_B), for example, *rank, diagonalisability, invertibility*.

1.7.1 Rank, classification of linear morphisms

Let $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ be a linear morphism and recall the definition of the kernel of f and the image of f (Definition 1.4.4).

[32] These facts, along with the trivial statement that A is similar to A, imply that the notion of similarity defines an *equivalence relation* on $Mat_n(\mathbb{K})$.

Definition 1.7.12. We define the *rank of* f, denoted $\operatorname{rank} f$, to be the number

$$\operatorname{rank} f = \dim \operatorname{im} f.$$

We define the *nullity of* f, denoted $\operatorname{nul} f$, to be the number

$$\operatorname{nul} f = \dim \ker f.$$

If A is an $m \times n$ matrix then we define the *rank of* A, denoted $\operatorname{rank} A$, to be the rank of the linear morphism T_A determined by A. Similary, we define the *nullity of* A, denoted $\operatorname{nul} A$, to be the nullity of the linear morphism T_A.

There exists a basic relationship between rank and nullity.

Theorem 1.7.13 (Rank Theorem)**.** *Let* $f \in \operatorname{Hom}_{\mathbb{K}}(V, W)$. *Then,*

$$\dim V = \operatorname{nul} f + \operatorname{rank} f.$$

Proof: By Corollary 1.5.19 we know that there is a subspace $U \subset V$ such that $V = \ker f \oplus U$. Let $\mathcal{B} = (b_1, \dots, b_r)$ be an ordered basis for U. Then, we claim that $\mathcal{C} = (f(b_1), \dots, f(b_r))$ is an ordered basis of $\operatorname{im} f$.

First, it is easy to see that the set $\{f(b_1), \dots, f(b_r)\} \subset W$ is a subset of $\operatorname{im} f$. If $v \in \operatorname{im} f$, then $v = z + u$, where $z \in \ker f$, $u \in U$ (since $V = \ker f \oplus U$). Moreover, if $u = \lambda_1 b_1 + \dots + \lambda_r b_r$ then

$$f(v) = f(z+u) = f(z) + f(u) = 0_W + f(\lambda_1 b_1 + \dots + \lambda_r b_r) = \lambda_1 f(b_1) + \dots + \lambda_r f(b_r) \in \operatorname{span}_{\mathbb{K}} \mathcal{C}.$$

Hence, since $\operatorname{im} f = \{f(v) \in W \mid v \in V\}$ then we must have $\operatorname{span}_{\mathbb{K}} \mathcal{C} = \operatorname{im} f$.

It remains to show that $\{f(b_1), \dots, f(b_r)\}$ is linearly independent: indeed, suppose we have a linear relation

$$\lambda_1 f(b_1) + \dots + \lambda_r f(b_r) = 0_W.$$

Then, since f is linear, this implies that $\lambda_1 b_1 + \dots + \lambda_r b_r \in \ker f$ and $\lambda_1 b_1 + \dots + \lambda_r b_r \in U$ (because $\mathcal{B}$ is a basis of U). Hence,

$$\lambda_1 b_1 + \dots + \lambda_r b_r \in \ker f \cap U = \{0_V\},$$

so that

$$\lambda_1 b_1 + \dots + \lambda_r b_r = 0_V.$$

Now, as $\mathcal{B}$ is linearly independent then

$$\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0.$$

Hence, $\mathcal{C}$ is linearly independent and therefore a basis of $\operatorname{im} f$.

Now, using Corollary 1.5.19, we see that

$$\dim V = \operatorname{nul} f + r = \operatorname{nul} f + \operatorname{rank} f,$$

by the previous discussion. □

Lemma 1.7.14. *Let* A *be an* $m \times n$ *matrix. Then, the rank of* A *is equal to the maximal number of linearly independent columns of* A.

Proof: Let us write

$$A = [a_1 \ a_2 \ \cdots \ a_n],$$

so that the i^{th} column of A is the vector $a_i \in \mathbb{K}^m$.

Consider the linear morphism $T_A \in \operatorname{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$. Then, we have defined $\operatorname{rank} A = \operatorname{rank} T_A = \dim \operatorname{im} T_A$. Then, since

$$T_A \left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) = x_1 a_1 + \dots + x_n a_n,$$

we see that

$$(*) \qquad \operatorname{span}_{\mathbb{K}}\{a_1, \dots, a_n\} = \operatorname{im} T_A.$$

Suppose that $A \neq 0_{m,n}$. Thus, one of the columns of A is nonzero. Suppose that $a_i \neq 0_{\mathbb{K}^m}$. Then, $\{a_i\}$ is a linearly independent set and can be extended to a basis of $\operatorname{im} T_A$ using vectors from $\{a_1, \dots, a_n\}$, by $(*)$. Hence, $\operatorname{rank} A = \dim \operatorname{im} T_A$ is equal to the number of columns of A that form a basis of $\operatorname{im} T_A$. Moreover, by Proposition 1.5.15, every linearly independent set in $\operatorname{im} T_A$ has size no greater than $\operatorname{rank} A$. In particular, every linearly independent subset of the columns of A has size no greater than $\operatorname{rank} A$ while there does exist some subset having size exactly $\operatorname{rank} A$.

If $A = 0_{m,n}$ then $T_A \in \operatorname{Hom}_{\mathbb{K}}(V, W)$ is the zero morphism and $\operatorname{rank} T_A = \dim\{0_W\} = 0$. The result follows. □

The proof that we have just given for the Rank Theorem implies the following result.

Theorem 1.7.15. *Let $f \in \operatorname{Hom}_{\mathbb{K}}(V, W)$ be a $\mathbb{K}$-linear morphism and denote $r = \operatorname{rank} f$. Then, there exists ordered bases $\mathcal{B} \subset V, \mathcal{C} \subset W$ such that*

$$[f]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix},$$

where $n = \dim V$, $m = \dim W$ and $0_{i,j} \in Mat_{i,j}(\mathbb{K})$ is the zero matrix.

Proof: Consider an ordered basis $\mathcal{B}_1 = (b_1, \dots, b_{n-r})$ of $\ker f$ and extend to an ordered basis

$$\mathcal{B} = (b_1, \dots, b_{n-r}, b_{n-r+1}, \dots, b_n)$$

of V. Then, as in the proof of the Rank Theorem, we see that $(f(b_{n-r+1}), \dots, f(b_n))$ is an ordered basis of $\operatorname{im} f$. Extend this to an ordered basis

$$\mathcal{C} = (f(b_{n-r+1}), \dots, f(b_n), c_1, \dots, c_{m-r}),$$

of W. Then,

$$[f]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}.$$

□

Corollary 1.7.16. *Let $A \in Mat_{m,n}(\mathbb{K})$ such that $\operatorname{rank} A = r$. Then, there exists $P \in \mathrm{GL}_n(\mathbb{K}), Q \in \mathrm{GL}_m(\mathbb{K})$ such that*

$$Q^{-1}AP = \begin{bmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}.$$

Corollary 1.7.17. *Let $A, B \in Mat_{m,n}(\mathbb{K})$. Then, A, B are the matrices of the same linear map with respect to different bases if and only if they have the same rank.*

Proof: By the previous Corollary we can find $Q_1, Q_2 \in \mathrm{GL}_m(\mathbb{K}), P_1, P_2 \in \mathrm{GL}_n(\mathbb{K})$ such that

$$Q_1^{-1}AP_1 = \begin{bmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} = Q_2^{-1}BP_2.$$

Then, we have

$$Q_2Q_1^{-1}AP_1P_2^{-1} = B.$$

Recall that $Q_2Q_1^{-1} = (Q_1Q_2^{-1})^{-1}$. Then, as $Q_1Q_2^{-1}$ and $P_1P_2^{-1}$ are invertible matrices (products of invertible matrices are invertible) their different sets of columns are linearly independent and therefore form ordered bases $\mathcal{C} \subset \mathbb{K}^m$ and $\mathcal{B} \subset \mathbb{K}^n$. Then, if we consider the linear map T_A, the above equation says that

$$P_{\mathcal{C} \leftarrow \mathcal{S}^{(m)}} [T_A]_{\mathcal{S}^{(n)}}^{\mathcal{S}^{(m)}} P_{\mathcal{S}^{(n)} \leftarrow \mathcal{B}} = B,$$

so that Proposition 1.7.6 implies that $[T_A]_{\mathcal{B}}^{\mathcal{C}} = B$. □

Remark 1.7.18. 1. The rank of a matrix A is usually defined to be the maximum number of linearly independent columns of A. However, we have shown that our definition is equivalent to this definition.

2. Theorem 1.7.15 is just a restatement in terms of linear morphisms of a fact that you might have come across before: every $m \times n$ matrix can be row-reduced to reduced echelon form using row operations. Moreover, if we allow 'column operations', then any $m \times n$ matrix can be row/column-reduced to a matrix of the form appearing in Theorem 1.7.15.

   This requires the use of *elementary (row-operation) matrices* and we will investigate this result during discussion.

3. Corollary 1.7.17 allows us to provide a classification of $m \times n$ matrices based on their rank: namely, we can say that A and B are *equivalent* if there exists $Q \in \mathrm{GL}_m(\mathbb{K})$, $P \in \mathrm{GL}_n(\mathbb{K})$ such that

$$B = Q^{-1}AP.$$

   Then, this notion of equivalence defines an *equivalence relation* on $Mat_{m,n}(\mathbb{K})$. Hence, we can partition $Mat_{m,n}(\mathbb{K})$ into dictinct equivalence classes. Corollary 1.7.17 says that the equivalence classes can be labelled by the rank of the matrices that each class contains.

1.8 Dual Spaces (non-examinable)

In this section we are going to try and understand a *'coordinate-free'* approach to solving systems of linear equations and to prove some basic results on row-reduction; in particularm we will prove that 'row-reduction works'. This uses the notion of the *dual space* of a $\mathbb{K}$-vector space V. We will also see the dual space appear when we are discussing bilinear forms and the *adjoint* of a linear morphism (Chapter 3).

Definition 1.8.1. Let V be a $\mathbb{K}$-vector space. We define the *dual space of* V, denoted V^*, to be the $\mathbb{K}$-vector space

$$V^* \stackrel{def}{=} \mathrm{Hom}_{\mathbb{K}}(V, \mathbb{K}).$$

Therefore, vectors in V^* are $\mathbb{K}$-linear morphisms from V to $\mathbb{K}$; we will call such a linear morphism a **linear form on** V.

Notice that $\dim_{\mathbb{K}} V^* = \dim_{\mathbb{K}} V$.

Example 1.8.2. Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = (b_1, \dots, b_n)$ an ordered basis of V. Then, for each $i = 1, \dots, n$, we define $b_i^* \in V^*$ to be the linear morphism defined as follows: since $\mathcal{B}$ is a basis we know that for every $v \in V$ we can write a unique expression

$$v = c_1 b_1 + \dots + c_n b_n.$$

Then, define

$$b_i^*(v) = c_i,$$

so that b_i^* is the function that 'picks out' the i^{th} entry in the $\mathcal{B}$-coordinate vector $[v]_{\mathcal{B}}$ of v.

Proposition 1.8.3. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = (b_1, \dots, b_n)$ an ordered basis of V. Then, the function*

$$\theta_{\mathcal{B}} : V \to V^* \ ; \ v = c_1 b_1 + \dots + c_n b_n \mapsto c_1 b_1^* + \dots + c_n b_n^*,$$

is a bijective $\mathbb{K}$-linear morphism. Moreover, $\mathcal{B}^* = (b_1^*, \dots, b_n^*)$ is a basis of V^* called the dual basis of $\mathcal{B}$.

Proof: Linearity is left as an exercise to the reader.

To show that $\theta_{\mathcal{B}}$ is bijective it suffices to show that $\theta_{\mathcal{B}}$ is injective, by Theorem 1.7.4. Hence, we will show that $\ker \theta_{\mathcal{B}} = \{0_V\}$: let $v \in \ker \theta_{\mathcal{B}}$ and suppose that

$$v = c_1 b_1 + \dots + c_n b_n.$$

Then,

$$0_{V^*} = \theta_{\mathcal{B}}(v) = c_1 b_1^* + \dots + c_n b_n^* \in V^*.$$

Hence, since this is an equality of morphisms, we see that evaluating both sides of this equality at b_i, and using the defintion of b_k^*, we have

$$0 = 0_{V^*}(b_i) = (c_1 b_1^* + ... + c_n b_n^*)(b_i) = c_1 b_1^*(b_i) + ... + c_n b_n^*(b_i) = c_i, \quad \text{for every } i,$$

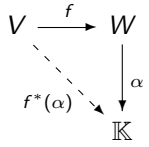
so that $c_1 = ... = c_n = 0 \in \mathbb{K}$. Hence, $v = 0$ and the result follows. □

Definition 1.8.4. Let $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ be a linear morphism between $\mathbb{K}$-vector spaces V, W. Then, we define the *dual of* f, denoted f^*, to be the function

$$f^* : W^* \to V^* \; ; \; \alpha \mapsto f^*(\alpha) = \alpha \circ f.$$

Remark 1.8.5.

1. Let's clarify just exactly what f^* is, for a given $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$: we have defined f^* as a function whose inputs are linear morphisms $\alpha : W \to \mathbb{K}$ and whose output is the linear morphism
$$f^*(\alpha) = \alpha \circ f : V \to W \to \mathbb{K} \; ; \; v \mapsto \alpha(f(v)).$$
Since the composition of linear morphisms is again a linear morphism we see that f^* is a well-defined function
$$f^* : W^* \to V^*.$$
We say that f^* **pulls back forms on** W **to forms on** V**.** Moreover, the function f^* is actually a linear morphism, so that $f^* \in \mathrm{Hom}_{\mathbb{K}}(W^*, V^*)$.[33]

2. Dual spaces/morphisms can be very confusing at first. It might help you to remember the following diagram


It now becomes clear why we say that f^* pulls back forms on W to forms on V.

3. The $(-)^*$ operation satisfies the following properties, which can be easily checked:
   - for $f, g \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ we have $(f+g)^* = f^* + g^* \in \mathrm{Hom}_{\mathbb{K}}(W^*, V^*)$; for $\lambda \in \mathbb{K}$ we have $(\lambda f)^* = \lambda f^* \in \mathrm{Hom}_{\mathbb{K}}(W^*, V^*)$,
   - if $f \in \mathrm{Hom}_{\mathbb{K}}(V, W), g \in \mathrm{Hom}_{\mathbb{K}}(W, X)$, then $(g \circ f)^* = f^* \circ g^* \in \mathrm{Hom}_{\mathbb{K}}(X^*, V^*)$; $\mathrm{id}_V^* = \mathrm{id}_{V^*} \in \mathrm{End}_{\mathbb{K}}(V^*)$.

4. Let $\mathcal{B} = (b_1, ..., b_n) \subset V, \mathcal{C} = (c_1, ..., c_m) \subset W$ be ordered bases and $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$. Let $\mathcal{B}^*, \mathcal{C}^*$ be the dual bases of $\mathcal{B}$ and $\mathcal{C}$. Then, we have that the matrix of f^* with respect to $\mathcal{C}^*, \mathcal{B}^*$ is
$$[f^*]_{\mathcal{C}^*}^{\mathcal{B}^*} = [[f^*(c_1^*)]_{\mathcal{B}^*} \; \cdots \; [f^*(c_m^*)]_{\mathcal{B}^*}],$$
an $n \times m$ matrix.

   Now, for each i, we have
$$f^*(c_i^*) = \lambda_{1i} b_1^* + ... + \lambda_{ni} b_n^*,$$
   so that
$$[f^*(c_i^*)]_{\mathcal{B}^*} = \begin{bmatrix} \lambda_{1i} \\ \vdots \\ \lambda_{ni} \end{bmatrix}, \quad \text{and} \quad \lambda_{ki} = f^*(c_i^*)(b_k) = c_i^*(f(b_k)).$$
   As $c_i^*(f(b_k))$ is the i^{th} entry in the $\mathcal{C}$-coordinate vector of $f(b_k)$, we see that the ik^{th} entry of $[f]_{\mathcal{B}}^{\mathcal{C}}$ is λ_{ki}, which is the ki^{th} entry of $[f^*]_{\mathcal{C}^*}^{\mathcal{B}^*}$. Hence, we have, if $A = [f]_{\mathcal{B}}^{\mathcal{C}}$, then
$$[f^*]_{\mathcal{C}^*}^{\mathcal{B}^*} = A^t.$$

[33] Check this.

Lemma 1.8.6. *Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$. Then,*

- *f is injective if and only if f^* is surjective.*
- *f is surjective if and only if f^* is injective.*
- *f is bijective if and only if f^* is bijective.*

Proof: The last statement is a consequence of the first two.

($\Rightarrow$) Suppose that f is injective, so that $\ker f = \{0_V\}$. Then, let $\beta \in V^*$ be a linear form on V, we want to find a linear form α on W such that $f^*(\beta) = \alpha$. Let $\mathcal{B} = (b_1, \dots, b_n)$ be an ordered basis of V, $\mathcal{B}^*$ the dual basis of V^*. Then, since f is injective, we must have that $f(\mathcal{B}) = (f(b_1), \dots, f(b_n))$ is a linearly independent subset of W[34]. Extend this to a basis $\mathcal{C} = (f(b_1, \dots, f(b_n), c_{n+1}, \dots, c_m)$ of W and consider the dual basis $\mathcal{C}^*$ of W^*.

In terms of the dual basis $\mathcal{B}^*$ we have

$$\beta = \lambda_1 b_1^* + \dots + \lambda_n b_n^* \in V^*.$$

Consider

$$\alpha = \lambda_1 f(b_1)^* + \dots \lambda_n f(b_n)^* + 0c_{n+1}^* + \dots + 0c_m^* \in W^*.$$

Then, we claim that $f^*(\alpha) = \beta$. To show this we must show that $f^*(\alpha)(v) = \beta(v)$, for every $v \in V$ (since $f^*(\beta), \alpha$ are both functions with domain V). We use the result (proved in Short Homework 4): if $f(b_i) = g(b_i)$, for each i, with f, g linear morphisms with domain V, then $f = g$. So, we see that

$$\begin{aligned} f^*(\alpha)(b_i) &= \lambda_1 f^*(f(b_1)^*)(b_i) + \dots + \lambda_n f^*(f(b_n)^*)(b_i) + 0_V, \quad \text{using linearity of } f^*, \\ &= \lambda_1 f(b_1)^*(f(b_i)) + \dots + \lambda_n f(b_n)^*(f(b_i)) = \lambda_i, \quad \text{since } f^* \text{ pulls back forms.} \end{aligned}$$

Then, it is easy to see that $\beta(b_i) = \lambda_i$, for each i. Hence, we must have $f^*(\alpha) = \beta$.

The remaining properties are left to the reader. In each case you will necessarily have to use some bases of V and W and their dual bases. □

Remark 1.8.7.

1. Lemma 1.8.6 allows us to try and show properties of a morphism by showing the 'dual' property of its dual morphism. You will notice in the proof that we had to make a choice of a basis of V and W and that this choice was arbitrary: for a general $\mathbb{K}$-vector space there is no 'canonical' choice of a basis. In fact, every proof of Lemma 1.8.6 must make use of a basis - there is no way that we can obtain these results without choosing a basis at some point. This is slightly annoying as this means there is no 'God-given' way to prove these statements, all such attempts must use some arbitrary choice of a basis.
2. Lemma 1.8.6 does **NOT** hold for infinite dimensional vector spaces. In fact, in the infinite dimensional case it is not true that V is isomorphic to V^*: the best we can do is show that there is an injective morphism $V \to V^*$. This a subtle and often forgotten fact.

In light of these remarks you should start to think that the passage from a vector space to its dual can cause problems because there is no 'God-given' way to choose a basis of V. However, these problems disappear if we *dualise twice*.

Theorem 1.8.8. *Let V be a finite dimensional $\mathbb{K}$-vector space. Then, there is a 'canonical isomorphism'*

$$\mathrm{ev} : V \to (V^*)^* \; ; \; v \mapsto (\mathrm{ev}_v : \alpha \mapsto \alpha(v))$$

Before we give the proof of this fact we will make clear what the function ev does: since V^* is a $\mathbb{K}$-vector space we can take its dual, to obtain $(V^*)^* \stackrel{def}{=} V^{**}$. Therefore, ev_v must be a linear form on V^*, so must take 'linear forms on V' as inputs, and give an output which is some value in $\mathbb{K}$. Given the linear form $\alpha \in V^*$, the output of $\mathrm{ev}_v(\alpha)$ is $\alpha(v)$, so we are 'evaluating α at v'.

[34] You have already showed this in Short Homework 3.

The reason we say that this isomorphism is 'canonical' is due to the fact that we did not need to use a basis to define ev - the same function ev works for *any* vector space V, so can be thought of as being 'God-given' or 'canonical' (there is no arbitrariness creeping in here).

Proof: ev is injective: suppose that $\mathrm{ev}_v = 0_{V^{**}}$, so that ev_v is the zero linear form on V^*. If $v \neq 0_V$ then we can extend the (linearly independent) set $\{v\}$ to a basis of V (simply take a maximal linearly independent subset of V that contains v). Then, $v^* \in V^*$ is the linear form that 'picks out the v-coefficient' of an arbitrary vector $u \in V$ when written as a linear combination using the basis containing v. Then, we must have

$$0 = ev_v(v^*) = v^*(v) = 1,$$

which is absurd. Hence, we can't have that $v \neq 0_V$ so that ev is injective.

Hence, since $\dim V = \dim V^* = \dim V^{**}$ we see that ev is an isomorphism. □

1.8.1 Coordinate-free systems of equations or Why row-reduction works

We know that a system of m linear equations in n variables is the same thing as a matrix equation

$$A\underline{x} = \underline{b},$$

where A is the coefficent matrix of the system and $\underline{x}$ is the vector of variables. We are going to try and consider systems of linear equations using linear forms.

Let $\mathcal{S}^{(n)} = (e_1 \dots, e_n)$ be the standard basis of $\mathbb{K}^n$, $\mathcal{S}^{(n),*}$ the dual basis. Then, if α is a linear form on $\mathbb{K}^n$ we see that

$$\alpha = \lambda_1 e_1^* + \dots + \lambda_n e_n^*.$$

Hence, if $\underline{x} = x_1 e_1 + \dots + x_n e_n \in \mathbb{K}^n$ then

$$\alpha(\underline{x}) = 0 \;\Leftrightarrow\; \lambda_1 x_1 + \dots + \lambda_n x_n = 0.$$

Hence, $\ker \alpha = \{\underline{x} \in \mathbb{K}^n \mid \lambda_1 x_1 + \dots + \lambda_n x_n = 0\}$.

Now, suppose that $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{K}^{n\,*}$ are linear forms. Then,

$$\bigcap_{i=1}^{m} \ker \alpha_i = \{\underline{x} \in \mathbb{K}^n \mid \alpha_i(\underline{x}) = 0, \text{ for every } i\}.$$

So, if $\alpha_i = \lambda_{i1} e_1^* + \dots + \lambda_{in} e_n^*$, then

$$\bigcap_{i=1}^{m} \ker \alpha_i = \left\{ \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n \mid \begin{array}{rl} \lambda_{11} x_1 + \dots + \lambda_{1n} x_n & = 0 \\ \vdots & \vdots \\ \lambda_{m1} x_1 + \dots + \lambda_{mn} x_n & = 0 \end{array} \right\}.$$

This is precisely the solution set of the matrix equation

$$A\underline{x} = \underline{0},$$

where $A = [\lambda_{ij}]$. Hence, we have translated our 'systems of linear equations' problem into one involving linear forms: namely, we want to try and understand $\bigcap_i \ker \alpha_i$, for some linear forms $\alpha_i \in \mathbb{K}^{n\,*}$.

Now, how can we interpret elementary row operations in this new framework? Of course, swapping rows is the same as just reordering the forms α_i. What happens if we scale a row by $\lambda \in \mathbb{K}$? This is the same as considering the linear form $\lambda\alpha \in \mathbb{K}^{n\,*}$. Similarly, adding row i to row j is the same as adding α_i to α_j to obtain the linear form $\alpha_i + \alpha_j$. In summary, performing elementary row operations is the same as forming linear combinations of the linear forms α_i.

The whole reason we row-reduce a matrix A to a reduced echelon form U is because the solution sets of $A\underline{x} = \underline{0}$ and $U\underline{x} = \underline{0}$ are the same (a fact we will prove shortly), and it is easier to determine solutions for the matrix equation defined by U. Since we obtain U by applying elementary row operations to A, this

is the same as doing calculations in $\text{span}_{\mathbb{K}}\{\alpha_i\} \subset \mathbb{K}^{n\,*}$, by what we have discussed above. Moreover, since U is in reduced echelon form this means that the rows of U are linearly independent (this is easy to see, by the definition of reduced echelon form) and because elementary row operations correspond to forming linear combinations of linear forms, we have that the linear forms that correspond to the rows of U must form a basis of the subspace $\text{span}_{\mathbb{K}}\{\alpha_i\} \subset \mathbb{K}^{n\,*}$.

Definition 1.8.9. Let V be a finite dimensional $\mathbb{K}$-vector space, V^* its dual space. Let $U \subset V$ be a subspace of V and $X \subset V^*$ a subspace of V^*. We define

$$\text{ann}_{V^*} U = \{\alpha \in V^* \mid \alpha(u) = 0, \text{ for every } u \in U\}, \text{ and}$$

$$\text{ann}_V X = \{v \in V \mid \text{ev}_v(\alpha) = 0, \text{ for every } \alpha \in X\},$$

the *annihilators of U (resp. X) in V^* (resp. V).* They are subspaces of V^* (resp. V), for any U (resp. X).

Proposition 1.8.10. *Let V be a $\mathbb{K}$-vector space, $U \subset V$ a subspace. Then,*

$$\dim V = \dim U + \dim \text{ann}_{V^*} U.$$

Proof: Let $\mathcal{A} = (a_1, \dots, a_k)$ be a basis of U and extend to a basis $\mathcal{B} = (a_1, \dots, a_k, a_{k+1}, \dots, a_n)$ of V. Then, it is easy to see that $a^*_{k+1}, \dots, a^*_n \in \text{ann}_{V^*} U$. Moreover, if $\alpha \in \text{ann}_{V^*} U$ then

$$\alpha = \lambda_1 a^*_1 + \dots + \lambda_n a^*_n,$$

and we must have, for every $i = 1, \dots, k$, that

$$0 = \alpha(a_i) = \lambda_i.$$

Hence, $\alpha \in \text{span}_{\mathbb{K}}\{a^*_{k+1}, \dots, a^*_n\}$ implying that $(a^*_{k+1}, \dots, a^*_n)$ is a basis of $\text{ann}_{V^*} U$. The result now follows. □

Corollary 1.8.11. *Let $f \in \text{Hom}_{\mathbb{K}}(V, W)$ be a linear morphism between finite dimensional vector spaces. Suppose that $A = [f]^{\mathcal{C}}_{\mathcal{B}} = [a_{ij}]$ is the matrix of f with respect to the bases $\mathcal{B} = (b_i) \subset V, \mathcal{C} = (c_j) \subset W$. Then,*

$$\text{rank } f = \text{ max. no. of linearly ind't columns of } A = \text{ max. no. of linearly ind't rows of } A.$$

Proof: The first equality was obtained in Lemma 1.7.14. The maximal number of linearly independent rows is equal to $\dim \text{span}_{\mathbb{K}}\{\alpha_i\}$, where

$$\alpha_i = a_{i1} b^*_1 + \dots + a_{in} b^*_n \in V^*.$$

Now, we have that

$$\text{ann}_V \text{span}_{\mathbb{K}}\{\alpha_i\} = \{v \in V \mid \text{ev}_v(\alpha_i) = 0, \text{ for every } i\} = \{v \in V \mid \alpha_i(v) = 0, \text{ for every } i\},$$

and this last set is nothing other than $\ker f$.[35] Thus, by the Rank Theorem (Theorem 1.7.13) we have

$$\dim V = \dim \text{im} f + \dim \ker f = \text{rank } f + \dim \text{ann}_V \text{span}_{\mathbb{K}}\{\alpha_i\} = \text{rank } f + (\dim V - \dim \text{span}_{\mathbb{K}}\{\alpha_i\}),$$

using Proposition 1.8.10 in the last equality. Hence, we find

$$\text{rank } f = \dim \text{span}_{\mathbb{K}}\{\alpha_i\},$$

which is what we wanted to show. □

Proposition 1.8.12 (Row-reduction works)**.** *Let $A \in Mat_{m,n}(\mathbb{K})$, U be its reduced echelon form. Then, $\underline{x}$ satisfies $U\underline{x} = \underline{0}$ if and only if $A\underline{x} = \underline{0}$.*

Proof: Let $\alpha_1, \dots, \alpha_m \in \mathbb{K}^{n\,*}$ be the linear forms corresponding to the rows of A, $\beta_1, \dots, \beta_r$ be the linear forms corresponding to the (nonzero) rows of U (we have just seen that $r = \text{rank } A$). Then, by our discussions above, we know (β_j) is a basis of $W = \text{span}_{\mathbb{K}}\{\alpha_i\}_{i=1}^m \subset \mathbb{K}^{n\,*}$. In particular, $\text{span}_{\mathbb{K}}\{\beta_j\} = W$. Now, we have

$$\text{ann}_{\mathbb{K}^n} W = \{\underline{x} \in \mathbb{K}^n \mid \alpha_i(\underline{x}) = 0, \text{ for every } i\} = \{\underline{x} \in \mathbb{K}^n \mid \beta_j(\underline{x}) = 0, \text{ for every } j\}.$$

The result follows from this equality of sets since this common set is the solution set of $A\underline{x} = \underline{0}$ and $U\underline{x} = \underline{0}$. □

[35] Think about this!

2 Jordan Canonical Form

In this chapter we are going to classify all $\mathbb{C}$-linear endomorphisms of a n-dimensional $\mathbb{C}$-vector space V. This means that we are going to be primarily studying $\mathrm{End}_{\mathbb{C}}(V)$, the $\mathbb{C}$-vector space of $\mathbb{C}$-endomorphisms of V (up to conjugation). For those of you that know about such things, we are going to identify the orbits of the group $\mathrm{GL}_{\mathbb{C}}(V)$ acting on the set $\mathrm{End}_{\mathbb{C}}(V)$ by conjugation. Since there exists an isomorphism

$$\mathrm{End}_{\mathbb{C}}(V) \to Mat_n(\mathbb{C}) \; ; \; f \mapsto [f]_{\mathcal{B}},$$

(once we choose an ordered basis $\mathcal{B}$ of V) this is the same thing as trying to classify all $n \times n$ matrices with $\mathbb{C}$-entries up to similarity.

You may recall that given any square matrix A with $\mathbb{C}$-entries we can ask whether A is *diagonalisable* and that there exists matrices that are not diagonalisable. For example, the matrix

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

is not diagonalisable.[36]

In fact, this example is typical, in the following sense: let $A \in Mat_2(\mathbb{C})$. Then, A is similar to one of the following types of matrices

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \; a, b \in \mathbb{C}, \quad \text{or} \quad \begin{bmatrix} c & 1 \\ 0 & c \end{bmatrix}, \; c \in \mathbb{C}.$$

In general, we have the following

Theorem (Jordan Canonical Form)**.** *Let $A \in Mat_n(\mathbb{C})$. Then, there exists $P \in \mathrm{GL}_n(\mathbb{C})$ such that*

$$P^{-1}AP = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & J_k \end{bmatrix},$$

where, for each $i = 1, \ldots, k$, we have an $n_i \times n_i$ matrix

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}, \quad \lambda_i \in \mathbb{C}.$$

Hence, every $n \times n$ matrix with $\mathbb{C}$-entries is similar to an *almost-diagonal* matrix.

We assume throughout this chapter that we are working with $\mathbb{C}$-vector spaces and $\mathbb{C}$-linear morphisms. Furthermore, we assume that all matrices have $\mathbb{C}$-entries.

2.1 Eigenthings

([1], p.108-113)

This section should be a refresher on the notions of *eigenvectors*, *eigenvalues* and *eigenspaces* of an $n \times n$ matrix A (equivalently, of a $\mathbb{C}$-linear morphism $f \in \mathrm{End}_{\mathbb{C}}(V)$).

Definition 2.1.1. Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ be a $\mathbb{C}$-linear endomorphism of the $\mathbb{C}$-vector space V. Let $\lambda \in \mathbb{C}$.

[36] Try and recall why this was true.

- The λ-*eigenspace of* f is the set

$$E_\lambda \stackrel{def}{=} \{v \in V \mid f(v) = \lambda v\}.$$

This is a vector subspace of V (possibly the zero subspace).

If $E_\lambda \neq \{0_V\}$ and $v \in E_\lambda$ is a nonzero vector, then we say that v is an *eigenvector of* f *with associated eigenvalue* λ.

- If A is an $n \times n$ matrix with $\mathbb{C}$-entries then we define the λ-*eigenspace of* A to be the λ-eigenspace of the linear morphism T_A. Similarly, we say that $v \in \mathbb{C}^n$ is an *eigenvector of* A *with associated eigenvalue* λ if v is an eigenvector of T_A with associated eigenvalue λ.

Lemma 2.1.2. *Let* $f \in \mathrm{End}_{\mathbb{C}}(V)$, $v_1, \ldots, v_k \in V$ *be eigenvectors of* f *with associated eigenvalues* $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$. *Assume that* $\lambda_i \neq \lambda_j$ *whenever* $i \neq j$. *Then,* $\{v_1, \ldots, v_k\}$ *is linearly independent.*

Proof: Let $S = \{v_1, \ldots, v_k\}$. Let $T \subset S$ denote a maximal linearly independent subset (we know that a linearly independent subset exists, just take $\{v_1\}$; then choose a linearly independent subset of largest size). We want to show that $T = S$. Suppose that $T \neq S$, we aim to provide a contradiction. As $T \neq S$, then we can assume, without loss of generality, that $v_k \notin T$.

We are going to show that $v_k \notin \mathrm{span}_{\mathbb{C}} T$, and then use Corollary 1.3.5 to deduce that $T \cup \{v_k\}$ is linearly independent, contradicting the maximality of T.

Suppose that $v_k \in \mathrm{span}_{\mathbb{C}} T$, we aim to provide a contradiction. So, as $v_k \in \mathrm{span}_{\mathbb{C}} T$ then

$$v_k = c_1 v_{i_1} + \ldots + c_s v_{i_s},$$

where $c_1, \ldots, c_s \in \mathbb{C}$, $v_{i_1}, \ldots, v_{i_s} \in T$. Apply f to both sides of this equation to obtain

$$\lambda_k v_k = c_1 \lambda_{i_1} v_{i_1} + \ldots + c_s \lambda_{i_s} v_{i_s}.$$

Taking this equation away from λ_k times the previous equation gives

$$0_V = c_1(\lambda_{i_1} - \lambda_k)v_{i_1} + \ldots + c_s(\lambda_{i_s} - \lambda_k)v_{i_s}.$$

This is a linear relation among vectors in T so must be the trivial linear relation since T is linearly independent. Hence, we have, for each $j = 1, \ldots, s$,

$$c_j(\lambda_{i_j} - \lambda_k) = 0,$$

and as $v_k \notin T$ (by assumption) we have $\lambda_{i_j} \neq \lambda_k$. Hence, we must have that $c_j = 0$, for every j. Then, we have $v_k = 0_V$, which is absurd as v_k is an eigenvector, hence nonzero by definition.

Therefore, our initial assumption that $v_k \in \mathrm{span}_{\mathbb{C}} T$ must be false, so that $v_k \notin \mathrm{span}_{\mathbb{C}} T$. As indicated above, this implies that $T \cup \{v_k\}$ is linearly independent, which contradicts the maximality of T. Therefore, T must be equal to S (otherwise $T \neq S$ and we run into the previous 'maximality' contradiction) so that S is linearly independent. □

Corollary 2.1.3. *Let* $\lambda_1, \ldots, \lambda_k$ *denote all eigenvalues of* $f \in \mathrm{End}_{\mathbb{C}}(V)$. *Then,*

$$E_{\lambda_1} + \ldots + E_{\lambda_k} = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k},$$

that is, the sum of all eigenspaces is a direct sum.

Proof: Left to the reader. □

Consider the case when

$$E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k} = V;$$

what does this tell us? In this case, we can find a basis of V consisting of eigenvectors of f (each λ_i-eigenspace E_{λ_i} is a subspace we can find a basis of it $\mathcal{B}_i$ say. Then, since we have in this case

$$\dim V = \dim E_{\lambda_1} + \ldots + \dim E_{\lambda_k},$$

we see that

$$\mathcal{B} = \mathcal{B}_1 \cup ... \cup \mathcal{B}_k,$$

is a basis of V.[37]) If we write $\mathcal{B} = (b_1, ..., b_n)$ (where $n = \dim V$) then we see that

$$[f]_{\mathcal{B}} = [[f(b_1)]_{\mathcal{B}} \ \cdots \ [f(b_n)]_{\mathcal{B}}],$$

and since $f(b_i)$ is a scalar multiple of b_i we see that $[f]_{\mathcal{B}}$ is a diagonal matrix.

Theorem 2.1.4. *Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ be such that*

$$E_{\lambda_1} \oplus ... \oplus E_{\lambda_k} = V.$$

Then, there exists a basis of V such that $[f]_{\mathcal{B}}$ is a diagonal matrix.

Corollary 2.1.5 (Diagonalisation)**.** *Let $A \in Mat_n(\mathbb{C})$ be such that there exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A. Then, there exists a matrix $P \in \mathrm{GL}_n(\mathbb{C})$ such that*

$$P^{-1}AP = D,$$

where D is a diagonal matrix. In fact, the entries on the diagonal of D are the eigenvalues of A.

Proof: Let $\mathcal{B} = (b_1, ..., b_n)$ be an ordered basis of $\mathbb{C}^n$ consisting of eigenvectors of A. Then, if $P = P_{\mathcal{S}^{(n)} \leftarrow \mathcal{B}}$ we have

$$P^{-1}AP = D,$$

by applying Corollary 1.7.7 to the morphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$. Here we note that $[T_A]_{\mathcal{B}} = D$ is a diagonal matrix by Theorem 2.1.4. □

Definition 2.1.6. We say that an endomorphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ is *diagonalisable* if there exists a basis $\mathcal{B} \subset V$ of V such that $[f]_{\mathcal{B}}$ is a diagonal matrix

We say that an $n \times n$ matrix $A \in Mat_n(\mathbb{C})$ is *diagonalisable* if T_A is diagonalisable. This is equivalent to: A is diagonalisable if and only if A is similar to a diagonal matrix (this is discussed in the following Remark).

Remark 2.1.7. Corollary 2.1.5 implies that if there exists a basis of $\mathbb{C}^n$ consisting of eigenvectors of A then A is diagonalisable. In fact, the converse is true: if A is diagonalisable and $P^{-1}AP = D$ then there is a basis of $\mathbb{C}^n$ consisting of eigenvectors of A. Indeed, if we let $\mathcal{B} = (b_1, ..., b_n)$ where b_i is the i^{th} column of P, then b_i is an eigenvector of A. Why does this hold? Since we have

$$P^{-1}AP = D = \mathrm{diag}(d_1, ..., d_n),$$

where $\mathrm{diag}(d_1, ..., d_n)$ denotes the diagonal $n \times n$ matrix with $d_1, ..., d_n$ on the diagonal, then we have

$$AP = PD.$$

Then, the i^{th} column of the matrix AP is Ab_i, so that $AP = PD$ implies that $Ab_i = d_i b_i$ (equate the columns of AP and PD). Therefore, each column of P is an eigenvector of A.

2.1.1 Characteristic polynomial, diagonalising matrices

Corollary 2.1.5 tells us conditions concerning when we can we can diagonalise a given matrix $A \in Mat_n(\mathbb{C})$ - we must find a basis of $\mathbb{C}^n$ consisting of eigenvectors of A. In order to this we need to determine how we can find *any* eigenvectors, let alone a basis consisting of eigenvectors.

Suppose that $v \in \mathbb{C}^n$ is an eigenvector of A with associated eigenvalue $\lambda \in \mathbb{C}$. This means we have

$$Av = \lambda v \implies (A - \lambda I_n)v = 0_{\mathbb{C}^n},$$

[37] Why must this be a basis?

that is, $v \in \ker T_{A-\lambda I_n}$. Conversely, any nonzero $v \in \ker T_{A-\lambda I_n}$ is an eigenvector of A with associated eigenvalue λ. Note that

$$E_\lambda = \ker T_{A-\lambda I_n}.$$

Since $T_{A-\lambda I_n} \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$ we know, by Proposition 1.7.4, that injectivity of $T_{A-\lambda I_n}$ is the same thing as bijectivity. Now, bijectivity of $T_{A-\lambda I_n}$ is the same thing as determining whether the matrix $A - \lambda I_n$ is invertible (using Theorem 1.7.4). Hence,

$$\ker T_{A-\lambda I_n} \neq \{0_{\mathbb{C}^n}\} \Leftrightarrow T_{A-\lambda I_n} \text{ not bijective } \Leftrightarrow \det(A - \lambda I_n) = 0.$$

Therefore, if v is an eigenvector of A with associated eigenvalue λ then we must have that $\det(A-\lambda I_n) = 0$ and $v \in \ker T_{A-\lambda I_n}$. Moreover, if $\lambda \in \mathbb{C}$ is such that $\det(A - \lambda I_n) = 0$ then $\ker T_{A-\lambda I_n} \neq \{0_{\mathbb{C}^n}\}$ and any nonzero $v \in \ker T_{A-\lambda I_n}$ is an eigenvector of A with associated eigenvalue λ.

Definition 2.1.8 (Characteristic polynomial)**.** Let $f \in \mathrm{End}_{\mathbb{C}}(V)$. Define the *characteristic polynomial of f*, denoted $\chi_f(\lambda)$, to be the polynomial in λ with complex coefficients

$$\chi_f(\lambda) = \det([f - \lambda \mathrm{id}_V]_{\mathcal{B}}),$$

where $\mathcal{B}$ is any ordered basis of V.[38]

If $A \in Mat_n(\mathbb{C})$ then we define the *characteristic polynomial of A*, denoted $\chi_A(\lambda)$, to be $\chi_{T_A}(\lambda)$. In this case, we have (using the standard basis $\mathcal{S}^{(n)}$ of $\mathbb{C}^n$)

$$\chi_A(\lambda) = \det(A - \lambda I_n).$$

Note that we are only considering λ as a 'variable' in the determinants, not an actual number. Also, note that the degree of $\chi_f(\lambda) = \dim V$[39] and the degree of $\chi_A(\lambda) = n$.

The *characteristic equation of f (resp. A)* is the equation

$$\chi_f(\lambda) = 0, \quad (\text{resp. } \chi_A(\lambda) = 0.)$$

Example 2.1.9. Let

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

Then,

$$A - \lambda I_2 = \begin{bmatrix} 1-\lambda & -3 \\ 2 & -1-\lambda \end{bmatrix}.$$

Hence, we have

$$\chi_A(\lambda) = (1-\lambda)(-1-\lambda) - 2.(-3) = \lambda^2 + 5.$$

Remark 2.1.10.

1. It should be apparent from the discussion above that the eigenvalues of a given linear morphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ (or matrix $A \in Mat_n(\mathbb{C})$) are precisely the zeros of the characteristic equation $\chi_f(\lambda) = 0$ (or $\chi_A(\lambda) = 0$).
2. Example 2.1.9 highlights an issue that can arise when we are trying to find eigenvalues of a linear morphism (or matrix). You'll notice that in this example there are **no $\mathbb{R}$-eigenvalues**: the eigenvalues are $\pm\sqrt{-5} \in \mathbb{C} \setminus \mathbb{R}$. Hence, we have complex eigenvalues that are not real. In general, given a matrix A with $\mathbb{C}$-entries (or a $\mathbb{C}$-linear morphism $f \in \mathrm{End}_{\mathbb{C}}(V)$) we will always be able to find eigenvalues - this follows from the **Fundamental Theorem of Algebra**:

[38] If $\mathcal{C}$ is any other basis of V then there is an invertible matrix P such that

$$[f - \lambda \mathrm{id}_V]_{\mathcal{C}} = P^{-1}[f - \lambda \mathrm{id}_V]_{\mathcal{B}} P.$$

Then, since $\det(AB) = \det A \det B$, for any matrices A, B, we see that $\det([f - \lambda \mathrm{id}_V]_{\mathcal{C}}) = \det([f - \lambda \mathrm{id}_V]_{\mathcal{B}})$ (where we have also used $\det P^{-1} = (\det P)^{-1}$).

[39] This will be shown in homework.

Theorem (Fundamental Theorem of Algebra)**.** *Let $p(T)$ be a nonconstant polynomial with $\mathbb{C}$-coefficients. Then, there exists $\lambda_0 \in \mathbb{C}$ such that $p(\lambda_0) = 0$. Hence, every such polynomial can be written as a product of linear factors*

$$p(T) = (T - \lambda_1)^{n_1}(T - \lambda_2)^{n_2} \cdots (T - \lambda_k)^{n_k}.$$

Note that this result is false if we wish to find a real root: for $p(T) = T^2 + 1$ there are no real roots (ie, no $\lambda_0 \in \mathbb{R}$ such that $p(\lambda_0) = 0$).

It is a consequence of this Theorem that we are considering in this section only $\mathbb{K} = \mathbb{C}$ as this guarantees that eigenvalues exist.

We are now in a position to find eigenvectors/eigenvalues of a given linear morphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ (or matrix $A \in Mat_n(\mathbb{C})$):

0. Find an ordered basis $\mathcal{B} = (b_1, \dots, b_n)$ of V to obtain $[f]_{\mathcal{B}}$. Let $A = [f]_{\mathcal{B}}$. **This step is not required if you are asked to find eigenthings for a given $A \in Mat_n(\mathbb{C})$.**
1. Determine the characteristic polynomial $\chi_A(\lambda)$ and solve the equation $\chi_A(\lambda) = 0$. The roots of this equation are the eigenvalues of A (and f), denote them $\lambda_1, \dots, \lambda_k$.
2. $v \in V$ is an eigenvector with associated eigenvalue λ_i if and only if $v \in \ker(f - \lambda_i \mathrm{id}_V)$ if and only if $[v]_{\mathcal{B}}$ is a solution to the matrix equation

$$(A - \lambda_i I_n)\underline{x} = \underline{0}.$$

Example 2.1.11. This follows on from Example 2.1.9 and we have already determined Step 1. above, we have

$$\lambda_1 = \sqrt{-5},\ \lambda_2 = -\sqrt{-5}.$$

If we wish to find eigenvectors with associated eigenvalue λ_1 then we consider the matrix

$$A - \lambda_1 I_2 = \begin{bmatrix} 1 - \sqrt{-5} & -3 \\ 2 & -1 - \sqrt{-5} \end{bmatrix} \sim \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix},$$

and so obtain that

$$\ker T_{A-\lambda_1 I_2} = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{C}^2 \mid x_1 - 3x_2 = 0 \right\} = \left\{ \begin{bmatrix} 3x \\ x \end{bmatrix} \mid x \in \mathbb{C} \right\}.$$

In particular, if we choose $x = 1$, we see that $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ is an eigenvector of A with associated eigenvalue $\sqrt{-5}$. Any eigenvector of A with associated eigenvalue $\sqrt{-5}$ is a nonzero vector in $\ker T_{A-\sqrt{-5}I_2}$.

Definition 2.1.12. Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ (or $A \in Mat_n(\mathbb{C})$). Supoose that

$$\chi_f(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}$$

so that $\lambda_1, \dots, \lambda_k$ are the eigenvalues of f.

- define the *algebraic multplicity of λ_i* to be n_i,
- define the *geometric multiplicity of λ_i* to be $\dim E_{\lambda_i}$.

Lemma 2.1.13. *Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ and λ be an eigenvalue of f. Then,*

'alg. multplicity of λ' $\geq$ 'geom. multiplicity of λ'

Proof: This will be proved later after we have introduced the polynomial algebra $\mathbb{C}[t]$ and the notion of a representation of $\mathbb{C}[t]$ (Definition 2.4.2) $\square$

Proposition 2.1.14. *Let $A \in Mat_n(\mathbb{C})$. Denote the eigenvalues of A by $\lambda_1, \dots, \lambda_k$. Then, A is diagonalisable if and only if, for every i, the algebraic multplicity of λ_i is equal to the geometric multiplicity of λ_i.*

Proof: ($\Rightarrow$) Suppose that A is diagonalisable and that

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}.$$

Then, by Remark 2.1.7, we can find a basis of eigenvectors of $\mathbb{C}^n$. Hence, we must have

$$\mathbb{C}^n = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}.$$

Then, by Lemma 2.1.13 and Corollary 1.5.19, we have

$$n = \dim E_{\lambda_1} + \dots + \dim E_{\lambda_k} \leq n_1 + \dots + n_k = n,$$

where we have used that the degree of the characteristic polynomial is n.
This implies that we must have $\dim E_{\lambda_i} = n_i$, for every i: indeed, we have

$$\dim E_{\lambda_1} + \dots \dim E_{\lambda_k} = n_1 + \dots + n_k,$$

with $\dim E_{\lambda_i} \leq n_i$, for each i. If $\dim E_{\lambda_i} < n_i$, for some i, then we would coontradict this previous equality. The result follows.

($\Leftarrow$) Assume that $\dim E_{\lambda_i} = n_i$, for every i. Then, we know that

$$V \supset E_{\lambda_1} + \dots + E_{\lambda_k} = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}.$$

Then, since

$$\dim(E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}) = \dim E_{\lambda_1} + \dots + \dim E_{\lambda_k} = n_1 + \dots + n_k = n,$$

we see that $V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$, by Corollary 1.5.17. Hence, there is a basis of V consisting of eigenvectors of A so that A is diagonalisable. □

As a consequence of Proposition 2.1.14 we are now in a position to determine (in practice) when a matrix A is diagonalisable. Following on from the above list to find eigenvectors we have

3. For each eigenvalue λ_i determine a basis of $\ker T_{A-\lambda_i I_n}$ (by row-reducing the matrix $A - \lambda_i I_n$ to reduced echelon form, for example). Denote this basis $\mathcal{B}_i = (b_1^{(i)}, \dots, b_{m_i}^{(i)})$.

4. If $|\mathcal{B}_i| = m_i = n_i$, for every i, then A is diagonalisable. Otherwise, A is not diagonalisable. Recall that in Step 1. above you will have determined $\chi_A(\lambda)$, and therefore n_i.

5. If A is diagonalisable then define the matrix P to be the $n \times n$ matrix

$$P = [b_1^{(1)} \cdots b_{n_1}^{(1)} b_1^{(2)} \cdots b_{n_2}^{(2)} \cdots b_1^{(k)} \cdots b_{n_k}^{(k)}].$$

   Then, Remark 2.1.7 implies that

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_1, \lambda_2, \dots, \lambda_2, \dots, \lambda_k, \dots, \lambda_k),$$

   with each eigenvalue λ_i appearing n_i times on the diagonal.

   Note that the order of the eigenvalues appearing on the diagonal depends on the ordering we put on $\mathcal{B}$.

Corollary 2.1.15. *Let $A \in Mat_n(\mathbb{C})$. Then, if A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$, then A is diagonalisable.*

Proof: Saying that A has n distinct eigenvalues is equivalent to saying that

$$\chi_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n),$$

so that the algebraic multiplicity n_i of each eigenvalue is 1. Furthermore, λ_i is an eigenvalue if and only if there exists a nonzero $v \in \mathbb{C}^n$ such that $Av = \lambda_i v$. Hence, we have

$$1 \leq \dim E_{\lambda_i} \leq n_i = 1,$$

by Lemma 2.1.13, so that $\dim E_{\lambda_i} = 1 = n_i$, for every i. Hence, A is diagonalisable by the previous Proposition. □

Example 2.1.16. Consider the matrix

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

from the previous examples. Then, we have $\chi_A(\lambda) = (\lambda - \sqrt{-5})(\lambda - (-\sqrt{-5}))$, so that Corollary 2.1.15 implies that A is diagonalisable.

In this section we have managed to obtain a useful criterion for when a given matrix A is diagonalisable. Moreover, this criterion is practically useful in that we have obtained a procedure that allows us to determine the diagonalisability of A by hand (or, at least, a criterion we could program a computer to undertake).

2.2 Invariant subspaces

([1], p.106-108)

In the proceeding sections we will be considering endomorphisms f of a $\mathbb{C}$-vector space V and some natural subspaces of V that we can associate to f. You may have seen some of these concepts before but perhaps not the terminology that we will adopt.

Definition 2.2.1 (Invariant subspace)**.** Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ be a linear endomorphism of V, $U \subset V$ a vector subspace of V. We say that U is *f-invariant* or *invariant with respect to f* if, for every $u \in U$ we have $f(u) \in U$.

If $A \in Mat_n(\mathbb{C})$, $U \subset \mathbb{C}^n$ a subspace, then we say that U is *A-invariant* or *invariant with respect to A* if U is T_A-invariant.

Example 2.2.2.

1. Any subspace $U \subset V$ is invariant with respect to $\mathrm{id}_V \in \mathrm{End}_{\mathbb{C}}(V)$. In fact, any subspace $U \subset V$ is invariant with respect to the endomorphism $c \cdot \mathrm{id}_V \in \mathrm{End}_{\mathbb{C}}(V)$, where

   $$(c \cdot \mathrm{id}_V)(v) = cv, \quad \text{for every } v \in V.$$

   In particular, every subspace is invariant with respect to the zero morphism of V.

2. Suppose that $V = U \oplus W$ and p_U, p_W are the projection morphisms introduced in Example 1.4.8. Then, U is p_U-invariant: let $u \in U$, we must show that $p_U(u) \in U$. Recall that if $v = u + w$ is the unique way of writing $v \in V$ as a linear combination of vectors in U and W (since $V = U \oplus W$), then

   $$p_U(v) = u, \quad p_W(v) = w.$$

   Hence, since $u \in V$ can be written as $u = u + 0_V$, then $p_U(u) = u \in U$, so that U is p_U-invariant. Also, if $w \in W$ then $w = 0_V + w$ (with $0_V \in U$), so that $p_U(w) = 0_V \in W$. Hence, W is also p_U-invariant. Similarly, we have U and W are both p_W-invariant.

   In general, if $V = U_1 \oplus \cdots \oplus U_k$, with $U_i \subset V$ a subspace, then each U_i is p_{U_j}-invariant, for any i, j.

3. Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ and suppose that λ is an eigenvalue of f. Then, E_λ is f-invariant: let $v \in E_\lambda$. Then, we have $f(v) = \lambda v \in E_\lambda$, since E_λ is a vector subspace of V.

Lemma 2.2.3. *Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ and $U \subset V$ an f-invariant subspace of V.*

- *Denote $f^k = f \circ f \circ \cdots \circ f$ (the k-fold composition of f on V) then U is also f^k-invariant.*
- *If U is also g-invariant, for some $g \in \mathrm{End}_{\mathbb{C}}(V)$, then U is $(f+g)$-invariant.*
- *If $\lambda \in \mathbb{C}$ then U is a λf-invariant subspace.*

Proof: Left to reader. □

Remark 2.2.4. It is important to note that the converse of the above statements in Lemma 2.2.3 do not hold.

For example, consider the matrix

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

and the associated endomorphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^2)$. Then, $T_A^2 = T_{A^2} = T_{I_2} = \mathrm{id}_{\mathbb{C}^2}$ (because $A^2 = I_2$), so that every subspace of $\mathbb{C}^2$ is A^2-invariant. However, the subspace $U = \mathrm{span}_{\mathbb{C}}(e_1)$ is not A-invariant since $Ae_1 = e_2$.

We can also see that $A + (-A) = 0_2$ so that every subspace of $\mathbb{C}^2$ is $(A + (-A))$-invariant, while $U = \mathrm{span}_{\mathbb{C}}(e_1)$ is neither A-invariant nor $(-A)$-invariant.

Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ and U be an f-invariant subspace. Suppose that $\mathcal{B}' = (b_1, \dots, b_k)$ is an ordered basis of U and extend to an ordered basis $\mathcal{B} = (b_1, \dots, b_k, b_{k+1}, \dots, b_n)$ of V. Then, the matrix of f relative to $\mathcal{B}$ is

$$[f]_{\mathcal{B}} = [[f(b_1)]_{\mathcal{B}} \ \cdots \ [f(b_k)]_{\mathcal{B}} \cdots \ [f(b_n)]_{\mathcal{B}}] = \begin{bmatrix} A & B \\ 0_{n-k,k} & C \end{bmatrix},$$

where $A \in Mat_k(\mathbb{C})$, $B \in Mat_{k,n-k}(\mathbb{C})$, $C \in Mat_{n-k,n-k}(\mathbb{C})$. This follows because $f(b_i) \in \mathrm{span}_{\mathbb{C}}\{b_1, \dots, b_k\}$, for each $i = 1, \dots, k$.

Moreover, we can see that if $V = U \oplus W$ with U *and* W *both* f*-invariant*, and if $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is an ordered basis of V, where $\mathcal{B}_1$ is an ordered basis of U, $\mathcal{B}_2$ is an ordered basis of W, then the matrix of f relative to $\mathcal{B}$ is

$$[f]_{\mathcal{B}} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where $A \in Mat_{\dim U}(\mathbb{C})$, $B \in Mat_{\dim W}(\mathbb{C})$.

Definition 2.2.5. Let $A \in Mat_n(\mathbb{C})$. We say that A is *block diagonal* if there are matrices $A_i \in Mat_{n_i}(\mathbb{C})$, $i = 1, \dots, k$, such that

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A_k \end{bmatrix}.$$

So, our previous discussion implies the following

Lemma 2.2.6. *Let $f \in \mathrm{End}_{\mathbb{C}}(V)$, $U_1, \dots, U_k \subset V$ subspaces of V that are all f-invariant and suppose that*

$$V = U_1 \oplus \cdots \oplus U_k.$$

Then, there exists an ordered basis $\mathcal{B}$ of V such that

$$[f]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A_k \end{bmatrix},$$

is a block diagonal matrix, with $A_i \in Mat_{\dim U_i}(\mathbb{C})$. In fact, we can assume that $\mathcal{B} = \mathcal{B}_1 \cup ... \cup \mathcal{B}_k$, with $\mathcal{B}_i$ an ordered basis of U_i, and that

$$A_i = [f_{|U_i}]_{\mathcal{B}_i},$$

where $f_{|U_i} : U_i \to U_i$ is the restriction of f to U_i.[40]

2.3 Nilpotent endomorphisms

([1], p.133-136)

In this section we will consider those linear endomorphisms $f \in \mathrm{End}_{\mathbb{C}}(V)$ whose only eigenvalue is 0. This necessarily implies that

$$\chi_f(\lambda) = \lambda^n.$$

We will see that for such endomorphisms there is a (ordered) basis $\mathcal{B}$ of V such that $[f]_{\mathcal{B}}$ is 'nearly diagonal'.

Definition 2.3.1. An endomorphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ is called *nilpotent* if there exists $r \in \mathbb{N}$ such that $f^r = 0_{\mathrm{End}_{\mathbb{C}}(V)}$, so that $f^r(v) = 0_V$, for every $v \in V$.

A matrix $A \in Mat_n(\mathbb{C})$ is called *nilpotent* if the endomorphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$ is nilpotent.

Lemma 2.3.2. *Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ be a nilpotent endomorphism. Then, the only eigenvalue of f is $\lambda = 0$ so that $\chi_f(\lambda) = \lambda^{\dim V}$.*

Proof: Suppose that $v \in V$ is an eigenvector of f with associated eigenvalue λ. Therefore, we have $v \neq 0$ and $f(v) = \lambda v$. Suppose that $f^r = 0$. Then,

$$0 = f^r(v) = f \circ \cdots \circ f(v) = f \circ \cdots \circ f(\lambda v) = \lambda^r v.$$

Thus, as $v \neq 0$ we must have $\lambda^r = 0$ (Proposition 1.2.5) implying that $\lambda = 0$. □

For a nilpotent endomorphism f (resp. matrix $A \in Mat_n(\mathbb{C})$) we define the *exponent of f (resp. of A)*, denoted $\eta(f)$ (resp. $\eta(A)$), to be the smallest $r \in \mathbb{N}$ such that $f^r = 0$ (resp. $A^r = 0$). Therefore, if $\eta(f) = r$ then there exists $v \in V$ such that $f^{r-1}(v) \neq 0_V$.

For $v \in V$ we define the *height of v (with respect to f)*, denoted $\mathrm{ht}(v)$, to be the smallest integer m such that $f^m(v) = 0_V$, while $f^{m-1}(v) \neq 0_V$. Hence, for every $v \in V$ we have $\mathrm{ht}(v) \leq \eta(f)$.

Define $H_k = \{v \in V \mid \mathrm{ht}(v) \leq k\}$, the set of vectors that have height no greater than k; this is a subspace of V.[41]

Let $f \in \mathrm{End}_{\mathbb{C}}(V)$ be a nilpotent endomorphism. Then, we obviously have $H_{\eta(f)} = V$, $H_0 = \{0_V\}$ and a sequence of subspaces

$$\{0_V\} = H_0 \subset H_1 \subset \cdots \subset H_{\eta(f)-1} \subset H_{\eta(f)} = V.$$

Let us denote

$$\dim H_i = m_i,$$

so that we have

$$0 = m_0 \leq m_1 \leq ... \leq m_{\eta(f)-1} \leq m_{\eta(f)} = \dim V.$$

We are going to construct a basis of V: for ease of notation we let $\eta(f) = k$. Assume that $k \neq 1$, so that f is not the zero endomorphism of V.

1. Let G_k be a *complementary subspace* of H_{k-1} so that

   $$H_k = H_{k-1} \oplus G_k,$$

   and let $(z_1, ..., z_{p_1})$ be an ordered basis of G_k. Then, since $z_j \in H_k \setminus H_{k-1}$ we have that $f^{k-1}(z_j) \neq 0_V$, for each j.

[40] This is a well-defined function since U_i is f-invariant.

[41] Exercise: show this.

2. Consider the vectors $f(z_1), f(z_2), \dots, f(z_{p_1})$. We have, for each j,

$$f^{k-1}(f(z_j)) = f^k(z_j) = 0_V, \quad \text{since } z_j \in H_k,$$

so that $f(z_j) \in H_{k-1}$, for each j. In addition, we can't have $f(z_j) \in H_{k-2}$, else

$$0_V = f^{k-2}(f(z_j)) = f^{k-1}(z_j),$$

implying that $z_j \in H_{k-1}$.

Moreover, the set $S_1 = \{f(z_1), f(z_2), \dots, f(z_{p_1})\} \subset H_{k-1} \setminus H_{k-2}$ is linearly independent: indeed, suppose that there is a linear relation

$$c_1 f(z_1) + \dots + c_{p_1} f(z_{p_1}) = 0_V.$$

with $c_1, \dots, c_{p_1} \in \mathbb{C}$. Then, since f is a linear morphism we obtain

$$f(c_1 z_1 + \dots + c_{p_1} z_{p_1}) = 0_V,$$

so that $c_1 z_1 + \dots + c_{p_1} z_{p_1} \in H_1 \subset H_{k-1}$.

Hence, we have $c_1 z_1 + \dots + c_{p_1} z_{p_1} \in H_{k-1} \cap G_k = \{0_V\}$, so that $c_1 z_1 + \dots + c_{p_1} z_{p_1} = 0_V$. Hence, because $\{z_1, \dots, z_{p_1}\}$ is linearly independent we must have $c_1 = \dots = c_{p_1} = 0 \in \mathbb{C}$. Thus, S_1 is linearly independent.

3. $\operatorname{span}_{\mathbb{C}} S_1 \cap H_{k-2} = \{0_V\}$: otherwise, we could find a linear combination

$$c_1 f(z_1) + \dots + c_{p_1} f(z_{p_1}) \in H_{k-2},$$

with some $c_i \neq 0$. Then, we would have

$$0_V = f^{k-2}(c_1 f(z_1) + \dots + c_{p_1} f(z_{p_1})) = f^{k-1}(c_1 z_1 + \dots + c_{p_1} z_{p_1}),$$

so that $c_1 z_1 + \dots + c_{p_1} z_{p_1} \in H_{k-1} \cap G_k = \{0_V\}$ which gives all $c_j = 0$, by linear independence of the z_j's. But this contradicts that some c_i is nonzero so that our initial assumption that $\operatorname{span}_{\mathbb{C}} S_1 \cap H_{k-2} \neq \{0_V\}$ is false.

Hence, we have

$$\operatorname{span}_{\mathbb{C}} S_1 + H_{k-2} = \operatorname{span}_{\mathbb{C}} S_1 \oplus H_{k-2} \subset H_{k-1}.$$

In particular, we see that $m_k - m_{k-1} \leq m_{k-1} - m_{k-2}$.

4. Let G_{k-1} be a complementary subspace of $H_{k-2} \oplus \operatorname{span}_{\mathbb{C}} S_1$ in H_{k-1}, so that

$$H_{k-1} = H_{k-2} \oplus \operatorname{span}_{\mathbb{C}} S_1 \oplus G_{k-1},$$

and let $(z_{p_1+1}, \dots, z_{p_2})$ be an ordered basis of G_{k-1}.

5. Consider the subset $S_2 = \{f^2(z_1), \dots, f^2(z_{p_1}), f(z_{p_1+1}), \dots, f(z_{p_2})\}$. Then, as in 2, 3, 4 above we have that

$$S_2 \subset H_{k-2} \setminus H_{k-3},$$

S_2 is linearly independent and $\operatorname{span}_{\mathbb{C}} S_2 \cap H_{k-3} = \{0_V\}$. Therefore, we have

$$\operatorname{span}_{\mathbb{C}} S_2 + H_{k-3} = \operatorname{span}_{\mathbb{C}} S_2 \oplus H_{k-3} \subset H_{k-2},$$

so that $m_{k-1} - m_{k-2} \leq m_{k-2} - m_{k-3}$.

6. Let G_{k-2} be a complementary subspace of $\text{span}_{\mathbb{C}} S_2 \oplus H_{k-3}$ in H_{k-2}, so that

$$H_{k-2} = H_{k-3} \oplus \text{span}_{\mathbb{C}} S_2 \oplus G_{k-2},$$

and $(z_{p_2+1}, \dots, z_{p_3})$ be an ordered basis of G_{k-2}.

7. Consider the subset $S_3 = \{f^3(z_1), \dots, f^3(z_{p_1}), f^2(z_{p_1+1}), \dots, f^2(z_{p_2}), f(z_{p_2+1}), \dots, f(z_{p_3})\}$. Again, it can be shown that

$$S_3 \subset H_{k-3} \setminus H_{k-4},$$

S_3 is linearly independent and $\text{span}_{\mathbb{C}} S_3 \cap H_{k-4} = \{0_V\}$. We obtain $m_{k-2} - m_{k-3} \leq m_{k-3} - m_{k-4}$.

8. Proceed in this fashion to obtain a basis of V. We denote the vectors we have obtained in a table

$$\begin{array}{cccccccccccc} z_1, & \dots & z_{p_1}, & & & & & & & & & \\ f(z_1), & \dots & f(z_{p_1}), & z_{p_1+1}, & \dots & z_{p_2}, & & & & & & \\ \vdots & & & & \vdots & & & & & & & \\ f^{k-1}(z_1), & \dots & f^{k-1}(z_{p_1}), & f^{k-2}(z_{p_1+1}), & \dots & f^{k-2}(z_{p_2}), & \dots & z_{p_{k-1}+1}, & \dots & z_{p_k}, \end{array} \tag{2.3.1}$$

where the vectors in the i^{th} row have height $k - i + 1$, so that vectors in the last row have height 1.

Also, note that each column determines an f-invariant subspace of V, namely the span of the vectors in the column.

Lemma 2.3.3. *Let W_i denote the span of the i^{th} column of vectors in the table above. Set $p_0 = 1$. Then,*

$$\dim W_i = k - j, \quad \text{if } p_j + 1 \leq i \leq p_{j+1}.$$

Proof: Suppose that $p_j + 1 \leq i \leq p_{j+1}$. Then, we have

$$W_i = \text{span}_{\mathbb{C}}\{z_i, f(z_i), \dots, f^{k-j-1}(z_i)\}.$$

Suppose that there exists a linear relation

$$c_0 z_i + c_1 f(z_i) + \dots + c_{k-j-1} f^{k-j-1}(z_i) = 0_V.$$

Then, applying f^{k-j-1} to both sides of this equation gives

$$c_0 f^{k-j-1}(z_i) + c_1 f^{k-j}(z_i) + \dots + c_{k-j-1} f^{2k-2j-2}(z_i) = 0_V.$$

Now, as z_i has height $k-j$ (this follows because the vector at the top of the i^{th} column is in the $(k-j)^{th}$ row, therefore as height $(k-j)$) the previous equation gives

$$c_0 f^{k-j-1}(z_i) + 0_V + \dots + 0_V = 0_V,$$

so that $c_0 = 0$, since $f^{k-j-1}(z_i) \neq 0_V$. Thus, we are left with a linear relation

$$c_1 f(z_i) + \dots + c_{k-j-1} f^{k-j-1}(z_i) = 0_V,$$

and applying f^{j-k-2} to this equation will give $c_1 = 0$, since $f(z_i)$ has height $k - j - 1$. Proceeding in this manner we find that $c_0 = c_1 = \dots c_{j-k-1} = 0$ and the result follows. □

Thus, the information recorded in (2.3.1) and Lemma 2.3.3 proves the following

Theorem 2.3.4. *Let $f \in \text{End}_{\mathbb{C}}(V)$ be a nilpotent endomorphism with exponent $\eta(f) = k$. Then, there exists integers $d_1, \dots, d_k \in \mathbb{Z}_{\geq 0}$ so that*

$$kd_1 + (k-1)d_2 + \dots + 2d_{k-1} + 1d_k = \dim V,$$

and f-invariant subspaces

$$W_1^{(k)}, \dots, W_{d_1}^{(k)}, W_1^{(k-1)}, \dots, W_{d_2}^{(k-1)}, \dots, W_1^{(1)}, \dots, W_{d_k}^{(1)} \subset V,$$

with $\dim_{\mathbb{C}} W_i^{(j)} = j$*, such that*

$$V = W_1^{(k)} \oplus \cdots \oplus W_{d_1}^{(k)} \oplus W_1^{(k-1)} \oplus \cdots \oplus W_{d_2}^{(k-1)} \oplus \cdots \oplus W_1^{(1)} \oplus \cdots \oplus W_{d_k}^{(1)}.$$

Moreover, there is an ordered basis $\mathcal{B}_i^{(j)}$ of $W_i^{(j)}$ such that

$$[f_{|W_i^{(j)}}]_{\mathcal{B}_i^{(j)}} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

We call such matrices 0-Jordan blocks*. Hence, we can write the matrix of f relative to $\mathcal{B} = \bigcup_{i,j} \mathcal{B}_i^{(j)}$ as a block diagonal matrix for which all of the blocks are* 0*-Jordan blocks and are of nonincreasing size as we move from left to right.*

Moreover, the geometric multiplicity of 0 *as an eigenvalue of f is equal to the number of blocks of the matrix $[f]_{\mathcal{B}}$ and this number equals the sum*

$$d_1 + d_2 + \dots + d_k = \dim E_0.$$

Proof: Everything except for the final statement follows from the construction of the basis $\mathcal{B}$ made prior to the Theorem.

The last statement is shown as follows: we have that $E_0 = H_1$, so that the 0-eigenspace of f consists of the set of all height 1 vectors in V.[42] Moreover, the construction of the basis $\mathcal{B}$ shows that a basis of H_1 is given by the bottom row of the table (2.3.1) and that this basis has the size specified. □

Corollary 2.3.5. *Let $A \in Mat_n(\mathbb{C})$ be a nilpotent matrix. Then, A is similar to a block diagonal matrix for which all of the blocks are* 0*-Jordan blocks.*

Proof: Consider the endomorphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$ and apply Theorem 2.3.4. Then, we have a basis $\mathcal{B}$ such that $[T_A]_{\mathcal{B}}$ takes the desired form. Now, use Corollary 1.7.7 and $[T_A]_{\mathcal{S}^{(n)}} = A$ to deduce the result. □

Definition 2.3.6. Let $n \in \mathbb{N}$. A *partition of n* is a decomposition of n into a sum of positive integers. If we have a partition of n

$$n = n_1 + \dots + n_l, \quad \text{with } n_1, \dots, n_l \in \mathbb{N}, \quad n_1 \leq n_2 \leq \dots \leq n_l,$$

then we denote this partition

$$1^{r_1} 2^{r_2} \cdots n_l^{r_{n_l}},$$

where we are assuming that 1 appears r_1 times in the partition of n, 2 appears r_2 times etc.

For example, consider the partition of 13

$$13 = 1 + 1 + 1 + 2 + 4 + 4,$$

then we denote this partition

$$1^3 2^1 4^2.$$

[42] Check this.

For a nilpotent endomorphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ we define its *nilpotent class* to be the set of all nilpotent endomorphisms g of V for which there is some ordered basis $\mathcal{C} \subset V$ with

$$[f]_{\mathcal{B}} = [g]_{\mathcal{C}},$$

where $\mathcal{B}$ is the basis described in Theorem 2.3.4.

We define the *partition associated to the nilpotent class of f*, denoted $\pi(A)$, to be the partition $1^{d_k}2^{d_{k-1}}\cdots k^{d_1}$ obtained in Theorem 2.3.4. We will also call this partition the *partition associated to f*.

For a matrix $A \in Mat_n(\mathbb{C})$ we define its nilpotent class (or *similarity class*) to be the nilpotent class of the endomorphism T_A. We define the *partition associated to A* to be the partition associated to T_A.

Theorem 2.3.7 (Classification of nilpotent endomorphisms)**.** *Let $f, g \in \mathrm{End}_{\mathbb{C}}(V)$ be nilpotent endomorphisms of V. Then, f and g lie in the same nilpotent class if and only if the partitions associated to f and g coincide.*

Corollary 2.3.8. *Let $A, B \in Mat_n(\mathbb{C})$ be nilpotent matrices. Then, f and g are similar if and only if the partitions associated to A and B coincide.*

Proof: We simply note that if T_A and T_B are in the same nilpotent class then there are bases $\mathcal{B}, \mathcal{C} \subset \mathbb{C}^n$ such that

$$[T_A]_{\mathcal{B}} = [T_B]_{\mathcal{C}}.$$

Hence, if $P_1 = P_{\mathcal{S}^{(n)} \leftarrow \mathcal{B}}$, $P_2 = P_{\mathcal{S}^{(n)} \leftarrow \mathcal{C}}$ then we must have

$$P_1^{-1} A P_1 = P_2^{-1} B P_2,$$

so that

$$P_2 P_1^{-1} A P_1 P_2^{-1} = B.$$

Now, since $P_2 P_1^{-1} = (P_1 P_2^{-1})^{-1}$ we have that A and B are similar precisely when T_A and T_B are in the same nilpotent class. The result follows. □

2.3.1 Determining partitions associated to nilpotent endomorphisms

Given a nilpotent endomorphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ (or nilpotent matrix $A \in Mat_n(\mathbb{C})$) how can we determine the partition associated to f (resp. A)?

Once we have chosen an ordered basis $\mathcal{B}$ of V we can consider the nilpotent matrix $[f]_{\mathcal{B}}$. Then, the problem of determining the partition associated to f reduces to determining the partition associated to $[f]_{\mathcal{B}}$. As such, we need only determine the partition associated to a nilpotent matrix $A \in Mat_n(\mathbb{C})$.

1. Determine the exponent of A, $\eta(A)$, by considering the products A^2, A^3, etc. The first r such that $A^r = 0$ is the exponent of A.
2. We can determine the subspaces H_i since

   $$H_i = \{\underline{x} \in \mathbb{C}^n \mid \mathrm{ht}(\underline{x}) \leq i\} = \ker T_{A^i}.$$

   In particular, we have that $\dim H_i$ is the number of non-pivot columns of A^i.
3. $d_1 = \dim H_{\eta(A)} - \dim H_{\eta(A)-1}$.
4. $d_2 = \dim H_{\eta(A)-1} - \dim H_{\eta(A)-2} - d_1$.
5. $d_3 = \dim H_{\eta(A)-2} - \dim H_{\eta(A)-3} - d_2$.
6. Thus, we can see that $d_i = \dim H_{\eta(A)-(i-1)} - \dim H_{\eta(A)-i} - d_{i-1}$, for $1 \leq i \leq \eta(A)$.

Hence, the partition associated to A is

$$\pi(A) : \ 1^{d_{\eta(A)}} 2^{d_{\eta(A)-1}} \cdots \eta(A)^{d_1}.$$

Example 2.3.9. Consider the endomorphism

$$f : \mathbb{C}^5 \to \mathbb{C}^5 \; ; \; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \mapsto \begin{bmatrix} x_2 \\ 0 \\ x_4 \\ 0 \\ 0 \end{bmatrix} .$$

Then, with respect to the standard basis $\mathcal{S}^{(5)}$ we have that

$$A \stackrel{def}{=} [f]_{\mathcal{S}^{(5)}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

You can check that $A^2 = 0$ so that $\eta(A) = 2$. Then,

- $d_1 = \dim H_2 - \dim H_1 = 5 - 3 = 2$, since $H_1 = \ker T_A$ has dimension 3 (there are 3 non-pivot columns of A).
- $d_2 = \dim H_1 - \dim H_0 - d_1 = 3 - 0 - 2 = 1$, since $H_0 = \{0\}$.

Hence, the partition associated to A is

$$\pi(A) : \; 12^2 \leftrightarrow 1 + 2 + 2 = 5;$$

there are three 0-Jordan blocks - two of size 2 and one of size 1.

You can check that the following matrix B is nilpotent

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

and that the partition associated to B is

$$\pi(B) : \; 1^3 2 \leftrightarrow 1 + 1 + 1 + 2 = 5$$

- We have $B^2 = 0$ so that $\eta(B) = 2$.
- $d_1 = \dim H_2 - \dim H_1 = 5 - 4 = 1$, since $H_1 = \ker T_B$ has dimension 4 (there are 4 non-pivot columns of B).
- $d_2 = \dim H_1 - \dim H_0 - d_1 = 4 - 0 - 1 = 3$, since $H_0 = \{0\}$.

Thus, A and B are not similar, by Corollary 2.3.8. However, since the matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} ,$$

has associated partition

$$\pi(C) : \; 1^3 2,$$

then we see that B is similar to C, by Corollary 2.3.8.

Moreover, there are four 0-Jordan blocks of B (and C) - one of size 2 and three of size 1.

2.4 Algebra of polynomials

([1], p.136-142)

In this section we will give a brief introduction to the algebraic properties of the polynomial algebra $\mathbb{C}[t]$. In particular, we will see that $\mathbb{C}[t]$ admits many similarities to the algebraic properties of the set of integers $\mathbb{Z}$.

Remark 2.4.1. Let us first recall some of the algebraic properties of the set of integers $\mathbb{Z}$.

- **division algorithm**: given two integers $w, z \in \mathbb{Z}$, with $|w| \leq |z|$, there exist $a, r \in \mathbb{Z}$, with $0 \leq r < |w|$ such that
$$z = aw + r.$$
Moreover, the 'long division' process allows us to determine a, r. Here r is the 'remainder'.

- **prime factorisation**: for any $z \in \mathbb{Z}$ we can write
$$z = \pm p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s},$$
where p_i are prime numbers. Moreover, this expression is essentially unique - it is unique up to ordering of the primes appearing.

- **Euclidean algorithm**: given integers $w, z \in \mathbb{Z}$ there exists $a, b \in \mathbb{Z}$ such that
$$aw + bz = \gcd(w, z),$$
where $\gcd(w, z)$ is the 'greatest common divisor' of w and z. In particular, if w, z share no common prime factors then we can write
$$aw + bz = 1.$$
The Euclidean algorithm is a process by which we can determine a, b.

We will now introduce the polynomial algebra in one variable. This is simply the set of all polynomials with complex coefficients and where we make explicit the $\mathbb{C}$-vector space structure and the multiplicative structure that this set naturally exhibits.

Definition 2.4.2. - The $\mathbb{C}$-*algebra of polynomials in one variable*, is the quadruple $(\mathbb{C}[t], \alpha, \sigma, \mu)$[43] where $(\mathbb{C}[t], \alpha, \sigma)$ is the $\mathbb{C}$-vector space of polynomials in t with $\mathbb{C}$-coefficients defined in Example 1.2.6, and
$$\mu : \mathbb{C}[t] \times \mathbb{C}[t] \to \mathbb{C}[t] \; ; \; (f, g) \mapsto \mu(f, g),$$
is the 'multiplication' function.

So, if
$$f = a_0 + a_1 t + ... + a_n t^n, \; g = b_0 + b_1 + ... + b_m t^m \in \mathbb{C}[t],$$
with $m \leq n$ say, then
$$\mu(f, g) = c_0 + c_1 t + ... + c_{m+n} t^{m+n},$$
where
$$c_i = \sum_{\substack{j+k=i, \\ 0 \leq j \leq n, \\ 0 \leq k \leq m}} a_j b_k.$$

[43] This is a particular example of a more general algebraic object called a $\mathbb{C}$-*algebra*: a $\mathbb{C}$-algebra is a set A that is a $\mathbb{C}$-vector space and for which there is a well-defined commutative multiplication map that interacts with addition in a nice way - for example, distributivity, associativity hold. One usually also requires that a $\mathbb{C}$-algebra A has a *multiplicative identity*, namely an element e such that $f \cdot e = e \cdot f = f$, for every $f \in A$. It is common to denote this element by 1.

We write

$$\mu(f,g) = f \cdot g, \quad \text{or simply } fg.$$

μ is nothing more than the function defining the 'usual' multiplication of polynomials with $\mathbb{C}$-coefficients. In particular, for every $f, g \in \mathbb{C}[t]$ we have $fg = gf$.

We will write $\mathbb{C}[t]$ instead of the quadruple above when discussing $\mathbb{C}[t]$ as a $\mathbb{C}$-algebra. Note that the polynomial $1 \in \mathbb{C}[t]$ satisfies the property that $1 \cdot f = f \cdot 1 = f$, for every $f \in \mathbb{C}[t]$.

- A *representation of* $\mathbb{C}[t]$ is a $\mathbb{C}$-linear morphism

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

for some finite dimensional $\mathbb{C}$-vector space V, such that

$$(*) \quad \rho(fg) = \rho(f) \circ \rho(g), \quad \text{and} \quad \rho(1) = \mathrm{id}_V,$$

where we are considering composition of linear endomorphisms of V on the RHS of the first equality.[44]

Remark 2.4.3. Suppose that

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

is a representation of $\mathbb{C}[t]$. Then, for any $f = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n \in \mathbb{C}[t]$, we have

$$\begin{aligned} \rho(f) = \rho(a_0 + a_1 t + ... + a_n t^n) &= a_0\rho(1) + a_1\rho(t) + ... + a_n\rho(t^n), \quad \text{as } \rho \text{ is } \mathbb{C}\text{-linear}, \\ &= a_0\mathrm{id}_V + a_1\rho(t) + a_2\rho(t)^2 + ... + a_n\rho(t)^n, \quad \text{by } (*), \end{aligned}$$

where we have written $\rho(t)^k = \rho(t) \circ \cdots \circ \rho(t)$, the k-fold composition of $\rho(t)$.

Hence, a represention of $\mathbb{C}[t]$ is the same thing as specifying a $\mathbb{C}$-linear endomorphism $\rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$: the multiplicative property of ρ then implies that $\rho(f)$ only depends on $\rho(t)$, for any $f \in \mathbb{C}[t]$.

Conversely, given a $\mathbb{C}$-linear endomorphism of V, $L \in \mathrm{End}_{\mathbb{C}}(V)$ say, then we can define a representation ρ_L of $\mathbb{C}[t]$ as follows: define

$$\rho_L : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V) \; ; \; a_0 + a_1 t + ... + a_n t^n \mapsto a_0\mathrm{id}_v + a_1 L + ... + a_n L^n \in \mathrm{End}_{\mathbb{C}}(V),$$

where $L^k = L \circ \cdots \circ L$ and the addition and scalar multiplication on the RHS is occuring in $\mathrm{End}_{\mathbb{C}}(V)$.

We are going to study an endomorphism $L \in \mathrm{End}_{\mathbb{C}}(V)$ by studying the representation ρ_L of $\mathbb{C}[t]$ it defines. If $A \in Mat_n(\mathbb{C})$ then we define ρ_A to be the representation defined by the endomorphism T_A of $\mathbb{C}^n$.

Suppose we are given a representation of $\mathbb{C}[t]$

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

and denote $n = \dim_{\mathbb{C}} V$, $L = \rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$ (so that $\rho = \rho_L$) and suppose that $L \neq \mathrm{id}_V$.[45]

We know that $\mathrm{End}_{\mathbb{C}}(V)$ is n^2-dimensional (since we know that $\mathrm{End}_{\mathbb{C}}(V)$ is isomorphic to $Mat_n(\mathbb{C})$). Therefore, there must exist a nontrivial linear relation

$$\lambda_0\mathrm{id}_V + \lambda_1 L + \lambda_2 L^2 + ... + \lambda_{n^2} L^{n^2} = 0_{\mathrm{End}_{\mathbb{C}}(V)},$$

[44] This means that ρ is a morphism of (unital) $\mathbb{C}$-algebras.

[45] If $L = \mathrm{id}_V$ then we call the representation ρ_{id_V} the *trivial representation*. In this case, we have that

$$\mathrm{im}\rho = \{c \cdot \mathrm{id}_V \in \mathrm{End}_{\mathbb{C}}(V) \mid c \in \mathbb{C}\} \subset \mathrm{End}_{\mathbb{C}}(V).$$

with $\lambda_i \in \mathbb{C}$, since the set $\{\mathrm{id}_V, L, L^2, \dots, L^{n^2}\}$ contains n^2+1 vectors. Thus, we have

$$\begin{aligned} 0_{\mathrm{End}_{\mathbb{C}}(V)} &= \lambda_0 \mathrm{id}_V + \lambda_1 L + \lambda_2 L^2 + \dots + \lambda_{n^2} L^{n^2} \\ &= \lambda_0 \rho(1) + \lambda_1 \rho(t) + \dots + \lambda_{n^2} \rho(t)^{n^2} \\ &= \rho(\lambda_0 + \lambda_1 t + \dots + \lambda_{n^2} t^{n^2}), \end{aligned}$$

so that the polynomial

$$f = \lambda_0 + \lambda_1 t + \dots + \lambda_{n^2} t^{n^2} \in \ker \rho.$$

In particular, we have that $\ker \rho \neq \{0_{\mathbb{C}[t]}\}$. We will now make a detailed study of the kernel of representations of $\mathbb{C}[t]$.

Keep the same notation as above. We have just seen that $\ker \rho$ is nonzero. Let $m_L \in \ker \rho$ be a nonzero polynomial for which $\rho(m_L) = 0_{\mathrm{End}_{\mathbb{C}}(V)}$ and such that m_L has minimal degree.[46] We must have $\deg m_L \neq 0$, otherwise m_L is a constant polynomial, say $m_L = c \cdot 1$ with $c \in \mathbb{C}$, $c \neq 0$, and $\rho(c \cdot 1) = c\rho(1) = c\mathrm{id}_V \neq 0_{\mathrm{End}_{\mathbb{C}}(V)}$, contradicting that $m_L \in \ker \rho$. Hence, we can assume that $\deg m_L = m > 0$.

Now, let $f \in \ker \rho$ be any other polynomial in the kernel of ρ. Denote $\deg f = p$. Thus, by our choice of m_L (it must have minimal degree) we see that $p \geq m$. Now use the division algorithm for polynomials[47] to find polynomials $g, h \in \mathbb{C}[t]$ such that

$$f = g m_L + h,$$

where $\deg h < m$.

Then, as $f \in \ker \rho$, we must have

$$0_{\mathrm{End}_{\mathbb{C}}(V)} = \rho(f) = \rho(g m_L + h) = \rho(g)\rho(m_L) + \rho(h) = 0_{\mathrm{End}_{\mathbb{C}}(V)} + \rho(h),$$

so that $h \in \ker \rho$. If h were a nonzero polynomial then we have obtained an element in $\ker \rho$ that has strictly smaller degree that m_L, contradicting our choice of m_L. Hence, we must have that $h = 0$ and $f = g m_L$. We say that m_L *divides* f.

We have just shown the following

Proposition 2.4.4. *Suppose that*

$$\rho : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

is a representation of $\mathbb{C}[t]$. *Denote* $L = \rho(t) \in \mathrm{End}_{\mathbb{C}}(V)$ *and suppose that* $m_L \in \ker \rho$ *is nonzero and has minimal degree. Then, for any* $f \in \ker \rho$ *there exists* $g \in \mathbb{C}[t]$ *such that*

$$f = g m_L.$$

Remark 2.4.5. Proposition 2.4.4 is stating the fact that the $\mathbb{C}$-algebra $\mathbb{C}[t]$ is a *principal ideal domain*, namely, every ideal in $\mathbb{C}[t]$ is generated by a single polynomial (ie, 'principal').

Definition 2.4.6. Let $L \in \mathrm{End}_{\mathbb{C}}(V)$ and consider the representation

$$\rho_L : \mathbb{C}[t] \to \mathrm{End}_{\mathbb{C}}(V),$$

defined above. We define the *minimal polynomial of* L, denoted $\mu_L \in \mathbb{C}[t]$, to be the unique nonzero polynomial $\mu_L \in \ker \rho$ that has minimal degree and has leading coefficient 1: this means that

$$\mu_L = a_0 + a_1 t + \dots + a_{m-1} t^{m-1} + t^m.$$

[46] Recall that the degree, $\deg f$, of a polynomial

$$f = a_0 + a_1 t + \dots + a_k t^k \in \mathbb{C}[t],$$

is defined to be $\deg f = k$. We have the property that

$$\deg fg = \deg f + \deg g.$$

[47] If you have not seen this before, don't worry, as I will cover this in class.

This polynomial is well-defined (ie, it's unique) by Proposition 2.4.4: if $m_L \in \ker \rho$ has minimal degree and leading coefficient $a \in \mathbb{C}$, then we have $\mu_L = a^{-1} m_L$. If $f \in \ker \rho$ is any other polynomial of minimal degree and with leading coefficient 1, then we must have $\deg f = \deg \mu_L$ and, by Proposition 2.4.4, we know that there exists $g \in \mathbb{C}[t]$ such that

$$f = g\mu_L.$$

Since $\deg f = \deg(g\mu_L) = \deg g + \deg \mu_L$ we must have that $\deg g = 0$, so that $g = c \cdot 1 \in \mathbb{C}[t]$. As both f and μ_L have leading coefficient 1, the only way this can hold is if $c = 1$, so that $f = \mu_L$.

For $A \in Mat_n(\mathbb{C})$ we write μ_A instead of μ_{T_A} and call it the *minimal polynomial of A*.

Corollary 2.4.7. *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$, μ_L be the minimal polynomial of L. For $f = a_0 + a_1 t + ... + a_k t^k \in \mathbb{C}[t]$ we denote*

$$f(L) = \rho_L(f) = a_0 \mathrm{id}_V + a_1 L + ... + a_k L^k \in \mathrm{End}_{\mathbb{C}}(V).$$

If $f(L) = 0_{\mathrm{End}_{\mathbb{C}}(V)}$ then $f = \mu_L g$, for some $g \in \mathbb{C}[t]$.

Proof: This is simply a restatement of Proposition 2.4.4. □

Example 2.4.8. 1. Consider the endomorphism T_A of $\mathbb{C}^3$ defined by the matrix

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

Then, you can check that the following relation holds

$$-A^3 + 2A^2 - A + 2I_3 = 0_3.$$

Consider the representation ρ_A defined by A. Then, since the above relation holds we must have

$$f = -\lambda^3 + 2\lambda^2 - \lambda + 2 \in \ker \rho_A.$$

You can check that we can decompose f as

$$f = (2 - \lambda)(\lambda - \sqrt{-1})(\lambda + \sqrt{-1}).$$

Hence, we must have that μ_A is one of the following polynomials[48]

$$(\lambda - \sqrt{-1})(\lambda + \sqrt{-1}),\ (2 - \lambda)(\lambda - \sqrt{-1}),\ (2 - \lambda)(\lambda + \sqrt{-1}),\ f.$$

In fact, we have $\mu_A = f$.

You may have noticed that $f = \chi_A(\lambda)$ - this is the *Cayley-Hamilton Theorem* (to be proved later and in homework): if $A \in Mat_n(\mathbb{C})$ then $\chi_A(\lambda) \in \ker \rho_A$, so that $\chi_A(A) = 0$ (using the above notation from Corollary 2.4.7).

- Consider the matrix

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

You can check that we have the relation

$$-A^3 + 3A^2 - 3A + I_3 = 0_3,$$

so that

$$f = -\lambda^3 + 3\lambda^2 - 3\lambda + 1 = (1 - \lambda)^3 \in \ker \rho_A.$$

[48] Why can't we have μ_A be one of $(2 - \lambda)$, $(\lambda - \sqrt{-1})$, $(\lambda + \sqrt{-1})$?

Now, we see that we must have μ_A being one of the following polynomials[49]

$$(1-\lambda)^2,\ f.$$

It can be checked that

$$A^2 - 2A + I_3 = 0_3,$$

so that

$$\mu_A = (1-\lambda)^2.$$

You will notice that

$$\chi_A(\lambda) = (1-\lambda)^3.$$

In both of these examples you can see that the roots of the minimal poynomial of A are precisely the eigenvalues of A (possibly with some repeated multiplicity). In fact, this is always true: for a matrix A the roots of μ_A are precisely the eigenvalues of A. This will be proved in the next section.

Recall that a polynomial $f \in \mathbb{C}[t]$ can be written as a product of linear factors

$$f = a(t-c_1)^{n_1}(t-c_2)^{n_2}\cdots(t-c_k)^{n_k},$$

where $a, c_1, \ldots, c_k \in \mathbb{C},\ n_1, \ldots, n_k \in \mathbb{N}$.

This is the analogue in $\mathbb{C}[t]$ of the 'prime factorisation' property of $\mathbb{Z}$ mentioned at the beginning of this section: the 'primes' of $\mathbb{C}[t]$ are degree 1 polynomials.

Definition 2.4.9. We say that the (nonzero) polynomials $f_1, \ldots, f_p \in \mathbb{C}[t]$ are *relatively prime* if there is no common linear factor for all of the f_i.

Example 2.4.10. The polynomials $f = t^2+1$ and $g = t^2-1$ are relatively prime. Indeed, we have

$$f = t^2+1 = (t-\sqrt{-1})(t+\sqrt{-1}),\ g = (t-1)(t+1),$$

so that there is no common linear factor of either.

However, the polynomials g and $h = t^n - 1$ are <u>not</u> relatively prime as

$$h = t^n - 1 = (t-1)(t-\omega)(t-\omega^2)\cdots(t-\omega^{n-1}),$$

where $\omega = \cos(2\pi/n) + \sqrt{-1}\sin(2\pi/n) \in \mathbb{C}$. Hence, the linear factor $(t-1)$ appears in both g and h.

We now give another basic algebraic property of the $\mathbb{C}$-algebra $\mathbb{C}[t]$ whose proof you would usually see in Math 113. As such, we will not prove this result here although the proof is exactly the same as the corresponding result for $\mathbb{Z}$ (with the appropriate modifications): it involves the $\mathbb{C}[t]$-analogue of the 'Euclidean algorithm' for $\mathbb{Z}$.

Lemma 2.4.11. *Let $f_1, \ldots, f_p \in \mathbb{C}[t]$ be a collection of relatively prime polynomials. Then, there exists $g_1, \ldots, g_p \in \mathbb{C}[t]$ such that*

$$f_1g_1 + \ldots + f_pg_p = 1 \in \mathbb{C}[t].$$

Example 2.4.12.

1. The polynomials $f = t^2+1, g = t^2-1$ are relatively prime and
$$\frac{1}{2}(t^2+1) - \frac{1}{2}(t^2-1) = 1.$$
2. The polynomials $f = t^2+1, g = t^3-1$ are relatively prime and
$$\frac{1}{2}(t-1)(t^3-1) - \frac{1}{2}(t^2-t-1)(t^2+1) = 1.$$

[49] Why can't we have $1-\lambda$?

Remark 2.4.13. The mathematical reason that $\mathbb{Z}$ and $\mathbb{C}[t]$ obey the same algebraic properties is that they are both examples of *Euclidean domains*: these are commutative rings for which there exists a division algorithm, Euclidean algorithm and the notion of prime elements.

More specifically, a Euclidean domain is a commutative ring without zerodivisors for which there exists a well defined 'degree' function. As a consequence of the existence of the degree function the division algorithm and Euclidean algorithm hold. Moreover, it can be show that such commutative rings are principal ideal domains and are therefore unique factorisation domains: this means that the 'unique factorisation' property holds.

2.5 Canonical form of an endomorphism

([1], p.142-146)

Throughout this section we fix a linear endomorphism $L \in \mathrm{End}_{\mathbb{C}}(V)$, for some finite dimensional $\mathbb{C}$-vector space V. We denote $n = \dim_{\mathbb{C}} V$.

We recall the notation from Corollary 2.4.7: for $L \in \mathrm{End}_{\mathbb{C}}(V)$, $f = a_0 + a_1 t + ... + a_k t^k \in \mathbb{C}[t]$, we define the endomorphism

$$f(L) = \rho_L(f) = a_0 \mathrm{id}_V + a_1 L + a_2 L^2 + ... + a_k L^k \in \mathrm{End}_{\mathbb{C}}(V),$$

where $L^i = L \circ \cdots \circ L$ is the i-fold composition of the endomorphism L.

Definition 2.5.1. Any nonzero $f \in \ker \rho_L$ is called an *annihilating polynomial of L*.

In particular, the minimal polynomial μ_L of L is an annihilating polynomial of L.

The following theorem is the culmination of our discussion regarding polynomials and representations of the polynomial algebra. It allows us to use the minimal polynomial of L to decompose V into a direct sum of L-invariant subspaces. Hence, we can find a basis of V for which the matrix of L with respect to this basis is block diagonal. We will then see that we can use our results on nilpotent endomorphisms to find a basis of V for which the matrix of L is 'almost diagonal' - this is the **Jordan canonical form** (Theorem 2.5.12).

Theorem 2.5.2. *Suppose that $f \in \ker \rho$ is an annihilating polynomial of L and that $f = f_1 f_2$, with f_1 and f_2 relatively prime. Then, we can write*

$$V = U_1 \oplus U_2,$$

with U_1 and U_2 both L-invariant (Definition 2.2.1), and such that

$$f_1(L)(u_2) = 0_V, \quad f_2(L)(u_1) = 0_V,$$

for every $u_1 \in U_1, u_2 \in U_2$.

Moreover,

$$U_1 = \ker f_2(L), \quad U_2 = \ker f_1(L).$$

Proof: As f_1 and f_2 are relatively prime we know that there exists $g_1, g_2 \in \mathbb{C}[t]$ such that

$$f_1 g_1 + f_2 g_2 = 1 \in \mathbb{C}[t].$$

This follows from Lemma 2.4.11. Hence, we have

$$\mathrm{id}_v = \rho_L(1) = \rho_L(f_1 g_1 + f_2 g_2) = \rho_L(f_1)\rho_L(g_1) + \rho_L(f_2)\rho_L(g_2) = f_1(L)g_1(L) + f_2(L)g_2(L).$$

Define

$$U_1 = \mathrm{im} f_1(L), \quad U_2 = \mathrm{im} f_2(L).$$

Then, since

$$f_1(L) \circ L = L \circ f_1(L), \quad f_2(L) \circ L = L \circ f_2(L),$$

(you should check this) we have that, if $u_1 = f_1(L)(x_1) \in U_1, u_2 = f_2(L)(x_2) \in U_2$, then

$$L(u_1) = L \circ f_1(L)(x_1) = f_1(L) \circ L(x_1) \in \mathrm{im} f_1(L) = U_1,\ L(u_2) = L \circ f_2(L)(x_2) = f_2(L) \circ L(x_2) \in \mathrm{im} f_2(L) = U_2.$$

Hence, U_1, U_2 are L-invariant.

Now, let $u_1 \in U_1 = \mathrm{im} f_1(L)$ so that $u_1 = f_1(L)(x_1)$, for some $x_1 \in V$. Then,

$$f_2(L)(u_1) = f_2(L)(f_1(L)(x_1)) = f(L)(x_1),$$

since $f_1 f_2 = f$ and $\rho_L(f) = \rho_L(f_1 f_2) = \rho_L(f_1)\rho_L(f_2)$ (ρ_L is a representation of $\mathbb{C}[t]$). Our assumption is that f is an annhiliating polynomial of L so that $f(L) = 0_{\mathrm{End}_{\mathbb{C}}(V)}$ and we obtain

$$f_2(L)(u_1) = f(L)(x_1) = 0_V.$$

Similarly we obtain that

$$f_1(L)(u_2) = 0_V, \quad \text{for every } u_2 \in U_2.$$

Let $v \in V$. Then,

$$v = \mathrm{id}_V(v) = (f_1(L)g_1(L) + f_2(L)g_2(L))(v) = f_1(L)(g_1(L)(v)) + f_2(L)(g_2(L)(v)) \in U_1 + U_2.$$

Hence, $V = U_1 + U_2$.

Now, let $x \in U_1 \cap U_2$. Therefore, we have $f_1(L)(x) = 0_V = f_2(L)(x)$ by what we showed above. Hence,

$$x = f_1(L)(g_1(L)(x)) + f_2(L)(g_2(L)(x)) = g_1(L)(f_1(L)(x)) + g_2(L)(f_2(L)(x)) = g_1(L)(0_V) + g_2(L)(0_V) = 0_V.$$

Here we have used that $h(L) \circ g(L) = g(L) \circ h(L)$, for any $g, h \in \mathbb{C}[t]$, which can be easily verified.

Hence, we have

$$V = U_1 \oplus U_2,$$

Finally, suppose that $f_2(L)(w) = 0_V$, for some $w \in V$. Then, we want to show that $w \in U_1$. Since $V = U_1 \oplus U_2$ then we have

$$w = u_1 + u_2,$$

where $u_1 \in U_1, u_2 \in U_2$. Thus, we have $x_1, x_2 \in V$ such that

$$u_1 = f_1(L)(x_1), \quad x_2 = f_2(L)(x_2).$$

Thus,

$$0_V = f_2(L)(w) = f_2(L)(u_1 + u_2) = f_2(L)(u_1) + f_2(L)(u_2) = 0_V + f_2(L)(u_2),$$

and

$$f_1(L)(u_2) = 0,$$

as $u_2 \in U_2$. Therefore,

$$u_2 = g_1(L)(f_1(L)(u_2)) + g_2(L)(f_2(L)(u_2)) = 0_V + 0_V = 0_V,$$

so that $w = u_1 \in U_1$. We obtain that $\ker f_1(L) = U_2$ similarly. □

Corollary 2.5.3 (Primary Decomposition Theorem)**.** *Let $f \in \mathbb{C}[t]$ be an annihilating polynomial of $L \in \mathrm{End}_{\mathbb{C}}(V)$. Suppose that f is decomposed into the following linear factors:*[50]

$$f = a(t - \lambda_1)^{n_1}(t - \lambda_2)^{n_2} \cdots (t - \lambda_k)^{n_k}.$$

Then, there are L-invariant subspaces $U_1, \ldots, U_k \subset V$ such that

$$V = U_1 \oplus \ldots \oplus U_k,$$

and such that each U_i is annihilated by the endomorphism

$$(L - \lambda_i \mathrm{id}_V)^{n_i} = (L - \lambda_i \mathrm{id}_V) \circ \cdots \circ (L - \lambda_i \mathrm{id}_V).$$

[50] This is always possible by the Fundamental Theorem of Algebra.

Proof: This is a direct consequence of Theorem 2.5.2: apply Theorem 2.5.2 to

$$f_1 = (t-\lambda_1)^{n_1},\ g_1 = (t-\lambda_2)^{n_2}\cdots(t-\lambda_k)^{n_k},$$

which are obviously relatively prime polynomials, to obtain

$$V = U_1 \oplus V_1,$$

where $U_1 = \ker f_1(L), V_1 = \ker g_1(L)$. Then, V_1 is L-invariant so that L restricts to a well-defined endomorphism of V_1, denoted $L_1 \in \mathrm{End}_{\mathbb{C}}(V_1)$. Then, g_1 is an annihilating polynomial of L_1.

Now, we can write

$$g_1 = f_2 g_2,$$

where

$$f_2 = (t-\lambda_2)^{n_2},\ g_2 = (t-\lambda_3)^{n_3}\cdots(t-\lambda_k)^{n_k}.$$

Then, f_2 and g_2 are relatively prime so we can apply Theorem 2.5.2 to V_1 to obtain

$$V_1 = U_2 \oplus V_2.$$

with $U_2 = \ker f_2(L), V_2 = \ker g_2(L)$. Then, V_2 is L_1-invariant (and also L-invariant, when we consider V_2 as a subspace of V) so that L_1 restricts to a well-defined endomorphism of V_2, denoted $L_2 \in \mathrm{End}_{\mathbb{C}}(V_2)$. Then, g_2 is an annihilating polynomial of L_2.

Proceeding in this way we see that we can write

$$V = U_1 \oplus \cdots \oplus U_k,$$

where $U_i = \ker(L - \lambda_i \mathrm{id}_V)^{n_i}$. □

Remark 2.5.4. Theorem 2.5.2 and the Primary Decomposition Theorem (Corollary 2.5.3) form the theoretical basis for the study of endomorphisms of a finite dimensional $\mathbb{C}$-vector space. These results allow us to deduce many properties of an endomorphism L if we know its minimal polynomial (or its characteristic polynomial). The next few Corollaries demonstrate this.

Corollary 2.5.5. *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$. Then, L is diagonalisable if and only if μ_L is a product of distinct linear factors, ie,*

$$\mu_L = (t-c_1)(t-c_2)\cdots(t-c_k),$$

with $c_i \neq c_j$ for $i \neq j$.

Proof: ($\Rightarrow$) Suppose that L is diagonalisable so that we have

$$E^L_{\lambda_1} \oplus \cdots \oplus E^L_{\lambda_k} = V,$$

with $E^L_{\lambda_i}$ the λ_i-eigenspace of L. Consider the polynomial

$$f = (t-\lambda_1)\cdots(t-\lambda_k) \in \mathbb{C}[t].$$

Then, we claim that $\rho_L(f) = 0 \in \mathrm{End}_{\mathbb{C}}(V)$: indeed, let $v \in V$ and write $v = e_1 + \ldots + e_k$ with $e_i \in E^L_{\lambda_i}$. Then, for each i, we have

$$\rho_L(f)(e_i) = (L-\lambda_1\mathrm{id}_V)\cdots(L-\lambda_k\mathrm{id}_V)(e_i) = 0_V,$$

because $(L-\lambda_s\mathrm{id}_V)(L-\lambda_t\mathrm{id}_V) = (L-\lambda_t\mathrm{id}_V)(L-\lambda_s\mathrm{id}_V)$, for every s,t.[51] Hence, we must have $\rho_L(f)(v) = 0_V$, for every $v \in V$, so that $\rho_L(f) = 0 \in \mathrm{End}_{\mathbb{C}}(V)$. Hence, by Proposition 2.4.4, there is some $g \in \mathbb{C}[t]$ such that

$$f = \mu_L g.$$

As f is a product of distinct linear factors the same must be true of μ_L.

[51] We can move $(L-\lambda_i\mathrm{id}_V)$ to the front of $\rho_L(f)$ and, since $L(e_i) = \lambda_i e_i$, we obtain $(L-\lambda_i\mathrm{id}_V)(e_i) = 0_V$.

($\Leftarrow$) Suppose that

$$\mu_L = (t - c_1) \cdots (t - c_k) \in \mathbb{C}[t].$$

Then, by Corollary 2.5.3, we obtain a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_k,$$

where $U_i = \ker(L - c_i \mathrm{id}_V)$. Hence,

$$U_i = \{v \in V \mid (L - c_i \mathrm{id}_V)(v) = 0_V\} = \{v \in V \mid L(v) = c_i v\} = E^L_{c_i}$$

is precisely the c_i-eigenspace of L. Thus, as we have written V as a direct sum of eigenspaces of L we must have that L is diagonalisable. □

Example 2.5.6. 1. Let $A \in Mat_n(\mathbb{C})$ be such that

$$A^k - I_n = 0_n,$$

for some $k \in \mathbb{N}$. Then, we see that

$$f = t^k - 1 \in \ker \rho_A,$$

where $\rho_A = \rho_{T_A}$ is the representation of $\mathbb{C}[t]$ defined by the endomorphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n)$. Therefore, the minimal polynomial of A, μ_A, must divide f so that there is $g \in \mathbb{C}[t]$ such that

$$f = \mu_A g.$$

Now, we have

$$f = (t - 1)(t - \omega) \cdots (t - \omega^{k-1}),$$

where $\omega = \cos(2\pi/k) + \sin(2\pi/k)\sqrt{-1}$; in particular, f has distinct linear factors. Thus, the same must be true of μ_A. Hence, by Corollary 2.5.5 we have that A is diagonalisable.

For those of you that are taking Math 113 this has an important consequence:

'every commutative finite group can be realised as a subgroup of D_n, for some n'

where D_n is the group of diagonal $n \times n$ complex matrices. This uses Cayley's theorem (for groups) and the fact that a family of commuting diagonalisable matrices can be simultaneously diagonalised (mentioned as a footnote on LH3).

2. More generally, $A \in Mat_n(\mathbb{C})$ is such that there exists a polynomial relation

$$0 = f(A) = \rho_A(f),$$

for some $f \in \mathbb{C}[t]$ with distinct linear factors, then A is diagonalisable. For example, if

$$A^2 - 3A + 2I_n = 0_n,$$

then A is diagonalisable.

The previous Corollary shows that the zeros of the minimal polynomial are eigenvalues of L, for L diagonalisable. In fact, **this is true for any $L \in \mathrm{End}_{\mathbb{C}}(V)$.**

Corollary 2.5.7. *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$ and $\mu_L \in \mathbb{C}[t]$ the minimal polynomial of L. Then, $\mu_L(c) = 0$ if and only if $c \in \mathbb{C}$ is an eigenvalue of L.*

Proof: Suppose that

$$\mu_L = (t - c_1)^{n_1} \cdots (t - c_k)^{n_k}.$$

Then, $\mu_L(c) = 0$ if and only if $c = c_i$, for some $i \in \{1, \dots, k\}$. We will show that each c_i is an eigenvalue of L and, conversely, if λ is an eigenvalue of L then $\lambda = c_i$, for some i. This shows that the set of eigenvalues of L is precisely $\{c_1, \dots, c_k\}$.

Let $U_1, \dots, U_k \subset V$ be the L-invariant subspaces such that

$$V = U_1 \oplus \cdots \oplus U_k,$$

from Corollary 2.5.3. Then, the proof of Corollary 2.5.3 shows that $U_i = \ker(L - c_i \mathrm{id}_V)^{n_i}$. As $n_i \geq 1$ we can find nonzero $w \in V$ such that $(L - c_i \mathrm{id}_V)(v) = 0_V$: namely, we take

$$w = (L - c_i \mathrm{id}_V)^{r-1}(v),$$

where $r = \mathrm{ht}(v)$ is equal to the height of any nonzero $v \in U_i$ with respect to the nilpotent endomorphism $(L_{|U_i} - c_i \mathrm{id}_{U_i}) \in \mathrm{End}_{\mathbb{C}}(U_i)$.[52] Hence,

$$(L - c_i \mathrm{id}_V)(w) = (L - c_i \mathrm{id}_V)^r(v) = 0_V,$$

so that w is eigenvector of L with associated eigenvalue c_i. In particular, c_i is an eigenvalue of L.

Conversely, suppose that $c \in \mathbb{C}$ is an eigenvalue of L and that v is an eigenvector such that $L(v) = cv$; in particular, $v \neq 0_V$. Then, since

$$V = U_1 \oplus \cdots \oplus U_k,$$

we have a unique expression

$$v = u_1 + \ldots + u_k, \ u_i \in U_i.$$

Then,

$$L(u_1) + \ldots + L(u_k) = L(v) = cv = cu_1 + \ldots + cu_k,$$

and since $L(u_i) \in U_i$ (each U_i is L-invariant) we must have $L(u_i) = cu_i$, for each i: this follows because every $z \in V$ can be written as a unique linear combination of vectors in $U_1, \dots, U_k$.

Let $\Gamma_1 = \{i \in \{1, \dots, k\} \mid u_i = 0_V\}$ and $\Gamma_2 = \{1, \dots, k\} \setminus \Gamma_1$: as $v \neq 0_V$ we must have $\Gamma_2 \neq \varnothing$. Thus, for every $i \in \Gamma_2$ we have that $u_i \in U_i$ is also an eigenvector of L with associated eigenvalue c. As

$$U_i = \ker(L - c_i \mathrm{id}_V)^{n_i},$$

we have, for each $i \in \Gamma_2$,

$$0_V = (L - c_i \mathrm{id}_V)^{n_i}(u_i) = \left(\sum_{p=0}^{n_i} \binom{n_i}{p} (-c_i)^p L^{n-p} \right)(u_i) = \sum_{p=0}^{n_i} \binom{n_i}{p} (-c_i)^p c^{n-p} u_i = (c - c_i)^{n_i} u_i.$$

Hence, we see that $c = c_i$, for each $i \in \Gamma_2$. Since $c_i \neq c_j$, if $i \neq j$, then we must have that $c = c_j$, for some j, so that any eigenvalue of L is equal to some c_j.

We have just shown that the set of eigenvalues of L is precisely $\{c_1, \dots, c_k\}$. Moreover, the set of roots of μ_L is also equal to this set and the result follows. □

Corollary 2.5.8. *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$ and $\mu_L \in \mathbb{C}[t]$ the minimal polynomial of L. Suppose that*

$$V = U_1 \oplus \cdots \oplus U_k,$$

is the direct sum decomposition from Corollary 2.5.3. Then, if c is an eigenvalue of L we must have that the c-eigenspace of L satisfies

$$E_c^L \subset U_j,$$

for some j. Furthermore, if c, c' are eigenvalues of L and $E_c^L, E_{c'}^L \subset U_j$, then $c = c'$.

[52] This is an endomorphism of U_i since U_i is L-invariant.

Proof: This follows from the latter part of the the previous proof of Corollary 2.5.7: if $v \in E_c^L$ is nonzero, so that $L(v) = cv$, then we have

$$v = u_1 + ... + u_k,\ u_i \in U_i,$$

as above. Moreover, if we define Γ_2 as before, then the latter part of the previous proof shows that $\Gamma_2 = \{j\}$, for some j. Thus,

$$v = u_j \in U_j.$$

Hence, $E_c^L \subset U_j$, for some j. The last statement follow from the proof of Corollary 2.5.7. □

Corollary 2.5.9 (Cayley-Hamilton Theorem)**.** *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$ and $\chi_L \in \mathbb{C}[t]$ the characteristic polynomial of L. Then,*

$$\chi_L(L) = \rho_L(\chi_L) = 0_{\mathrm{End}_{\mathbb{C}}(V)} \in \mathrm{End}_{\mathbb{C}}(V).$$

Proof: This is a consequence of Corollary 2.5.7. The roots of the minimal polynomial of L, μ_L, are precisely the eigenvalues of L. The roots of χ_L are also the eigenvalues of L. Therefore, we see that

$$\mu_L = (t-\lambda_1)^{m_1} \cdots (t-\lambda_k)^{m_k}, \text{ and } \chi_L = (t-\lambda_1)^{n_1} \cdots (t-\lambda_k)^{n_k}.$$

We are going to show that $m_i \leq n_i$, for each i. First we need the following Lemma (which can be easily proved by induction on k and expanding the determinant across the top row)

Lemma 2.5.10. *Let $A \in Mat_n(\mathbb{C})$ and suppose that*

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$$

with $A_i \in Mat_k(\mathbb{C}), A_2 \in Mat_{n-k}(\mathbb{C})$. Then, $\chi_A(\lambda) = \chi_{A_1}(\lambda)\chi_{A_2}(\lambda)$

If $\mathcal{B} = \mathcal{B}_1 \cup ... \cup \mathcal{B}_k$ is a basis of V, with each $\mathcal{B}_i \subset U_i$, then the matrix $[L]_{\mathcal{B}}$ is block diagonal

$$[L]_{\mathcal{B}} = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix}.$$

As a consequence of Lemma 2.5.10 we have that

$$\chi_L = \chi_{A_1}\chi_{A_2} \cdots \chi_{A_k}.$$

Moreover, it follows from the proof of Corollary 2.5.7 and Corollary 2.5.8 that the only eigenvalue of A_i is λ_i. Hence, using Lemma 2.5.10 we must have that

$$\chi_{A_i} = (t-\lambda_i)^{n_i}.$$

It is a further consequence of Lemma 2.5.10 that $\dim U_i = n_i$.

Since the endomorphism $N_i = L_{|U_i} - \lambda_i \mathrm{id}_{U_i} \in \mathrm{End}_{\mathbb{C}}(U_i)$ is nilpotent (Corollary 2.5.3) the structure theorem for nilpotent endomorphisms (Theorem 2.3.4) shows that $\eta(N_i) \leq n_i$, where $\eta(N_i)$ is the exponent of N_i.

By construction, we have that

$$U_i = \ker(L - \lambda_i \mathrm{id}_V)^{m_i},$$

which implies that $\eta(N_i) \leq m_i$. In fact, $\eta(N_i) = m_i$, for every i: otherwise, we must have $\eta(N_i) < m_i$, for some i, so that for every $u \in U_i$,

$$(L - \lambda \mathrm{id}_V)^{\eta(N_i)}(u) = 0_V.$$

Consider the polynomial

$$g = (t-\lambda_1)^{m_1} \cdots (t-\lambda_{i-1})^{m_{i-1}} (t-\lambda_i)^{\eta(N_i)} (t-\lambda_{i+1})^{m_{i+1}} \cdots (t-\lambda_k)^{m_k} \in \mathbb{C}[t].$$

We have that $\deg g < \deg \mu_L$ as $\eta(N_i) < m_i$. Then, for any $v \in V$, if we write $v = u_1 + ... + u_k$, then we see that

$$\begin{aligned}\rho_L(g)(v) &= \rho_L(g)(u_1 + ... + u_k) \\ &= \rho_L(g)(u_1) + ... + \rho_L(g)(u_k) \\ &= 0_V + ... + 0_V = 0_V,\end{aligned}$$

because

$$(L - \lambda_j \mathrm{id}_V)^{m_j}(u_j) = 0_V, \quad \text{for } j \neq i, \quad \text{and} \quad (L - \lambda_i \mathrm{id}_V)^{\eta(N_i)}(u_i) = 0_V.$$

But then this contradicts the definition of μ_L being a nonzero element of $\ker \rho_L$ of minimal degree. Hence, our initial assumption the $\eta(N_i) < m_i$, for some i, cannot hold so that $\eta(N_i) = m_i$, for every i.

Therefore, $m_i \leq n_i$, for every i, so that μ_L divides χ_L: there exists $f \in \mathbb{C}[t]$ such that

$$\chi_L = \mu_L f \in \mathbb{C}[t].$$

Hence, we obtain

$$\rho_L(\chi_L) = \rho_L(\mu_L f) = \rho_L(\mu_L)\rho_L(f) = 0_{\mathrm{End}_{\mathbb{C}}(V)} \in \mathrm{End}_{\mathbb{C}}(V),$$

where we use that $\mu_L \in \ker \rho_L$. □

Remark 2.5.11. The Cayley-Hamilton theorem is important as it gives us an upper bound on the degree of the minimal polynomial: we know that the minimal polynomial of L must have degree at most n^2 (because the set $\{\mathrm{id}_v, L, \dots, L^{n^2}\} \subset \mathrm{End}_{\mathbb{C}}(V)$ must be linearly dependent), so that $\deg \mu_L \leq n^2$. However, the Cayley-Hamilton theorem says that we actually have $\deg \mu_L \leq n$ thereby limiting the possibilities for μ_L.

2.5.1 The Jordan canonical form

Let us denote

$$N_i = L_{|U_i} - \lambda_i \mathrm{id}_{U_i} \in \mathrm{End}_{\mathbb{C}}(U_i).$$

Since each U_i is L-invariant it is also N_i-invariant (Lemma 2.2.3). Moreover, Corollary 2.5.3 implies that the restriction of N_i to U_i is a nilpotent endomorphism of U_i. Hence, by Theorem 2.3.4, we can find a basis $\mathcal{B}_i \subset U_i$ of U_i such that the matrix of the restriction of N_i with respect to $\mathcal{B}_i$ has the canonical form

$$\begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & J_{p_i} \end{bmatrix},$$

with each J_a a 0-Jordan block and such that the size of J_i is at least as large as the size of J_{i+1}. Let $\mathcal{B} = \mathcal{B}_1 \cup ... \cup \mathcal{B}_k$ be the subsequent ordered basis of V we obtain.

As we have

$$V = U_1 \oplus \cdots \oplus U_k,$$

then for each $v \in V$, we have

$$v = u_1 + ... + u_k, \ u_i \in U_i.$$

Thus, applying L to v gives

$$L(v) = L(u_1) + ... + L(u_k) = \lambda_1 u_1 + N_1(u_1) + ... + \lambda_k u_k + N_k(u_k).$$

Hence, the matrix of L with respect to the basis $\mathcal{B}$ takes the form

$$[L]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & A_k \end{bmatrix},$$

where, for each $i = 1, \dots, k$, we have

$$A_i = \lambda_i I_{\dim U_i} + \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & J_{p_i} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & & & & & \\ 0 & \lambda_i & 1 & \cdots & 0 & & & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots & & & & & \\ 0 & \cdots & \cdots & \lambda_i & 1 & & & & & \\ 0 & \cdots & \cdots & 0 & \lambda_i & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & \lambda_i & 1 & \cdots & 0 \\ & & & & & & 0 & \lambda_i & \cdots & 0 \\ & & & & & & \vdots & \vdots & \vdots & \vdots \\ & & & & & & 0 & \cdots & \cdots & 1 \\ & & & & & & 0 & \cdots & 0 & \lambda_i \\ & & & & & & & & & \ddots \\ & & & & & & & & & & \lambda_i \end{bmatrix} \tag{2.5.1}$$

Theorem 2.5.12 (Jordan Canonical Form)**.** *Let $L \in \mathrm{End}_{\mathbb{C}}(V)$, V a finite dimensional $\mathbb{C}$-vector space. Then, there exists an ordered basis $\mathcal{B} \subset V$ such that $[L]_{\mathcal{B}}$ is a matrix of the form 2.5.1 above. We call $\mathcal{B}$ a* Jordan basis of L.

Proof: Since the minimal polynomial μ_L of L is an annihilating polynomial of L we can use Primary Decomposition (Corollary 2.5.3) to obtain a direct sum decomposition of V,

$$V = U_1 \oplus \ldots \oplus U_k.$$

Now, the previous discussion implies the existence of $\mathcal{B}$ so that $[L]_{\mathcal{B}}$ takes the desired form. □

Corollary 2.5.13. *Let $A \in Mat_n(\mathbb{C})$. Then, A is similar to a matrix of the form 2.5.1 above.*

Proof: Consider the endomorphism $T_A \in \mathrm{End}_{\mathbb{C}^n}$. Then, there is an ordered basis $\mathcal{B}$ of $\mathbb{C}^n$ such that $[T_A]_{\mathcal{B}}$ takes the desired form, by Theorem 2.5.12. Since $[T_A]_{\mathcal{S}^{(n)}} = A$, we have that A and $[T_A]_{\mathcal{B}}$ are similar (Corollary 1.7.7). □

Remark 2.5.14. 1. The Jordan canonical form is a remarkable result. However, practically it is quite difficult to determine the Jordan basis of L. The use of the Jordan canonical form is mostly in theoretical applications where you are (perhaps) only concerned with knowing what the matrix of an endomorphism looks like with respect to some basis of V. The fact that a Jordan basis exists allows us to consider only 'almost diagonal' matrices, for which it can be quite easy to show that certain properties hold true.

2. The Jordan canonical form allows us to classify *similarity classes* of matrices: a similarity class is the set of all matrices which are similar to a particular matrix. Since similiarity is an equivalence relation we can partition $Mat_n(\mathbb{C})$ into disjoint similarity classes. Then, the Jordan canonical form tells us that each similarity class is labelled by a set of eigenvalues (the entries on the diagonal of the Jordan form lying in that similarity class) and the partitions of each block. Two matrices are similar if and only if these pieces of data are equal.

3. In group-theoretic language, we see that the Jordan canonical form allows us to classify the orbits of $\mathrm{GL}_n(\mathbb{C})$ acting on the set $Mat_n(\mathbb{C})$. Furthermore, this is actually the same thing as classifying the Ad-orbits of the algebraic group $\mathrm{GL}_n(\mathbb{C})$ acting on its Lie algebra $\mathfrak{gl}_n(\mathbb{C})$ via the Adjoint representation.

Example 2.5.15. Consider the following matrix

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

Then, you can check that

$$\chi_A(t) = -(t-2)^2(t+3).$$

Since

$$A^2 + A - 6I_3 \neq 0_3,$$

it is not possible for A to be diagonalisable as this is the only possibility for the minimal polynomial μ_A with distinct linear factors.

Therfore, it must be the case that there exists $P \in \mathrm{GL}_3(\mathbb{C})$ such that

$$P^{-1}AP = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix},$$

as this is the only possibility for the Jordan canonical form of A. Let's determine a basis $\mathcal{B} \subset \mathbb{C}^3$ such that

$$P_{\mathcal{B}\leftarrow\mathcal{S}^{(3)}}[T_A]_{\mathcal{S}^{(3)}}P_{\mathcal{S}^{(3)}\leftarrow\mathcal{B}} = [T_A]_{\mathcal{B}} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

As

$$\mu_A = (t-2)^2(t+3),$$

is an annihilating polynomial of A and $f_1 = (t-2)^2, f_2 = (t+3)$ are relatively prime, then we can find A-invariant subspaces $U_1, U_2 \subset \mathbb{C}^3$ such that

$$\mathbb{C}^3 = U_1 \oplus U_2,$$

and where

$$U_1 = \ker T_{(A-2I_3)^2}, \ U_2 = \ker T_{A+3I_3}.$$

You can check that

$$U_2 = E_{-3} = \mathrm{span}_{\mathbb{C}} \left\{ \begin{bmatrix} -5/28 \\ -13/28 \\ 1 \end{bmatrix} \right\},$$

so that A defines an endomorphism $T_2 : U_2 \to U_2 \ ; \ \underline{x} \mapsto A\underline{x}$ of U_2 and if $\mathcal{B}_2 = \left(\begin{bmatrix} -5/28 \\ -13/28 \\ 1 \end{bmatrix} \right) \subset U_2$ then

$$[T_2]_{\mathcal{B}_2} = [-3].$$

We also know that A defines and endomorphism $T_1 : U_1 \to U_1 \ ; \ \underline{x} \mapsto A\underline{x}$. Now, since

$$(A - 2I_3)^2 = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 1 & 3 \\ -5 & -1 & -6 \end{bmatrix},$$

we find that

$$U_1 = \ker T_{(A-2I_3)^2} = \mathrm{span}_{\mathbb{C}} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

So, if we let

$$\mathcal{C}_1 = \left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) (= (c_1, c_2)),$$

then

$$[T_2]_{\mathcal{C}_1} = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}.$$

If we set

$$N_1 = [T_2]_{\mathcal{C}_1} - 2I_2 = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix},$$

then we see that $N_1^2 = 0_2$, so that N_1 is nilpotent. Moreover, using our results on nilpotent matrices, if we set $P = [N_1 e_2 \ e_2]$ then we have

$$P^{-1} N_1 P = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Hence, we have

$$[T_1]_{\mathcal{B}} = N_1 + 2I_2 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$

Therefore, if we let

$$\mathcal{B}_1 = (c_1 + c_2, c_2) = \left(\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right),$$

and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ then we have

$$[T_A]_{\mathcal{B}} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

In particular, if we set

$$P = \begin{bmatrix} 1 & 0 & -5/28 \\ -1 & 1 & -13/28 \\ -1 & 0 & 1 \end{bmatrix},$$

then

$$P^{-1} A P = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

3 Bilinear Forms & Euclidean/Hermitian Spaces

Bilinear forms are a natural generalisation of linear forms and appear in many areas of mathematics. Just as linear algebra can be considered as the study of 'degree one' mathematics, bilinear forms arise when we are considering 'degree two' (or quadratic) mathematics. For example, an inner product is an example of a bilinear form and it is through inner products that we define the notion of length in analytic geometry - recall that the length of a vector $\underline{x} \in \mathbb{R}^n$ is defined to be $\sqrt{x_1^2 + ... + x_n^2}$ and that this formula holds as a consequence of Pythagoras' Theorem. In addition, the 'Hessian' matrix that is introduced in multivariable calculus can be considered as defining a bilinear form on tangent spaces and allows us to give well-defined notions of length and angle in tangent spaces to geometric objects. Through considering the properties of this bilinear form we are able to deduce geometric information - for example, the local nature of critical points of a geometric surface.

In this final chapter we will give an introduction to arbitrary bilinear forms on $\mathbb{K}$-vector spaces and then specialise to the case $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. By restricting our attention to thse number fields we can deduce some particularly nice classification theorems. We will also give an introduction to Euclidean spaces: these are $\mathbb{R}$-vector spaces that are equipped with an inner product and for which we can 'do Euclidean geometry', that is, all of the geometric Theorems of Euclid will hold true in any arbitrary Euclidean space. We will discuss the notions of orthogonality (=perpendicularity) and try to understand those linear transformations of a Euclidean space that are length-preserving. We will then generalise to $\mathbb{C}$-vector spaces and consider Hermitian spaces and unitary morphisms - these are the complex analogues of Euclidean spaces, where we make use of the 'conjugation' operation that exists on $\mathbb{C}$.

3.1 Bilinear forms

([1], p.179-182) **Throughout this section $\mathbb{K}$ can be ANY number field. V will always denote a finite dimensional $\mathbb{K}$-vector space.**

In this section we will give the basic definitions of bilinear forms and discuss the basic properties of symmetric and alternating bilinear forms. We will see that matrices are useful in understanding bilinear forms and provide us with a tool with which we can determine properties of a given bilinear form

Definition 3.1.1. Let V be a finite dimensional $\mathbb{K}$-vector space. A $\mathbb{K}$-*bilinear form on* V is a function

$$B : V \times V \to \mathbb{K} \; ; \; (u, v) \mapsto B(u, v),$$

such that

(BF1) for every $u, v, w \in V, \lambda \in \mathbb{K}$ we have

$$B(u + \lambda v, w) = B(u, w) + \lambda B(v, w),$$

(BF2) for every $u, v, w \in V, \lambda \in \mathbb{K}$ we have

$$B(u, v + \lambda w) = B(u, v) + \lambda B(u, w).$$

We say that a $\mathbb{K}$-bilinear form on V, B, is *symmetric* if

$$B(u, v) = B(v, u), \quad \text{for every } u, v \in V.$$

We say that a $\mathbb{K}$-bilinear form on V, B, is *antisymmetric* if

$$B(u, v) = -B(v, u), \quad \text{for every } u, v \in V.$$

We denote the set of all $\mathbb{K}$-bilinear forms on V by $\text{Bil}_{\mathbb{K}}(V)$. This is a $\mathbb{K}$-vector space (the $\mathbb{K}$-vector space structure will be discussed in a worksheet/homework).

Remark 3.1.2. 1. The conditions BF1, BF2 that a bilinear form B must satisfy can be restated as saying that

'B is linear in each argument.'

2. We will refer to $\mathbb{K}$-bilinear forms on V as simply 'bilinear forms on V', when there is no confusion on $\mathbb{K}$, or even more simply as 'bilinear forms', when there is no confusion on V.

Example 3.1.3. 1. Let V be a finite dimensional $\mathbb{K}$-vector space and let $\alpha_1, \alpha_2 \in V^* = \mathrm{Hom}_{\mathbb{K}}(V, \mathbb{K})$ be two linear forms. Then,

$$B_{\alpha_1,\alpha_2} : V \times V \to \mathbb{K} \; ; \; (u, v) \mapsto \alpha_1(u)\alpha_2(v),$$

is a bilinear form.[53]

In fact, every bilinear form is a sum of bilinear forms of this type. This requires introducing the notion of **tensor product** which is beyond the scope of this course. You can learn about this in Math 250A, the introductory graduate algebra course.

2. Consider the function

$$B : \mathbb{Q}^3 \times \mathbb{Q}^3 \to \mathbb{Q}, \; ; \; \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right) \mapsto x_1y_1 + 3x_2y_3 - x_3y_1 + 2x_1y_3.$$

Then, it can be checked that B is a $\mathbb{Q}$-bilinear form on $\mathbb{Q}^3$. It is neither symmetric nor antisymmetric.[54]

3. Consider the 'dot product' on $\mathbb{R}^n$

$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \; ; \; (\underline{x}, \underline{y}) \mapsto \underline{x} \cdot \underline{y} = x_1y_1 + ... + x_ny_n.$$

Then, this function is a (symmetric) bilinear form on $\mathbb{R}^n$.

4. The function

$$D : \mathbb{K}^2 \times \mathbb{K}^2 \; ; \; (\underline{x}, \underline{y}) \mapsto \det([\underline{x} \; \underline{y}]),$$

where $[\underline{x} \; \underline{y}]$ is the 2×2 matrix with columns $\underline{x}, \underline{y}$, is an antisymmetric bilinear form on $\mathbb{K}^2$.

5. Let $A \in Mat_n(\mathbb{K})$. Then, we have a bilinear form

$$B_A : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K} \; ; \; (\underline{x}, \underline{y}) \mapsto \underline{x}^t A \underline{y},$$

where $\underline{x}^t = [x_1 \; \cdots \; x_n]$ is the row vector determined by the column vector $\underline{x}$. That B_A is a bilinear form follows from basic matrix arithmetic.

B_A is symmetric if and only if A is symmetric.

B_A is antisymmetric if and only if A is antisymmetric.

It will be shown in homework that,

every bilinear form B on $\mathbb{K}^n$ is of the form $B = B_A$, for some $A \in Mat_n(\mathbb{K})$.

As an example, consider the bilinear form B on $\mathbb{Q}^3$ from Example 2 above. Then, we have

$$B = B_A, \quad \text{where} \quad A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 3 \\ -1 & 0 & 0 \end{bmatrix}.$$

[53] Check this.
[54] Why?

Indeed, we have

$$[x_1\ x_2\ x_3]\begin{bmatrix}1 & 0 & 2\\ 0 & 0 & 3\\ -1 & 0 & 0\end{bmatrix}\begin{bmatrix}y_1\\ y_2\\ y_3\end{bmatrix} = [x_1\ x_2\ x_3]\begin{bmatrix}y_1+2y_3\\ 3y_3\\ -y_1\end{bmatrix} = x_1y_1 + 2x_1y_3 + 3x_2y_3 - x_3y_1.$$

Definition 3.1.4. Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = (b_1, \dots, b_n) \subset V$ an ordered basis and $B \in \mathrm{Bil}_{\mathbb{K}}(V)$. Then, we define *the matrix of B relative to $\mathcal{B}$* to be the matrix

$$[B]_{\mathcal{B}} = [a_{ij}] \in Mat_n(\mathbb{K}), \quad \text{where } a_{ij} = B(b_i, b_j).$$

Moreover, if $B' \in \mathrm{Bil}_{\mathbb{K}}(V)$ is another bilinear form then[55]

$$[B]_{\mathcal{B}} = [B']_{\mathcal{B}} \Leftrightarrow B = B'.$$

Hence, there is a well-defined function

$$[-]_{\mathcal{B}} : \mathrm{Bil}_{\mathbb{K}}(V) \to Mat_n(\mathbb{K})\ ;\ B \mapsto [B]_{\mathcal{B}}.$$

Note that this function is dependent on the choice of $\mathcal{B}$.

Proposition 3.1.5. *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$, $\mathcal{B} \subset V$ an ordered basis. Then,*

a) $[-]_{\mathcal{B}} : \mathrm{Bil}_{\mathbb{K}}(V) \to Mat_n(\mathbb{K})$ is a bijective $\mathbb{K}$-linear morphism.

b) Let $A \in Mat_n(\mathbb{K})$ and $B_A \in \mathrm{Bil}_{\mathbb{K}}(\mathbb{K}^n)$ be the bilinear form on $\mathbb{K}^n$ defined by A. Then,

$$[B_A]_{\mathcal{S}^{(n)}} = A.$$

c) Let $B \in \mathrm{Bil}_{\mathbb{K}}(\mathbb{K}^n)$ and denote

$$A = [B]_{\mathcal{S}^{(n)}} \in Mat_n(\mathbb{K}).$$

Then, $B_A = B$.

Proof: This is a homework exercise. □

Lemma 3.1.6. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B} = (b_1, \dots, b_n) \subset V$ an ordered basis of V and $B \in \mathrm{Bil}_{\mathbb{K}}(V)$. Then, for any $u, v \in V$ we have*

$$[u]_{\mathcal{B}}^t [B]_{\mathcal{B}} [v]_{\mathcal{B}} = B(u, v) \in \mathbb{K}.$$

Moreover, if $A \in Mat_n(\mathbb{K})$ is such that

$$[u]_{\mathcal{B}}^t A [v]_{\mathcal{B}} = B(u, v),$$

for every $u, v \in V$, then $A = [B]_{\mathcal{B}}$.

Proof: Let $u, v \in V$ and suppose that

$$u = \sum_{i=1}^{n} \lambda_i b_i,\ v = \sum_{j=1}^{n} \mu_j b_j,$$

so that

$$[u]_{\mathcal{B}}^t = [\lambda_1\ \dots\ \lambda_n],\ [v]_{\mathcal{B}} = \begin{bmatrix}\mu_1\\ \vdots\\ \mu_n\end{bmatrix}.$$

[55] Why is this true?

Then, we have

$$\begin{aligned} B(u,v) &= B(\sum_{i=1}^{n} \lambda_i b_i, \sum_{j=1}^{n} \mu_j b_j) \\ &= \sum_{i=1}^{n} \lambda_i B(b_i, \sum_{j=1}^{n} \mu_j b_j), \quad \text{by BF1,} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j B(b_i, b_j), \quad \text{by BF2.} \end{aligned}$$

Also, we see that

$$[u]_{\mathcal{B}}^t [B]_{\mathcal{B}} [v]_{\mathcal{B}} = [\lambda_1 \ \dots \ \lambda_n][B]_{\mathcal{B}} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j B(b_i, b_j).^{56}$$

The result follows.

The last statement can be checked by using the fact that

$$x_{ij} = e_i^t X e_j = B(b_i, b_j),$$

for any $X = [x_{ij}] \in Mat_n(\mathbb{K})$. □

Remark 3.1.7. 1. Suppose that V is a $\mathbb{K}$-vector space and $\mathcal{B} = (b_1, \dots, b_n) \subset V$ is an ordered basis. Let $A = [B]_{\mathcal{B}}$ be the matrix of B relative to $\mathcal{B}$. We can interpret Lemma 3.1.6 using the following commutative diagram

$$\begin{array}{ccc} V \times V & \xrightarrow{B} & \mathbb{K} \\ {\scriptstyle [-]_{\mathcal{B}} \times [-]_{\mathcal{B}}} \downarrow & \circlearrowleft & \nearrow {\scriptstyle B_A} \\ \mathbb{K}^n \times \mathbb{K}^n & & \end{array}$$

Here we have

$$[-]_{\mathcal{B}} \times [-]_{\mathcal{B}} : V \times V \to \mathbb{K}^n \times \mathbb{K}^n \ ; \ (u,v) \mapsto ([u]_{\mathcal{B}}, [v]_{\mathcal{B}}),$$

and B_A is the bilinear form on $\mathbb{K}^n$ defined by A (Example 3.1.3).

The last statement in Lemma 3.1.6 tells us that if $X \in Mat_n(\mathbb{K})$ is such that we have the commutative diagram

$$\begin{array}{ccc} V \times V & \xrightarrow{B} & \mathbb{K} \\ {\scriptstyle [-]_{\mathcal{B}} \times [-]_{\mathcal{B}}} \downarrow & \circlearrowleft & \nearrow {\scriptstyle B_X} \\ \mathbb{K}^n \times \mathbb{K}^n & & \end{array}$$

then $X = [B]_{\mathcal{B}}$.

What happens if we choose a different ordered basis $\mathcal{C} \subset V$, **how can we compare $[B]_{\mathcal{B}}$ and $[B]_{\mathcal{C}}$?**

Proposition 3.1.8. *Let V be a $\mathbb{K}$-vector space, $\mathcal{B}, \mathcal{C} \subset V$ ordered bases and $B \in \mathrm{Bil}_{\mathbb{K}}(V)$. Then, if $P = P_{\mathcal{C} \leftarrow \mathcal{B}}$ then*

$$P^t [B]_{\mathcal{C}} P = [B]_{\mathcal{B}},$$

where P^t is the transpose of P.

[56] Check this.

Proof: By Lemma 3.1.6 we know that if we can show that

$$B(u,v) = [u]_{\mathcal{B}}^t P^t [B]_{\mathcal{C}} P [v]_{\mathcal{B}},$$

for every $u, v \in V$, then we must have that

$$[B]_{\mathcal{B}} = P^t [B]_{\mathcal{C}} P.$$

Now, for any $v \in V$ we have that $P[v]_{\mathcal{B}} = [v]_{\mathcal{C}}$, since P is the change of coordinate morphism from $\mathcal{B}$ to $\mathcal{C}$. Thus, for any $u, v \in V$, we have

$$[u]_{\mathcal{B}}^t P^t [B]_{\mathcal{C}} P [v]_{\mathcal{B}} = (P[u]_{\mathcal{B}})^t [B]_{\mathcal{C}} P [v]_{\mathcal{B}} = [u]_{\mathcal{C}}^t [B]_{\mathcal{C}} [v]_{\mathcal{C}} = B(u,v),$$

where we have used that $(XY)^t = Y^t X^t$ and the defining property of $[B]_{\mathcal{C}}$. The result follows. □

3.1.1 Nondegenerate bilinear forms

We will now introduce the important notion of *nondegeneracy* of a bilinear form. Nondegenerate bilinear forms arise throughout mathematics. For example, an inner product is an example of a nondegenerate bilinear form, as is the Lorentzian metric from Einstein's Theory of Special Relativity.

Definition 3.1.9. Let V be a finite dimensional $\mathbb{K}$-vector space, $B \in \mathrm{Bil}_{\mathbb{K}}(V)$. Then, we say that B is *nondegenerate* if the following property holds:

(ND) $\quad B(u,v) = 0,$ for every $u \in V \implies v = 0_V$.

If B is not nondegenerate then we say that B is *degenerate*.

Lemma 3.1.10. *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$, $\mathcal{B} \subset V$ be an ordered basis. Then, B is nondegenerate if and only if $[B]_{\mathcal{B}}$ is an invertible matrix.*

Proof: Suppose that B is nondegenerate. We will show that $A = [B]_{\mathcal{B}}$ is invertible by showing that $\ker T_A = \{\underline{0}\}$. So, suppose that $\underline{x} \in \mathbb{K}^n$ is such that

$$A\underline{x} = \underline{0}.$$

Then, for every $\underline{y} \in \mathbb{K}^n$ we have

$$0 = \underline{y}^t \underline{0} = \underline{y}^t A \underline{x} = B_A(\underline{y}, \underline{x}).$$

As $[-]_{\mathcal{B}} : V \to \mathbb{K}^n$ is an isomorphism we have $\underline{x} = [v]_{\mathcal{B}}$ for some unique $v \in V$. Moreover, if $\underline{y} \in \mathbb{K}^n$ then there is some unique $u \in V$ such that $\underline{y} = [v]_{\mathcal{B}}$. Hence, we have just shown that

$$0 = B_A(\underline{y}, \underline{x}) = [u]_{\mathcal{B}}^t [B]_{\mathcal{B}} [v]_{\mathcal{B}} = B(u,v),$$

by Lemma 3.1.6. Therefore, since B is nondegenerate

$$B(u,v) = 0, \quad \text{for every } u \in V \implies v = 0_V,$$

Hence, $\underline{x} = [v]_{\mathcal{B}} = \underline{0}$ so that $\ker T_A = \{\underline{0}\}$ and A must be invertible.

Conversely, suppose that $A = [B]_{\mathcal{B}}$ is invertible. We want to show that B is nondegenerate so that we must show that if

$$B(u,v) = 0, \quad \text{for every } u \in V,$$

then $v = 0_V$. Suppose that $B(u,v) = 0$, for every $u \in V$. Then, by Lemma 3.1.6, this is the same as

$$0 = B(u,v) = [u]_{\mathcal{B}}^t A [v]_{\mathcal{B}}, \quad \text{for every } u \in V.$$

In particular, if we consider $e_i = [b_i]_{\mathcal{B}}$ then we have

$$0 = e_i^t A [v]_{\mathcal{B}}, \quad \text{for every } i, \implies A[v]_{\mathcal{B}} = \underline{0}.$$

As A is invertible this implies that $[v]_{\mathcal{B}} = \underline{0}$ so that $v = 0_V$, since $[-]_{\mathcal{B}}$ is an isomorphism. □

Corollary 3.1.11. *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be a nondegenerate bilinear form. Then,*

$$B(u,v) = 0, \quad \textit{for every } v \in V, \implies u = 0_V.$$

BEWARE: this condition is (similar but) different to the one defining nondegeneracy in Definition 3.1.9. Of course, if B is symmetric then this follows from Definition 3.1.9.

Proof: This will be a homework exercise. □

Example 3.1.12. 1. Consider the bilinear form

$$B : \mathbb{Q}^3 \times \mathbb{Q}^3 \to \mathbb{Q} \; ; \; (\underline{x}, \underline{y}) \mapsto x_1y_2 + x_3y_2 + x_2y_1.$$

Then, B is degenerate: indeed, we have

$$A = [B]_{\mathcal{S}^{(n)}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

which is non-invertible.

2. The dot product on $\mathbb{R}^n$ is nondegenerate. This will be shown in a proceeding section.

3. Consider the bilinear form

$$B : Mat_2(\mathbb{Q}) \times Mat_2(\mathbb{Q}) \to \mathbb{Q} \; ; \; (X, Y) \mapsto \mathrm{tr}(XY).$$

Then, B is nondegenerate. Suppose that $X \in Mat_2(\mathbb{Q})$ is such that

$$B(X, Y) = 0, \quad \text{for every } Y \in Mat_2(\mathbb{Q}).$$

Then, in particular, we have

$$B(X, e_{ij}) = 0, \; i, j \in \{1, 2\}.$$

Hence,

$$x_{11} = B(X, e_{11}) = 0, \; x_{12} = B(X, e_{21}) = 0, \; x_{21} = B(X, e_{12}) = 0, \; x_{22} = B(X, e_{22}) = 0,$$

so that $X = 0_2 \in Mat_2(\mathbb{Q})$.

Proposition 3.1.13. *Let V be a $\mathbb{K}$-vector space, $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ a nondegenerate bilinear form. Then, B induces an isomorphism of $\mathbb{K}$-vector spaces*

$$\sigma_B : V \to V^* \; ; \; v \mapsto \sigma_B(v),$$

where

$$\sigma_B(v) : V \to \mathbb{K} \; ; \; u \mapsto \sigma_B(v)(u) = B(u, v).$$

Proof: It is left as an exercise to check that σ_B is well-defined, ie, that σ_B is $\mathbb{K}$-linear and $\sigma_B(v) \in V^*$, for every $v \in V$.

Since we know that $\dim V = \dim V^*$ it suffices to show that σ_B is injective. So, suppose that $v \in \ker \sigma_B$. Then, $\sigma_B(v) = 0 \in V^*$, so that $\sigma_B(v)$ is the zero linear form. Hence, we have $\sigma_B(v)(u) = 0$, for every $u \in V$. Thus, using nondegeneracy of B we have

$$0 = \sigma_B(v)(u) = B(u, v), \quad \text{for every } u \in V, \implies v = 0_V.$$

Hence, σ_B is injective and the result follows. □

Remark 3.1.14. 1. We could have also defined an isomorphism

$$\hat{\sigma}_B : V \to V^*,$$

where

$$\hat{\sigma}_B(v)(u) = B(v,u), \text{ for every } u \in V.$$

If B is symmetric then we have

$$\sigma_B = \hat{\sigma}_B,$$

but this is not the case in general.

2. In fact, Proposition 3.1.13 has a converse: suppose that σ_B induces an isomorphism

$$\sigma_B : V \to V^*.$$

Then, B is nondegenerate. This follows because σ_B is *injective*.[57]

3. Suppose that $\mathcal{B} = (b_1, \dots, b_n) \subset V$ is an ordered basis of V and $\mathcal{B}^* = (b_1^*, \dots, b_n^*) \subset V^*$ is the dual basis (Proposition 1.8.3). **What is the matrix $[\sigma_B]_{\mathcal{B}}^{\mathcal{B}^*}$ of σ_B with respect to $\mathcal{B}$ and $\mathcal{B}^*$?**

By definition we have

$$[\sigma_B]_{\mathcal{B}}^{\mathcal{B}^*} = [[\sigma_B(b_1)]_{\mathcal{B}^*} \cdots [\sigma_B(b_n)]_{\mathcal{B}^*}].$$

Now, for each i, $\sigma_B(b_i) \in V^*$ is a linear form on V so we need to know what it does to elements of V. Suppose that

$$v = \lambda_1 b_1 + \dots + \lambda_n b_n \in V.$$

Then,

$$\sigma_B(b_i)(v) = B(\sum_{k=1}^{n} \lambda_k b_k, b_i) = \sum_{k=1}^{n} \lambda_k B(b_k, b_i),$$

and

$$\left(\sum_{j=1}^{n} B(b_j, b_i) b_j^*\right)(v) = \left(\sum_{j=1}^{n} B(b_j, b_i) b_j^*\right)\left(\sum_{k=1}^{n} \lambda_k b_k\right) = \sum_{k=1}^{n} \lambda_k B(b_k, b_i),$$

so that we must have

$$\sigma_B(b_i) = \sum_{j=1}^{n} B(b_j, b_i) b_j^*.$$

Hence,

$$\boxed{[\sigma_B]_{\mathcal{B}}^{\mathcal{B}^*} = [B]_{\mathcal{B}}.}$$

It is now clear that B is nondegenerate precisely when the morphism σ_B is an isomorphism.

Definition 3.1.15. Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$. Let $E \subset V$ be a nonempty subset. Then, we define the *(right) B-complement of E in V* to be the set

$$E_r^{\perp} = \{v \in V \mid B(u,v) = 0 \text{ for every } u \in E\};$$

this is a subspace of V.[58]

Similarly, we define the *(left) B-complement of E in V* to be the set

$$E_l^{\perp} = \{v \in V \mid B(v,u) = 0, \text{ for every } u \in E\};$$

[57] Some people actually use this property to *define* nondegeneracy: they say that B is nondegenerate if σ_B is injective. If you think about it, you will see that these two definitions are saying the exact same thing.

[58] Check this.

this is a subspace of V.[59]

If B is (anti-)symmetric then we have that

$$E_l^\perp = E_r^\perp.$$

In this case we write $E^\perp$.

Remark 3.1.16. Let $E \subset V$ be a nonempty subset and $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be (anti-)symmetric. Then, it is not hard to see that

$$E^\perp = \mathrm{span}_{\mathbb{K}}(E)^\perp.$$

Indeed, we obviously have

$$\mathrm{span}_{\mathbb{K}}(E)^\perp \subset E^\perp,$$

since if $B(u,v) = 0$, for every $u \in \mathrm{span}_{\mathbb{K}}(E)$, then this must also hold for those $u \in E$. Hence, $v \in \mathrm{span}_{\mathbb{K}}(E)^\perp \implies v \in E^\perp$. Conversely, if $v \in E^\perp$, so that $B(e,v) = 0$, for every $e \in E$, then if $w = c_1 e_1 + ... + c_k e_k \in \mathrm{span}_{\mathbb{K}}(E)$, then

$$B(w,v) = B(c_1 e_1 + ... c_k e_k, w) = c_1 B(e_1, v) + ... + c_k B(e_k, w) = 0 + ... + 0 = 0.$$

Proposition 3.1.17. *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be (anti-)symmetric and nondegenerate, $U \subset V$ a subspace of V. Then,*

$$\dim U + \dim U^\perp = \dim V.$$

Proof: As B is nondegenerate we can consider the isomorphism

$$\sigma_B : V \to V^*,$$

from Proposition 3.1.13. We are going to show that

$$\sigma_B(U^\perp) = \mathrm{ann}_{V^*}(U) = \{\alpha \in V^* \mid \alpha(u) = 0, \text{ for every } u \in U\}.$$

Indeed, suppose that $w \in U^\perp$. Then, for every $u \in U$, we have

$$\sigma_B(w)(u) = B(u,w) = 0,$$

so that $\sigma_B(w) \in \mathrm{ann}_{V^*}(U)$. Conversely, let $\alpha \in \mathrm{ann}_{V^*}(U)$. Then, $\alpha = \sigma_B(w)$, for some $w \in V$, since σ_B is an isomorphism. Hence, for every $u \in U$, we must have

$$0 = \alpha(u) = \sigma_B(w)(u) = B(u,w),$$

so that $w \in U^\perp$ and $\alpha = \sigma_B(w) \in \sigma_B(U^\perp)$.

Hence, using Proposition 1.8.10, we have

$$\dim U^\perp = \dim \sigma_B(U^\perp) = \dim \mathrm{ann}_{V^*}(U) = \dim V - \dim U.$$

The result follows. □

3.1.2 Adjoints

Suppose that $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ is a nondegenerate symmetric bilinear form on V. Then, we have the isomorphism

$$\sigma_B : V \to V^*,$$

given above.

Consider a linear endomorphism $f \in \mathrm{End}_{\mathbb{K}}(V)$. Then, we have defined the dual of f (Definition 1.8.4)

$$f^* : V^* \to V^* \ ; \ \alpha \mapsto f^*(\alpha) = \alpha \circ f.$$

[59] Check this.

We are going to define a new morphism $f^+ : V \to V$ called the *adjoint of* f: in order to define a morphism we have to define a function and then show that it is linear.

So, given the input $v \in V$ what is the output $f^+(v) \in V$? We have $\sigma_B(v) \in V^*$ is a linear form on V and we define

$$\alpha_v = f^*(\sigma_B(v)) \in V^*.$$

As σ_B is an isomorphism, there must exist a unique $w \in V$ such that $\sigma_B(w) = \alpha_v$. We define $f^+(v) = w$: that is, $f^+(v) \in V$ is the unique vector in V such that

$$\sigma_B(f^+(v)) = f^*(\sigma_B(v)).$$

Hence, for every $u \in V$ we have that

$$\sigma_B(f^+(v))(u) = f^*(\sigma_B(v))(u) \implies \sigma_B(f^+(v))(u) = \sigma_B(v)(f(u)) \implies B(u, f^+(v)) = B(f(u), v).$$

Moreover, since we have

$$f^+ = \sigma_B^{-1} \circ f^* \circ \sigma_B,$$

then we see that f^+ is a linear morphism (it is the composition of linear morphisms, hence must be linear).

Definition 3.1.18. Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be symmetric and nondegenerate. Suppose that $f \in \mathrm{End}_{\mathbb{K}}(V)$. Then, we define *the adjoint of f (with respect to B)*, denoted f^+, to be the linear morphism

$$f^+ = \sigma_B^{-1} \circ f^* \circ \sigma_B \in \mathrm{End}_{\mathbb{K}}(V).$$

It is the unique endomorphism of V such that

$$B(u, f^+(v)) = B(f(u), v), \quad \text{for every } u, v \in V.$$

We will usually just refer to f^+ as the adjoint of f, the bilinear form B being implicitly assumed known.

Remark 3.1.19. The adjoint of a linear morphism can be quite difficult to understand at first. In particular, given an ordered basis $\mathcal{B} \subset V$, what is $[f^+]_{\mathcal{B}}$?

We use the fact that

$$f^+ = \sigma_B^{-1} \circ f^* \circ \sigma_B,$$

so that

$$\boxed{[f^+]_{\mathcal{B}} = [\sigma_B^{-1} \circ f^* \circ \sigma_B]_{\mathcal{B}} = [\sigma_B^{-1}]^{\mathcal{B}}_{\mathcal{B}^*}[f^*]_{\mathcal{B}^*}[\sigma_B]^{\mathcal{B}^*}_{\mathcal{B}} = [B]^{-1}_{\mathcal{B}}[f]^t_{\mathcal{B}}[B]_{\mathcal{B}}}$$

Hence, if $B = B_A \in \mathrm{Bil}_{\mathbb{K}}(\mathbb{K}^n)$, for some symmetric $A \in \mathrm{GL}_n(\mathbb{K})$, and $f = T_C$, where $C \in Mat_n(\mathbb{K})$, then we have

$$f^+ = T_X, \quad \text{where } X = A^{-1}C^tA.$$

Example 3.1.20. Consider the bilinear form $B = B_A \in \mathrm{Bil}_{\mathbb{Q}}(\mathbb{Q}^3)$, where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \in \mathrm{GL}_3(\mathbb{Q}).$$

Let $f \in \mathrm{End}_{\mathbb{Q}}(\mathbb{Q}^3)$ be the linear morphism

$$f : \mathbb{Q}^3 \to \mathbb{Q}^3 \; ; \; \underline{x} \mapsto C\underline{x},$$

where

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

Then, the adjoint of f is the morphism

$$f^+ : \mathbb{Q}^3 \to \mathbb{Q}^3 \; ; \; \underline{x} \mapsto \begin{bmatrix} 1 & -3 & -1 \\ 1 & 5 & 0 \\ 0 & 2 & 3 \end{bmatrix} \underline{x}.$$

As a verification, you can check that

$$B\left(\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix} 1 & -3 & -1 \\ 1 & 5 & 0 \\ 0 & 2 & 3 \end{bmatrix} \begin{bmatrix}-1\\0\\-1\end{bmatrix}\right) = B\left(\begin{bmatrix} 1 & 0 & 1 \\ -1 & 3 & 0 \\ -3 & 2 & 5 \end{bmatrix} \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\end{bmatrix}\right).$$

3.2 Real and complex symmetric bilinear forms

Throughout the remainder of these notes we will assume that $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Throughout this section we will assume that all bilinear forms are symmetric.

When we consider symmetric bilinear forms on real or complex vector spaces we obtain some particularly nice results.[60] For a $\mathbb{C}$-vector space V and symmetric bilinear form $B \in \mathrm{Bil}_{\mathbb{C}}(V)$ we will see that there is a basis $\mathcal{B} \subset V$ such that

$$[B]_{\mathcal{B}} = I_{\dim V}.$$

First we introduce the important *polarisation identity*.

Lemma 3.2.1 (Polarisation identity)**.** *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be a symmetric bilinear form. Then, for any $u, v \in V$, we have*

$$B(u,v) = \frac{1}{2}\left(B(u+v, u+v) - B(u,u) - B(v,v)\right).$$

Proof: Left as an exercise for the reader. □

Corollary 3.2.2. *Let $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ be symmetric and nonzero. Then, there exists some nonzero $v \in V$ such that $B(v,v) \neq 0$.*

Proof: Suppose that the result does not hold: that is, for every $v \in V$ we have $B(v,v) = 0$. Then, using the polarisation identity (Lemma 3.2.1) we have, for every $u, v \in V$,

$$B(u,v) = \frac{1}{2}\left(B(u+v, u+v,) - B(u,u) - B(v,v)\right) = \frac{1}{4}(0-0-0) = 0.$$

Hence, we must have that $B = 0$ is the zero bilinear form, which contradicts our assumption on B. Hence, ther must exist some $v \in V$ such that $B(v,v) \neq 0$. □

This seemingly simple result has some profound consequences for nondegenerate complex symmetric bilinear forms.

Theorem 3.2.3 (Classification of nondegenerate symmetric bilinear forms over $\mathbb{C}$)**.** *Let $B \in \mathrm{Bil}_{\mathbb{C}}(V)$ be symmetric and nondegenerate. Then, there exists an ordered basis $\mathcal{B} \subset V$ such that*

$$[B]_{\mathcal{B}} = I_{\dim V}.$$

Proof: By Corollary 3.2.2 we know that there exists some nonzero $v_1 \in V$ such that $B(v_1, v_1) \neq 0$ (we know that B is nonzero since it is nondegenerate). Let $E_1 = \mathrm{span}_{\mathbb{C}}\{v_1\}$ and consider $E_1^{\perp} \subset V$.

We have $E_1 \cap E_1^{\perp} = \{0_V\}$: indeed, let $x \in E_1 \cap E_1^{\perp}$. Then, $x = cv_1$, for some $c \in \mathbb{C}$. As $x \in E_1^{\perp}$ we must have

$$0 = B(x, v_1) = B(cv_1, v_1) = cB(v_1, v_1),$$

so that $c = 0$ (as $B(v_1, v_1) \neq 0$). Thus, by Proposition 3.1.17, we must have

$$V = E_1 \oplus E_1^{\perp}.$$

[60] Actually, all results that hold for $\mathbb{C}$-vector space also hold for $\mathbb{K}$-vector spaces, where $\mathbb{K}$ is an algebraically closed field. To say that $\mathbb{K}$ is algebraically closed means that the Fundamental Theorem of Algebra holds for $\mathbb{K}[t]$; equivalently, every polynomial $f \in \mathbb{K}[t]$ can be written as a product of linear factors.

Moreover, B restricts to a nondegenerate symmetric bilinear form on $E_1^\perp$: indeed, the restriction is

$$B_{|E_1^\perp} : E_1^\perp \times E_1^\perp \to \mathbb{C}\ ;\ (u, u') \mapsto B(u, u'),$$

and this is a symmetric bilinear form. We need to check that it is nondegenerate. Suppose that $w \in E_1^\perp$ is such that, for every $z \in E_1^\perp$ we have

$$B(z, w) = 0.$$

Then, for any $v \in V$, we have $v = cv_1 + z, z \in E_1^\perp, c \in \mathbb{C}$, so that

$$B(v, w) = B(cv_1 + z, w) = cB(v_1, w) + B(z, w) = 0 + 0 = 0,$$

where we have used the assumption on w and that $w \in E_1^\perp$. Hence, using nongeneracy of B on V we see that $w = 0_V$. Hence, we have that B is also nondegenerate on $E_1^\perp$.

As above, we can now find $v_2 \in E_1^\perp$ such that $B(v_2, v_2) \neq 0$ and, if we denote $E_2 = \text{span}_{\mathbb{C}}\{v_2\}$, then

$$E_1^\perp = E_2 \oplus E_2^\perp,$$

where $E_2^\perp$ is the B-complement of E_2 **in** $E_1^\perp$. Hence, we have

$$V = E_1 \oplus E_2 \oplus E_2^\perp.$$

Proceeding in the manner we obtain

$$V = E_1 \oplus \cdots \oplus E_n,$$

where $n = \dim V$, and where $E_i = \text{span}_{\mathbb{C}}\{v_i\}$. Moreover, by construction we have that

$$B(v_i, v_j) = 0, \quad \text{for } i \neq j.$$

Define

$$b_i = \frac{1}{\sqrt{B(v_i, v_i)}} v_i;$$

we know that the square root $\sqrt{B(v_i, v_i)}$ exists (and is nonzero) since we are considering $\mathbb{C}$-scalars.[61] Then, it is easy to see that

$$B(b_i, b_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Finally, since

$$V = \text{span}_{\mathbb{C}}\{b_1\} \oplus \cdots \oplus \text{span}_{\mathbb{C}}\{b_n\},$$

we have that $\mathcal{B} = (b_1, \dots, b_n)$ is an ordered basis such that

$$[B]_{\mathcal{B}} = I_n.$$

□

Corollary 3.2.4. *Let $A \in \text{GL}_n(\mathbb{C})$ be a symmetric matrix (so that $A = A^t$). Then, there exists $P \in \text{GL}_n(\mathbb{C})$ such that*

$$P^t A P = I_n.$$

Proof: This is just Theorem3.2.3 and Proposition 3.1.8 applied to the bilinear form $B_A \in \text{Bil}_{\mathbb{C}}(\mathbb{C}^n)$. The assumptions on A ensure that B_A is symmetric and nondegenerate. □

Corollary 3.2.5. *Suppose that $X, Y \in \text{GL}_n(\mathbb{C})$ are both symmetric. Then, there is a nondegenerate bilinear form $B \in \text{Bil}_{\mathbb{C}}(\mathbb{C}^n)$ and bases $\mathcal{B}, \mathcal{C} \subset \mathbb{C}^n$ such that*

$$X = [B]_{\mathcal{B}},\ Y = [B]_{\mathcal{C}}.$$

[61] This is a consequence of the Fundamental Theorem of Algebra: for any $c \in \mathbb{C}$ we have that

$$t^2 - c = 0,$$

has a solution.

Proof: By the previous Corollary we can find $P, Q \in \mathrm{GL}_n(\mathbb{C})$ such that

$$P^t X P = I_n = Q^t Y Q \implies (Q^{-1})^t P^t X P Q^{-1} = Y \implies (PQ^{-1})^t X P Q^{-1} = Y.$$

Now, let $B = B_X \in \mathrm{Bil}_{\mathbb{C}}(\mathbb{C}^n)$, $\mathcal{B} = \mathcal{S}^{(n)}$ and $\mathcal{C} = (c_1, \dots, c_n)$, where c_i is the i^{th} column of PQ^{-1}. Then, the above identity states that

$$[B]_{\mathcal{C}} = P^t_{\mathcal{B}\leftarrow\mathcal{C}} [B]_{\mathcal{B}} P_{\mathcal{B}\leftarrow\mathcal{C}} = Y.$$

The result follows. □

The situation is not as simple for an $\mathbb{R}$-vector space V and nondegenerate symmetric bilinear form $B \in \mathrm{Bil}_{\mathbb{R}}(V)$, however we can still obtain a nice classification result.

Theorem 3.2.6 (Sylvester's law of inertia)**.** *Let V be an $\mathbb{R}$-vector space, $B \in \mathrm{Bil}_{\mathbb{R}}(V)$ a nondegenerate symmetric bilinear form. Then, there is an ordered basis $\mathcal{B} \subset V$ such that $[B]_{\mathcal{B}}$ is a diagonal matrix*

$$[B]_{\mathcal{B}} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

where $d_i \in \{1, -1\}$.

Moreover, if $p =$ the number of 1s appearing on the diagonal and $q =$ the number of -1s appearing on the diagonal, then p and q are invariants of B: this means that if $\mathcal{C} \subset V$ is any other basis of V such that

$$[B]_{\mathcal{C}} = \begin{bmatrix} e_1 & & & \\ & e_2 & & \\ & & \ddots & \\ & & & e_n \end{bmatrix},$$

where $e_j \in \{1, -1\}$, and p' (resp. q') denotes the number of 1s (resp. -1s) on the diagonal. Then,

$$p = p', \ q = q'.$$

Proof: The proof is similar to the proof of Theorem 3.2.3: we determine $v_1, \dots, v_n \in V$ such that

$$V = \mathrm{span}_{\mathbb{R}}\{v_1\} \oplus \cdots \oplus \mathrm{span}_{\mathbb{R}}\{v_n\},$$

and with $B(v_i, v_j) = 0$, whenever $i \neq j$. However, we now run into a problem: what if $B(v_i, v_i) < 0$? We can't find a real square root of a negative number so we can't proceed as in the complex case. However, if we define

$$\delta_i = \sqrt{|B(v_i, v_i)|}, \quad \text{for every } i,$$

then we can obtain a basis $\mathcal{B} = (b_1, \dots, b_n)$, where we define

$$b_i = \frac{1}{\delta_i} v_i.$$

Then, we see that

$$B(b_i, b_j) = \begin{cases} 0, & i \neq j, \\ \pm 1, & i = j, \end{cases}$$

and $[B]_{\mathcal{B}}$ is of the required form.

Let us reorder $\mathcal{B}$ so that, for $i = 1, \dots, p$, we have $B(b_i, b_i) > 0$. Then, if we denote

$$P = \mathrm{span}_{\mathbb{R}}\{b_1, \dots, b_p\}, \quad \text{and} \quad Q = \mathrm{span}_{\mathbb{R}}\{b_{p+1}, \dots, b_n\},$$

we have

$$\dim P = p, \ \dim Q = q \ (= n - p).$$

We see that the restriction of B to P satisfies

$$B(u,u) > 0, \text{ for every } u \in P,$$

and that if $P \subset P'$, $P \neq P'$, with $P' \subset V$ a subspace, then there is some $v \in P'$ such that $B(v,v) \leq 0$: indeed, as $v \notin P$ then we have

$$v = \lambda_1 b_1 + ... + \lambda_p b_p + \mu_1 b_{p+1} + ... + \mu_q b_n,$$

and some $\mu_j \neq 0$. Then, since $P \subset P'$ we must have $b_{p+j} \in P'$ and

$$B(b_{p+j}, b_{p+j}) < 0.$$

Hence, we can see that p is the dimension of the largest subspace U of V for which the restriction of B to U satisfies $B(u,u) > 0$, for every $u \in U$.

Similarly, we can define q to be the dimension of the largest subspace $U' \in V$ for which the restriction of B to U' satisfies $B(u',u') < 0$, for every $u' \in U'$.

Therefore, we have defined p and q only in terms of B so that they are invariants of B. □

Corollary 3.2.7. *For every symmetric $A \in \mathrm{GL}_n(\mathbb{R})$, there exists $X \in GL_n(\mathbb{R})$ such that*

$$X^t A X = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

with $d_i \in \{1, -1\}$.

Definition 3.2.8. Suppose that $B \in \mathrm{Bil}_{\mathbb{R}}(V)$ is nondegenerate and symmetric and that p, q are as in Theorem 3.2.6. Then, we define the *signature of B*, denoted $\mathrm{sig}(B)$, to be the number

$$\mathrm{sig}(B) = p - q.$$

It is an invariant of B: for any basis $\mathcal{B} \subset V$ such that

$$[B]_{\mathcal{B}} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

with $d_i \in \{1, -1\}$, the quantity $p - q$ is the same.

3.2.1 Computing the canonical form of a real nondegenerate symmetric bilinear form

([1], p.185-191)

Suppose that $B \in \mathrm{Bil}_{\mathbb{R}}(V)$ is **symmetric** and **nondegenerate**, with V a finite dimensional $\mathbb{R}$-vector space. Suppose that $\mathcal{B} \subset V$ is an ordered basis such that

$$[B]_{\mathcal{B}} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

where $d_i \in \{1, -1\}$. Such a basis exists by Theorem 3.2.6. How do we determine $\mathcal{B}$?

Suppose that $\mathcal{C} \subset V$ is **any** ordered basis. Then, we know that

$$P^t_{\mathcal{C} \leftarrow \mathcal{B}} [B]_{\mathcal{C}} P_{\mathcal{C} \leftarrow \mathcal{B}} = [B]_{\mathcal{B}},$$

by Proposition 3.1.8. Hence, the problem of determining $\mathcal{B}$ is equivalent to the problem of determining $P_{\mathcal{C}\leftarrow\mathcal{B}}$ (since we already know $\mathcal{C}$ and we can use $P_{\mathcal{C}\leftarrow\mathcal{B}}$ to determine $\mathcal{B}$[62]).

Therefore, suppose that $A = [a_{ij}] \in \mathrm{GL}_n(\mathbb{R})$ is symmetric. We want to determine $P \in \mathrm{GL}_n(\mathbb{R})$ such that

$$P^tAP = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

where $d_i \in \{1, -1\}$.

Consider the column vector of variables

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Then, we have

$$\underline{x}^tA\underline{x} = a_{11}x_1^2 + ... + a_{nn}x_n^2 + 2\sum_{i<j} a_{ij}x_ix_j.$$[63]

By performing the 'completing the square' process for each variable x_i we will find variables

$$\begin{aligned} y_1 = &\quad q_{11}x_1 + q_{12}x_2 + ... + q_{1n}x_n, \\ y_2 = &\quad q_{21}x_1 + q_{22}x_1 + ... + q_{2n}x_n \\ &\vdots \\ y_n = &\quad q_{n1}x_1 + q_{n2}x_2 + ... + q_{nn}x_n \end{aligned}$$

such that

$$\underline{x}^tA\underline{x} = y_1^2 + ... + y_p^2 - y_{p+1}^2 - ... - y_n^2.$$

Then, $P = [q_{ij}]^{-1}$ is the matrix we are looking for.

Why? The above system of equations corresponds to the matrix equation

$$\underline{y} = Q\underline{x}, \quad Q = [q_{ij}] \in \mathrm{GL}_n(\mathbb{R}),$$

which we can consider as a change of coordiante transformation $P_{\mathcal{B}\leftarrow\mathcal{S}^{(n)}}$ from the standard basis $\mathcal{S}^{(n)} \subset \mathbb{R}^n$ to a basis $\mathcal{B}$ (we consider $\underline{x}$ to be the $\mathcal{S}^{(n)}$-coordinate vector of the corresponding element of $\mathbb{R}^n$). Then, we see that

$$(P\underline{y})^tA(P\underline{y}) = \underline{x}^tA\underline{x} = y_1^2 + ... + y_p^2 - y_{p+1}^2 - ... - y_n^2,$$

where $P = Q^{-1}$. As

$$\underline{y}^tP^tAP\underline{y} = (P\underline{y})^tA(P\underline{y}) = y_1^2 + ... + y_p^2 - y_{p+1}^2 - ... - y_n^2 = \underline{y}^t \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{bmatrix} \underline{y},$$

we see that P^tAP is of the desired form. Moreover, $\mathcal{B}$ is the required basis.

It is better to indicate this method through an example.

[62] Why?

[63] The assignment $\underline{x} \mapsto \underline{x}^tA\underline{x}$ is called a *quadratic form*. The study of quadratic forms and their properties is primarily determined by the symmetric bilinear forms defined by A.

Example 3.2.9. 1. Let

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

so that A is symmetric and invertible. Consider the column vector of variable $\underline{x}$ as above. Then, we have

$$\underline{x}^t A \underline{x} = x_1^2 + 2x_2^2 - x_4^2 - 2x_1x_3 + 4x_1x_4 + 2x_2x_3 - 4x_2x_4.$$

Let's complete the square with respect to x_1: we have

$$\begin{aligned} & x_1^2 + 2x_2^2 - x_4^2 - 2x_1x_3 + 4x_1x_4 + 2x_2x_3 - 4x_2x_4 \\ = & x_1^2 - 2x_1(x_3 - 2x_4) + (x_3 - 2x_4)^2 - (x_3 - 2x_4)^2 + 2x_2^2 - x_4^2 + 2x_2x_3 - 4x_2x_4 \\ = & (x_1 - (x_3 - 2x_4))^2 + 2x_2^2 - x_3^2 - 5x_4^2 + 2x_2x_3 - 4x_2x_4 + 4x_3x_4 \end{aligned}$$

Now we set

$$y_1 = x_1 - x_3 + 2x_4.$$

Then, complete the square with respect to the remaining x_2 terms: we have

$$\begin{aligned} & y_1^2 + 2x_2^2 - x_3^2 - 5x_4^2 + 2x_2x_3 - 4x_2x_4 + 4x_3x_4 \\ = & y_1^2 + 2(x_2^2 + x_2(x_3 - 2x_4) + \frac{1}{4}(x_3 - 2x_4)^2) - \frac{1}{2}(x_3 - 2x_4)^2 - x_3^2 - x_4^2 - 4x_3x_4 \\ = & y_1^2 + 2(x_2 + \frac{1}{2}(x_3 - 2x_4))^2 - \frac{3}{2}x_3^2 - 7x_4^2 - 2x_3x_4 \end{aligned}$$

Now we set

$$y_2 = \sqrt{2}\,(x_2 + \tfrac{1}{2}x_3 - x_4).$$

We obtain

$$x_1^2 + 2x_2^2 - x_4^2 - 2x_1x_3 + 4x_1x_4 + 2x_2x_3 - 4x_2x_4 = y_1^2 + y_2^2 - \frac{3}{2}x_3^2 - 7x_4^2 - 2x_3x_4.$$

Completing the square with respect to x_3 we obtain

$$\begin{aligned} & y_1^2 + y_2^2 - \frac{3}{2}x_3^2 - 7x_4^2 - 2x_3x_4 \\ = & y_1^2 + y_2^2 - \frac{3}{2}(x_3^2 + \frac{14}{3}x_3x_4 + \frac{49}{9}x_4^2) + \frac{49}{6}x_4^2 \\ = & y_1^2 + y_2^2 - \frac{3}{2}(x_3 + \frac{7}{3}x_4)^2 + \frac{49}{6}x_4^2. \end{aligned}$$

Then, set

$$\begin{aligned} y_3 = & \sqrt{\tfrac{3}{2}}(x_3 + \tfrac{7}{3}x_4), \\ y_4 = & \tfrac{7}{\sqrt{6}}x_4 \end{aligned}.$$

So, if we let

$$Q = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} & -\sqrt{2} \\ 0 & 0 & \sqrt{\frac{3}{2}} & \frac{7}{\sqrt{6}} \\ 0 & 0 & 0 & \frac{7}{\sqrt{6}} \end{bmatrix},$$

then we have

$$\underline{y} = Q\underline{x}.$$

Hence, if we define $P = Q^{-1}$, then we have that

$$P^tAP = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & 1 \end{bmatrix}.$$

Hence, we have that $p = 3, q = 1$ and that if $B_A \in \mathrm{Bil}_{\mathbb{R}}(\mathbb{R}^4)$ then

$$\mathrm{sig}(B_A) = 3 - 1 = 2.$$

2. Consider the matrix

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

which is symmetric and invertible. Consider the column vector of variables $\underline{x}$ as before. Then, we have

$$\underline{x}^tA\underline{x} = -x_1^2 + 2x_2x_3.$$

Proceeding as before, we 'complete the square' with respect to x_2 (we don't need to complete the square for x_1): we have

$$\begin{aligned} & -x_1^2 + 2x_2x_3 \\ = & -x_1^2 + \frac{1}{2}(x_2 + x_3)^2 - \frac{1}{2}(x_2 - x_3)^2 \end{aligned}$$

Hence, if we let

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= \tfrac{1}{\sqrt{2}}(x_2 + x_3) \\ y_3 &= \tfrac{1}{\sqrt{2}}(x_2 - x_3) \end{aligned}$$

then we have

$$\underline{x}^tA\underline{x} = -y_1^2 + y_2^2 - y_3^2.$$

Furthermore, if we let

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix},$$

and defined $P = Q^{-1}$, then

$$P^tAP = \begin{bmatrix} -1 & & \\ & 1 & \\ & & -1 \end{bmatrix}.$$

Hence, $p = 1, q = 2$ and

$$\mathrm{sig}(B_A) = -1.$$

3.3 Euclidean spaces

Throughout this section V will be a finite dimensional $\mathbb{R}$-vector space and $\mathbb{K} = \mathbb{R}$.

Definition 3.3.1. Let $B \in \mathrm{Bil}_{\mathbb{R}}(V)$ be a symmetric bilinear form. We say that B is an *inner product on V* if B satisfies the following property:

$$B(v, v) \geq 0, \quad \text{for every } v \in V, \text{ and } \quad B(v, v) = 0 \Leftrightarrow v = 0_V.$$

If $B \in \mathrm{Bil}_{\mathbb{K}}(V)$ is an inner product on V then we will write

$$\langle u, v \rangle \stackrel{def}{=} B(u, v).$$

Remark 3.3.2. Suppose that $\langle , \rangle$ is an inner product on V. Then, we have the following properties:

i) $\langle \lambda u + v, w \rangle = \lambda \langle u, w \rangle + \langle v, w \rangle$, for every $u, v, w \in V, \lambda \in \mathbb{K}$,

ii) $\langle u, \lambda v + w \rangle = \lambda \langle u, v \rangle + \langle u, w \rangle$, for every $u, v, w \in V, \lambda \in \mathbb{K}$,

iii) $\langle u, v \rangle = \langle v, u \rangle$, for every $u, v \in V$.

iv) $\langle v, v \rangle \geq 0$, for every $v \in V$, with equality precisely when $v = 0_V$.

Property *iv*) is often referred to as the **positive-definite** property of an inner product.

Definition 3.3.3. A *Euclidean space, or inner product space,* is a pair $(V, \langle , \rangle)$, where V is a finite dimensional $\mathbb{R}$-vector space and $\langle , \rangle$ is an inner product on V.

Given an inner product space $(V, \langle , \rangle)$ we define the *norm function on V (with respect to $\langle , \rangle$)* to be the function

$$||.|| : V \to \mathbb{R}_{\geq 0} \; ; \; v \mapsto ||v|| = \sqrt{\langle v, v \rangle}.$$

For any $v \in V$ we define the *length of v (with respect to $\langle , \rangle$)* to be $||v|| \in \mathbb{R}_{\geq 0}$.

Let $(V_1, \langle , \rangle_1), (V_2, \langle , \rangle_2)$ be inner product spaces. Then, we say that a linear morphism

$$f : V_1 \to V_2,$$

is a *Euclidean morphism* if, for every $u, v \in V_1$ we have

$$\langle u, v \rangle_1 = \langle f(u), f(v) \rangle_2.$$

A Euclidean morphism whose underlying linear morphism is an isomorphism is called a *Euclidean isomorphism*.

If $f : (V, \langle , \rangle) \to (V, \langle , \rangle)$ is a Euclidean morphism such that the domain and codomain are the same Euclidean space, then we say that f is an *orthogonal morphism, or an orthogonal transformation*. We denote the set of all orthgonal transformations of $(V, \langle , \rangle)$ by $O(V, \langle , \rangle)$, or simply $O(V)$ when there is no confusion.

Example 3.3.4. 1. We define *n-dimensional Euclidean space,* denoted $\mathbb{E}^n$, to be the Euclidean space $(\mathbb{R}^n, \cdot)$, where $\cdot$ is the usual 'dot product' from analytic geometry: that is, for $\underline{x}, \underline{y} \in \mathbb{R}^n$ we have

$$\underline{x} \cdot \underline{y} \stackrel{def}{=} \underline{x}^t \underline{y} = x_1 y_1 + ... + x_n y_n.$$

It easy to check that $\cdot$ is bilinear and symmetric and, moreover, we have

$$\underline{x} \cdot \underline{x} = \underline{x}^t \underline{x} = x_1^2 + ... + x_n^2 \geq 0,$$

with equality precisely when $\underline{x} = \underline{0}$.

Given $\underline{x} \in \mathbb{E}^n$, the length of $\underline{x}$ is

$$||\underline{x}|| = \sqrt{x_1^2 + ... + x_n^2}.$$

2. Consider the symmetric bilinear form $B_A \in \mathrm{Bil}_{\mathbb{R}}(\mathbb{R}^3)$ where

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

Then, you can check that

$$\underline{x} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{R}^3,$$

has the property that

$$B_A(\underline{x}, \underline{x}) = -2 < 0,$$

so that B_A is not an inner product on $\mathbb{R}^3$.

3. Let $B_A \in \mathrm{Bil}_{\mathbb{R}}(\mathbb{R}^4)$ be the symmetric bilinear form defined by

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

Then, B_A is an inner product: indeed, let $\underline{x} \in \mathbb{R}^3$. Then, we have

$$B_A(\underline{x}, \underline{x}) = x_1^2 + 2x_1x_2 + 2x_2^2 + 2x_3^2 + 2x_3x_4 + x_4^2 = (x_1 + x_2)^2 + x_2^2 + x_3^2 + (x_3 + x_4)^2 \geq 0,$$

and we have $B_A(\underline{x}, \underline{x}) = 0$ precisely when

$$x_1 + x_2 = 0,\ x_2 = 0,\ x_3 = 0,\ x_3 + x_4 = 0,$$

so that $x_1 = x_2 = x_3 = x_4 = 0$ and $\underline{x} = 0$.

With respect to this inner product, the vector

$$\underline{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^4,$$

has length

$$||\underline{x}|| = \sqrt{\langle \underline{x}, \underline{x} \rangle} = \sqrt{2}.$$

Hence, $(\mathbb{R}^4, B_A)$ is a Euclidean space.

4. In fact, a symmetric bilinear form B on an n-dimensional $\mathbb{R}$-vector space V is an inner product precisely when $\mathrm{sig}(B) = n$.[64]

5. Consider the linear morphism $T_A \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$, where

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Then, T_A is an orthogonal transformation of $\mathbb{E}^2$: indeed, for any $\underline{x}, \underline{y} \in \mathbb{R}^2$, we have

$$T_A(\underline{x}) \cdot T_A(\underline{y}) = (A\underline{x})^t(A\underline{y}) = \underline{x}^t A^t A \underline{y} = \underline{x}^t \underline{y} = \underline{x} \cdot \underline{y},$$

since $A^{-1} = A^t$.

This example highlights a more general property of orthogonal transformations of $\mathbb{E}^n$ to be discussed later:

$$A \in O(\mathbb{E}^n) \textbf{ if and only if } A^{-1} = A^t.$$[65]

6. If $(V, \langle , \rangle)$ is a Euclidean space then id_V is always an orthogonal transformation.

Remark 3.3.5. 1. A Euclidean space is simply a $\mathbb{R}$-vector space V equipped with an inner product. This means that it is possible for the same $\mathbb{R}$-vector space V to have two distinct Euclidean space structures (ie, we can equip the same $\mathbb{R}$-vector space with two distinct inner products). However, as we will see shortly, given a $\mathbb{R}$-vector space V there is *essentially* only one Euclidean space structure on V: this means that we can find a Euclidean isomorphism between the two distinct Euclidean space structures on V.

[64] This is shown in a few paragraphs.

2. It is important to remember that the norm function $||.||$ is **not linear**. In fact, the norm function is **not additive**: indeed, let $v \in V$ be nonzero. Then,

$$0 = ||0_v|| = ||v + (-v)||,$$

so that if $||.||$ were additive then we would have $||v|| + ||-v|| = 0$, for every $v \in V$. As $||v||, ||-v|| \geq 0$ then we would have that

$$||v|| = ||-v|| = 0, \text{ for every } v \in V.$$

That is, every $v \in V$ would have length 0. However, the only $v \in V$ that can have length 0 is $v = 0_V$.

Moreover, for any $v \in V, \lambda \in \mathbb{K}$, we have

$$||\lambda v|| = |\lambda| ||v||.$$

Theorem 3.3.6. *Let $(V, \langle , \rangle)$ be a Euclidean space. Then,*

a) for any $u, v \in V$ we have

$$||u + v|| \leq ||u|| + ||v||. \qquad \textit{(triangle inequality)}$$

b) $||v|| = 0$ if and only if $v = 0_V$.

c) if $\langle u, v \rangle = 0$ then

$$||u||^2 + ||v||^2 = ||u + v||^2. \qquad \textit{(Pythagoras' theorem)}$$

d) for any $u, v \in V$ we have

$$|\langle u, v \rangle| \leq ||u|| ||v||. \qquad \textit{(Cauchy-Schwarz inequality)}$$

Proof: Left as an exercise for the reader. □

We will now show that there is essentially only one Euclidean space structure that we can give an arbitrary finite dimensional $\mathbb{R}$-vector space. Moreover, this Euclidean space structure is well-known to us all.

Lemma 3.3.7. *Suppose that $\langle , \rangle$ is an inner product on V. Then, $\langle , \rangle \in \text{Bil}_{\mathbb{R}}(V)$ is nondegenerate.*

Proof: We need to show the following property of $\langle , \rangle$:

$$\text{if } v \in V \text{ is such that } \langle u, v \rangle = 0, \text{ for every } u \in V, \text{ then } v = 0_V.$$

So, suppose that $v \in V$ is such that $\langle u, v \rangle = 0$, for every $u \in V$. In particular, we must have

$$\langle v, v \rangle = 0 \implies v = 0_V,$$

by the defining property of an inner product (Remark 3.3.2, iv)). Hence, $\langle , \rangle$ is nondegenerate. □

Hence, using Sylvester's law of inertia (Theorem 3.2.6), we know that for a Euclidean space $(V, \langle , \rangle)$ there is an ordered basis $\mathcal{B} \subset V$ such that

$$[\langle , \rangle]_{\mathcal{B}} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}, \quad \text{where } d_i \in \{1, -1\}, \ n = \dim V.$$

Moreover, since $\langle , \rangle$ is an inner product we must have that $\text{sig}(\langle , \rangle) = n$: indeed, we have

$$\text{sig}(\langle , \rangle) = p - q \in \{-n, -(n-1), \ldots, n-1, n\},$$

so that $\text{sig}(\langle , \rangle) = n$ if and only if $q = 0$, so that there are no -1s appearing on the diagonal of $[\langle , \rangle]_{\mathcal{B}}$. If some $d_i = -1$ then we would have

$$0 \leq \langle b_i, b_i \rangle = -1,$$

which is impossible. Hence, we must have $d_1 = d_2 = \ldots = d_n = 1$, so that

$$[\langle , \rangle]_{\mathcal{B}} = I_n.$$

Theorem 3.3.8 (Classification of Euclidean spaces)**.** *Let $(V, \langle , \rangle)$ be a Euclidean space, $n = \dim V$. Then, there is a Euclidean isomorphism*

$$f : (V, \langle , \rangle) \to \mathbb{E}^n.$$

Proof: Let $\mathcal{B} \subset V$ be an ordered basis such that

$$[\langle , \rangle]_{\mathcal{B}} = I_n.$$

Then, let

$$f = [-]_{\mathcal{B}} : V \to \mathbb{R}^n,$$

be the $\mathcal{B}$-coordinate morphism. Then, this is an isomorphism of $\mathbb{R}$-vector spaces so that we need only show that

$$\langle u, v \rangle = [u]_{\mathcal{B}} \cdot [v]_{\mathcal{B}},$$

for every $u, v \in V$. Now, let $u, v \in V$ and suppose that

$$u = \sum_{i=1}^{n} \lambda_i b_i, \ u = \sum_{j=1}^{n} \mu_j b_j.$$

Then,

$$\langle u, v \rangle = \langle \sum_{i=1}^{n} \lambda_i b_i, \sum_{j=1}^{n} \mu_j b_j \rangle = \sum_{i,j} \lambda_i \mu_j \langle b_i, b_j \rangle = \sum_{i=1}^{n} \lambda_i \mu_i,$$

where we have used bilinearity of $\langle , \rangle$ and that

$$\langle b_i, b_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Now, we also have

$$[u]_{\mathcal{B}} \cdot [v]_{\mathcal{B}} = [u]_{\mathcal{B}}^t [v]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \mu_i = \langle u, v \rangle,$$

and the result follows. □

Corollary 3.3.9. *Let $(V_1, \langle , \rangle_1), (V_2, \langle , \rangle_2)$ be Euclidean spaces. Then, if $\dim V_1 = \dim V_2$ then $(V_1, \langle , \rangle_1)$ and $(V_2, \langle , \rangle_2)$ are Euclidean-isomorphic.*

Proof: By Theorem 3.3.8 we have Euclidean isomorphisms

$$f_1 : (V_1, \langle , \rangle_1) \to \mathbb{E}^n, \ f_2 : (V_2, \langle , \rangle_2) \to \mathbb{E}^n.$$

Then, as the composition of two Euclidean isomorphisms is again a Euclidean isomorphism[66] then we obtain an isomorphism

$$f_2^{-1} \circ f_1 : (V_1, \langle , \rangle_1) \to (V_2, \langle , \rangle_2).$$

□

In fact, the condition defining a Euclidean morphism (not necessarily an isomorphism) is extremely strong: if $(V_1, \langle , \rangle_1)$ and $(V_2, \langle , \rangle_2)$ are Euclidean spaces and $f : V_1 \to V_2$ is a Euclidean morphism, then it is easy to check that we must have

$$||v|| = ||f(v)||, \text{ for every } v \in V,$$

so that f is *length preserving*. If you think about what this means geometrically then we obtain that

'Euclidean morphisms are always injective'

since no nonzero vector can be mapped to 0_{V_2} by f. As a consequence, we obtain

[66] Check this.

Proposition 3.3.10. *Let $(V_1, \langle , \rangle_1), (V_2, \langle , \rangle_2)$ be Euclidean spaces of the same dimension. Then, if there exists a Euclidean morphism*

$$f : V_1 \to V_2,$$

it must automatically be a Euclidean isomorphism.

Corollary 3.3.11. *Let $(V, \langle , \rangle)$ be a Euclidean space. Then, every Euclidean endomorphism*

$$f : V \to V$$

is an orthogonal transformation (= Euclidean isomorphism). Hence, we have

$$O(V) = \{f \in \mathrm{End}_{\mathbb{R}}(V) \mid f \text{ is Euclidean}\}.$$

Definition 3.3.12. The set of orthogonal transformations of $\mathbb{E}^n$ is called the *orthogonal group of size n* and is denoted $O(n)$.

Suppose that $g \in O(n)$ is an orthogonal transformation of $\mathbb{E}^n$ and identify g with its standard matrix $[g]_{\mathcal{S}^{(n)}}$. Then, we must have, for every $\underline{x}, \underline{y} \in \mathbb{R}^n$, that

$$\underline{x} \cdot \underline{y} = (g\underline{x}) \cdot (g\underline{y}) = (g\underline{x})^t (g\underline{y}) = \underline{x}^t g^t g \underline{y},$$

so that

$$\underline{x}^t \underline{y} = \underline{x}^t g^t g \underline{y},$$

for every $\underline{x}, \underline{y} \in \mathbb{R}^n$. Hence, by Lemma 3.1.6, we must have that

$$g^t g = I_n.$$

Hence, we see that we can identify

$$[-]_{\mathcal{S}^{(n)}} : O(n) \to \{X \in Mat_n(\mathbb{R}) \mid X^t X = I_n\}.$$

Moreover, this identification satisfies the following properties:

- $[\mathrm{id}_{\mathbb{E}^n}]_{\mathcal{S}^{(n)}} = I_n$,
- for every $f, g \in O(n)$, $[f \circ g]_{\mathcal{S}^{(n)}} = [f]_{\mathcal{S}^{(n)}} [g]_{\mathcal{S}^{(n)}}$.

Hence, the correspondence

$$[-]_{\mathcal{S}^{(n)}} : O(n) \to \{X \in Mat_n(\mathbb{R}) \mid X^t X = I_n\},$$

is an *isomorphism of groups*.

From now on, when we consider orthogonal transformations $g \in O(n)$ we will identify g with its standard matrix. Then, the previous discussion shows that $g \in \mathrm{GL}_n(\mathbb{R})$ and $g^t g = I_n$.

Let's think a little bit more about the condition

$$A^t A = I_n,$$

for $A \in Mat_n(\mathbb{R})$.

i) If A is such that $A^t A = I_n$ then we must have that $\det(A)^2 = 1$, since $\det(A) = \det(A^t)$. In particular, $\det(A) \in \{1, -1\}$[67] so that $A \in \mathrm{GL}_n(\mathbb{R})$: the inverse of A is $A^{-1} = A^t$. Furthermore, this implies that we must have

$$AA^t = AA^{-1} = I_n,$$

so that

[67] It is NOT true that if $A \in \mathrm{GL}_n(\mathbb{R})$ such that $\det A = 1$ then $A \in O(n)$. For example, consider

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

Then, it is not the case that $A^t A = I_2$ so that $A \notin O(2)$.

$$\boxed{A^tA = I_n \textbf{ if and only if } AA^t = I_n.}$$

ii) Let us write

$$A = [a_1 \ \cdots \ a_n],$$

so that the i^{th} column of A is a_i. Then, as $A \in \mathrm{GL}_n(\mathbb{R})$ we have that $\{a_1, \dots, a_n\}$ is linearly independent and defines a basis of $\mathbb{R}^n$. Moreover, as the i^{th} row of A^t is a_i^t, then the condition $A^tA = I_n$ implies that

$$a_i \cdot a_j = a_i^t a_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} .$$

In particular, we see that **each column of A has length** 1[68] (with respect to the inner product $\cdot$), and that the $\cdot$-complement of a_i is precisely

$$\mathrm{span}_{\mathbb{R}}\{a_j \mid j \neq i\}.$$

iii) A matrix $A \in Mat_n(\mathbb{R})$ such that

$$A^tA = I_n,$$

will be called an *orthogonal matrix*.

iv) A matrix $A \in Mat_n(\mathbb{R})$ is an orthogonal matrix if and only if for every $\underline{x}, \underline{y} \in \mathbb{R}^n$ we have

$$(A\underline{x}) \cdot (A\underline{y}) = \underline{x} \cdot \underline{y}.$$

We can interpret this result using the slogan

$$\boxed{\textbf{'orthogonal transformations are the 'rigid' transformations'}}$$

Example 3.3.13. 1. Let $\theta \in \mathbb{R}$ and consider the matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in Mat_2(\mathbb{R}).$$

Then, you may know already that R_θ corresponds to the 'rotate by θ counterclockwise' morphism of $\mathbb{R}^2$. If not, then this is easily seen: since R_θ defines a linear transformation of $\mathbb{R}^2$ we need only determine what happens to the standard basis of $\mathbb{R}^2$. We have

$$R_\theta e_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \ R_\theta e_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

and by considering triangles and the unit circle the result follows.

You can check easily that

$$R_\theta^t R_\theta = I_2,$$

so that $\mathbb{R}_\theta \in O(2)$.

In fact, it can be shown that every orthogonal transformation of $\mathbb{R}^2$ that has determinant 1 is of the form R_θ, for some θ. Moreover, every orthogonal transformation of $\mathbb{R}^2$ is of one of the following forms:

$$R_\theta, \ \text{ or } \ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R_\theta.$$

[68] Similarly, we obtain that each row must have legnth 1

3.3.1 Orthogonal complements, bases and the Gram-Schmidt process

Definition 3.3.14. Let $(V, \langle , \rangle)$ be a Euclidean space, $S \subset V$ a nonempty subset. We define the *orthogonal complement of* S, denoted $S^\perp$, to be the $\langle , \rangle$-complement of S defined in Definition 3.1.15. Hence,

$$S^\perp = \{v \in V \mid \langle v, s\rangle = 0, \text{ for every } s \in S\} = \{v \in V \mid \langle s, v\rangle = 0, \text{ for every } s \in S\}.$$

$S^\perp$ is a subspace of V, for any subset $S \subset V$.[69]

Proposition 3.3.15. *Let $(V, \langle , \rangle)$ be a Euclidean space and $U \subset V$ a subspace. Then,*

$$V = U \oplus U^\perp.$$

Proof: We know that $\dim V = \dim U + \dim U^\perp$ by Proposition 3.1.17. Hence, if we show that $U \cap U^\perp = \{0_V\}$ then we must have

$$V = U + U^\perp = U \oplus U^\perp.$$[70]

Assume that $v \in U \cap U^\perp$. Then, $v \in U$ and $v \in U^\perp$ so that

$$0 = \langle v, v\rangle \implies v = 0_V,$$

since $\langle , \rangle$ is an inner product. The result follows. □

Remark 3.3.16. 1. Just as we have shown before, we have

$$S^\perp = (\text{span}_{\mathbb{R}} S)^\perp.$$

2. If we are thinking geometrically (as we should do whenever we are given any Euclidean space V) then we see that the orthogonal complement $U^\perp$ of a subspace U is the subspace of V which is 'perpendicular' to U. For example, consider the Euclidean space $\mathbb{E}^3$, U is the 'x-axis', which we'll denote L. Then, the subspace that is perpendicular to the x-axis is the $x = 0$-plane Π. Indeed, we have

$$L = \left\{ \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3 \right\}, \text{ and } \Pi = \left\{ \begin{bmatrix} 0 \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \right\}.$$

It is easy to check that $\Pi = L^\perp$.[71]

Definition 3.3.17. Let $(V, \langle , \rangle)$ be a Euclidean space, $U \subset V$ a subspace and $v \in V$. Then, we define the *projection of v onto U* to be the vector $\text{proj}_U v$ defined as follows: using Proposition 3.3.15 we know that $V = U \oplus U^\perp$ so that there exists (unique!) $u \in U, z \in U^\perp$ such that $v = u + z$. Then, we define

$$\text{proj}_U v \stackrel{def}{=} u \in U.$$

Remark 3.3.18. In fact, the assignment

$$\text{proj}_U : V \to U \; ; \; v \mapsto \text{proj}_U v,$$

is precisely the 'projection onto U' morphism defined earlier. As a consequence we see that

$$\text{proj}_U(v + v') = \text{proj}_U v + \text{proj}_U v', \text{ and } \text{proj}_U \lambda v = \lambda \, \text{proj}_U v.$$

We can think of $\text{proj}_U v$ in more geometric terms.

[69] Check this.
[70] This follows from the dimension formula.
[71] Do this!

Proposition 3.3.19. *Let $(V, \langle , \rangle)$ be a Euclidean space, $U \subset V$ a subspace and $v \in V$. Then,* $\text{proj}_U v \in U$ *is the unique vector in U such that*

$$||\,\text{proj}_U v - v|| \leq ||u - v||,\ u \in U.$$

Hence, we can say that $\text{proj}_U v$ *is* the closest vector to v in U.

Proof: Let $u \in U$. Then, we have

$$(\text{proj}_U v - v) + (u - \text{proj}_U v) = (u - v),$$

and, since $\text{proj}_U v - v \in U^\perp$ (Definition 3.3.17) and $u - \text{proj}_U v \in U$, then

$$||u - v||^2 = ||\,\text{proj}_U v - v||^2 + ||u - \text{proj}_U v||^2 \geq ||\,\text{proj}_U v - v||^2,$$

where we have used Pythagoras' theorem (Theorem 3.3.6). Hence, we have

$$||u - v|| \geq ||\,\text{proj}_U v - v||, \quad \text{for any } u \in U.$$

Suppose that $w \in U$ is such that

$$||w - v|| \leq ||u - v||, \quad \text{for any } u \in U.$$

This implies that we must have

$$||w - v|| = ||\,\text{proj}_U v - v||,$$

by what we have just shown.

Now, using Pythagoras' theorem, and that $v - \text{proj}_U v \in U^\perp, \text{proj}_U v - w \in U$, we obtain

$$||v - w||^2 = ||v - \text{proj}_U v + \text{proj}_U v - w||^2 = ||v - \text{proj}_U v||^2 + ||\,\text{proj}_U v - w||^2 \implies ||\,\text{proj}_U v - w||^2 = 0,$$

and $\text{proj}_U v = w$. Hence, $\text{proj}_U v$ is the unique element of U satisfying the above inequality. □

Example 3.3.20. Consider the Euclidean space $\mathbb{E}^2$ and let $L \subset \mathbb{R}^2$ be a line through the origin. Suppose that $v \in \mathbb{R}^2$ is an arbitrary vector. What does $\text{proj}_L v$ look like geometrically?

Using Proposition 3.3.19 we know that $w = \text{proj}_L v \in L$ is the unique vector in L that is closest to v.

- if $v \in L$ then $\text{proj}_L v = v$, as $v \in L$ is the closest vector v (trivially).
- if $v \notin L$ then consider the line L' perpendicular to L and for which the endpoint of the vector v lies on L' (so it might not be the case that L' is a line through the origin). The point of intersection $L \cap L'$ defines the vector $\text{proj}_L v$.

In fact, it is precisely this geometric intuition that guides the definition of $\text{proj}_L v$: we have defined $\text{proj}_L v \in L$ as the unique vector such that

$$v = \text{proj}_L v + z,\ z \in L^\perp.$$

Definition 3.3.21. Let $(V, \langle , \rangle)$ be a Euclidean space. We say that a subset $S \subset V$ is an *orthogonal set* if, for every $s, t \in S$, $s \neq t$, we have

$$\langle s, t \rangle = 0.$$

Lemma 3.3.22. *Let $S \subset V$ be an orthogonal set of nonzero vectors. Then, S is linearly independent.*

Proof: Left as en exercise for the reader. □

Lemma 3.3.23. *Let $S = \{s_1, \ldots, s_k\} \subset V$ be an orthogonal set and such that S contains only nonzero vectors. Then, for any $v \in V$, we have*

$$\text{proj}_{\text{span}_\mathbb{R} S}\, v = \frac{\langle v, s_1 \rangle}{\langle s_1, s_1 \rangle} s_1 + \ldots + \frac{\langle v, s_k \rangle}{\langle s_k, s_k \rangle} s_k.$$

Proof: Since S is linearly independent we have that S forms a basis of $\text{span}_{\mathbb{R}} S$. Hence, for any $v \in V$, we can write

$$\text{proj}_{\text{span}_{\mathbb{R}} S} v = \lambda_1 s_1 + ... + \lambda_k s_k,$$

for unique $\lambda_1, ..., \lambda_k \in \mathbb{R}$. Hence, for each $i = 1, ..., k$, we have

$$\langle \text{proj}_{\text{span}_{\mathbb{R}} S} v, s_i \rangle = \lambda_i \langle s_i, s_i \rangle,$$

using that S is orthogonal. Hence, we have that

$$\lambda_i = \frac{\langle \text{proj}_{\text{span}_{\mathbb{R}} S} v, s_i \rangle}{\langle s_i, s_i \rangle}.$$

Now, since $v - \text{proj}_{\text{span}_{\mathbb{R}} S} v \in (\text{span}_{\mathbb{R}} S)^{\perp}$ we see that, for each i,

$$0 = \langle v - \text{proj}_{\text{span}_{\mathbb{R}} S} v, s_i \rangle = \langle v, s_i \rangle - \langle \text{proj}_{\text{span}_{\mathbb{R}} S} v, s_i \rangle \implies \langle v, s_i \rangle = \langle \text{proj}_{\text{span}_{\mathbb{R}} S}, s_i \rangle.$$

The result follows. □

Definition 3.3.24. Let $(V, \langle , \rangle)$ be a Euclidean space. A basis $\mathcal{B} \subset V$ is called an *orthogonal basis* if it is an orthogonal set.

An orthogonal basis $\mathcal{B}$ is called *orthonormal* if, for every $b \in \mathcal{B}$, we have $||b|| = 1$.

Remark 3.3.25. 1. Recall that we defined an orthogonal matrix $A \in Mat_n(\mathbb{R})$ to be a matrix such that

$$A^t A = I_n.$$

The remarks at the end of the previous section imply that **the columns of an orthogonal matrix define an orthonormal basis**.

2. Not every basis in a Euclidean space is an orthogonal basis: for example, consider the Euclidean space $\mathbb{E}^2$. Then,

$$\mathcal{B} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = (b_1, b_2),$$

is a basis of $\mathbb{R}^2$ but we have

$$b_1 \cdot b_2 = 1 \neq 0.$$

3. It is not true that any orthogonal set $E \subset V$ defines an orthogonal basis of $\text{span}_{\mathbb{R}} E$: for example, let $v \in V$ be nonzero and consider the subset $E = \{0_V, v\}$. Then, E is orthogonal[72] but E is not a basis, as E is a linearly dependent set. However, if E contains nonzero vectors and is orthogonal then E is an orthogonal basis of $\text{span}_{\mathbb{R}} E$, by Lemma 3.3.22.

At first glance it would appear to be quite difficult to determine an orthogonal (or orthonormal) basis of V. This is essentially the same problem as coming up with an orthogonal matrix. Moreover, it is hard to determine whether orthogonal bases even exist!

It is a quite remarkable result that given **ANY** basis $\mathcal{B}$ of a Euclidean space $(V, \langle , \rangle)$ we can determine an **orthonormal** basis $\mathcal{B}'$ of V. This is the **Gram-Schmidt process**.

Theorem 3.3.26 (Gram-Schmidt process)**.** *Let $(V, \langle , \rangle)$ be a Euclidean space, $\mathcal{B} = (b_1, ..., b_n) \subset V$ an arbitrary ordered basis of V. Then, there exists an orthonormal basis $\mathcal{B}' = (b'_1, ..., b'_n) \subset V$.*

Proof: Consider the following algorithm: define

$$c_1 = b_1.$$

We inductively define c_i: for $2 \leq i \leq n$ define

$$c_i = b_i - \text{proj}_{E_{i-1}} b_i,$$

where $E_{i-1} \overset{def}{=} \text{span}_{\mathbb{R}}\{c_1, ..., c_{i-1}\}$.

If $i < j$ then

$$\langle c_i, c_j \rangle = 0,$$

since $c_j \in E_{j-1}^{\perp}$ by construction[73], and $c_i \in E_{j-1}$.

[72] Check this.

[73] Think about why this is true. What is the definition of c_j?

Hence, $\mathcal{C} = (c_1, \dots, c_n)$ is an orthogonal basis. To obtain an orthonormal basis $\mathcal{B}' = (b_1', \dots, b_n')$ given an orthogonal basis $\mathcal{C}$, we simply set

$$b_i' = \frac{c_i}{||c_i||}.$$

Then, we have

$$||b_i'|| = 1,$$

and $\mathcal{B}'$ is an orthonormal basis. □

Corollary 3.3.27. *Let $(V, \langle , \rangle)$ be a Euclidean space, $E \subset V$ an orthogonal set consisting of nonzero vectors. Then, E can be extended to an orthogonal basis of V.*

Proof: Left as an exercise for the reader. □

Remark 3.3.28. 1. Let's illuminate exactly what we have done in the proof of Theorem 3.3.26, making use of Lemma 3.3.23.

Let $\mathcal{B} = (b_1, \dots, b_n)$ be **any** basis. We can organise the algorithm from Theorem 3.3.26 into a table

$$\begin{aligned} c_1 &= b_1 \\ c_2 &= b_2 - \tfrac{\langle b_2, c_1 \rangle}{\langle c_1, c_1 \rangle} c_1 \\ c_3 &= b_3 - \tfrac{\langle b_3, c_1 \rangle}{\langle c_1, c_1 \rangle} c_1 - \tfrac{\langle b_3, c_2 \rangle}{\langle c_2, c_2 \rangle} c_2 \\ &\vdots \\ c_n &= b_n - \tfrac{\langle b_n, c_1 \rangle}{\langle c_1, c_1 \rangle} c_1 - \dots - \tfrac{\langle b_n, c_{n-1} \rangle}{\langle c_{n-1}, c_{n-1} \rangle} c_{n-1} \end{aligned}$$

Then $\mathcal{C} = (c_1, \dots, c_n)$ is an orthogonal basis of V. To obtain an orthonormal basis of V we set

$$b_i' = \frac{c_i}{||c_i||}, \text{ for each } i.$$

Then, $\mathcal{B}' = (b_1', \dots, b_n')$ is orthonormal.

In practice in can be quite painful to actually perform the Gram-Schmidt process (if $\dim V$ is large). However, it is important to know that the Gram-Schmidt process allows us to show that **orthonormal bases exist**.

2. If $\mathcal{B}$ is orthogonal to start with then the basis $\mathcal{C}$ we obtain after performing the Gram-Schmidt process is just $\mathcal{C} = \mathcal{B}$.

3. It is important to remember that **the Gram-Schmidt process depends on the inner product $\langle , \rangle$ used to define the Euclidean space $(V, \langle , \rangle)$.**

Example 3.3.29. Let $V = \mathbb{E}^2$ and consider the basis

$$\mathcal{B} = \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right).$$

Let's perform the Gram-Schmidt process to obtain an orthogonal basis $\mathcal{C} = (c_1, c_2)$ of $\mathbb{E}^2$. We have

$$\begin{aligned} c_1 &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ c_2 &= \begin{bmatrix} 2 \\ 5 \end{bmatrix} - \frac{\begin{bmatrix} 2 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} - \tfrac{2.1+5.(-1)}{1^2+(-1)^2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \tfrac{3}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 7/2 \end{bmatrix} \end{aligned}$$

Then, you can check that

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 7/2 \\ 7/2 \end{bmatrix} = 7/2 - 7/2 = 0.$$

If we define

$$b_1' = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \; b_2' = \frac{2}{7\sqrt{2}} \begin{bmatrix} 7/2 \\ 7/2 \end{bmatrix},$$

we have that $\mathcal{B}' = (b_1', b_2')$ is orthonormal.

Corollary 3.3.30 (QR factorisation)**.** *Let $A \in \mathrm{GL}_n(\mathbb{R})$. Then, there exists an orthogonal matrix $Q \in O(n)$ and a upper-triangular matrix R such that*

$$A = QR.$$

Proof: This is a simple restatement of the Gram-Schmidt process. Suppose that

$$A = [a_1 \ \cdots \ a_n].$$

Then $\mathcal{B} = (a_1, \dots, a_n)$ is an ordered basis of $\mathbb{R}^n$. Apply the Gram-Schmidt process (with respect to the dot product) to obtain an orthonormal basis $\mathcal{B}' = (b_1, \dots, b_n)$ as above. Then, we have

$$\begin{aligned} b_1 &= \tfrac{1}{r_1} a_1 \\ b_2 &= \tfrac{1}{r_2} (a_2 - (a_2 \cdot b_1) b_1) \\ &\vdots \\ b_n &= \tfrac{1}{r_n} (a_n - (a_n \cdot b_1) b_1 - \dots - (a_n \cdot b_{n-1}) b_{n-1}) \end{aligned}$$

where $r_i \in \mathbb{R}_{>0}$ is the length of the c_i vectors from the Gram-Schmidt process. We have also slightly modified the Gram-Schmidt process (in what way?) but you can check that $(b_1, \dots, b_n)$ is an orthonormal basis.[74]

By moving all b_i terms to the left hand side of the above equations we obtain the table

$$\begin{aligned} r_1 b_1 &= a_1 \\ (a_2 \cdot b_1) b_1 + r_2 b_2 &= a_2 \\ &\vdots \\ (a_n \cdot b_1) b_1 + \dots + (a_n \cdot b_{n-1}) b_{n-1} + r_n b_n &= a_n \end{aligned}$$

and we can rewrite these equations using matrices: if

$$Q = [b_1 \ \cdots \ b_n] \in O(n), \quad R = \begin{bmatrix} r_1 & a_2 \cdot b_1 & a_3 \cdot b_1 & \cdots & a_n \cdot b_1 \\ 0 & r_2 & a_3 \cdot b_2 & \cdots & a_n \cdot b_2 \\ 0 & 0 & r_3 & \cdots & a_n \cdot b_3 \\ \vdots & & & \ddots & \vdots \\ 0 & & \cdots & & r_n \end{bmatrix},$$

then we see that the above equations correspond to

$$QR = A.$$

□

3.4 Hermitian spaces

In this section we will give a (very) brief introduction to the definition and fundamental properties of Hermitian forms and Hermitian spaces. A Hermitian form can be considered as a 'quasi-bilinear form' on complex vector spaces.

Definition 3.4.1. Let V be a $\mathbb{C}$-vector space. A function

$$H : V \times V \to \mathbb{C} \ ; \ (u, v) \mapsto H(u, v),$$

is called a *Hermitian form on V* if

(HF1) for any $u, v, w \in V$, $\lambda \in \mathbb{C}$,

$$H(u + \lambda v, w) = H(u, w) + \lambda H(v, w),$$

[74] Do this!

(HF2) for any $u, v \in V$,

$$H(u,v) = \overline{H(v,u)}, \qquad (\textit{Hermitian symmetric})$$

where, if $z = a + \sqrt{-1}b \in \mathbb{C}$, we define the *complex conjugate of* z to be the complex number

$$\overline{z} = a - \sqrt{-1}b \in \mathbb{C}.$$

We denote the set of all Hermitian forms on V by $\mathrm{Herm}(V)$.

Remark 3.4.2. It is a direct consequence of the above definition that if H is a Hermitian form on V we have

$$H(u, v + \lambda w) = H(u,v) + \overline{\lambda} H(v,w),$$

for any $u, v, w \in V, \lambda \in \mathbb{C}$.

We say that a Hermitian form is

'linear in the first argument, antilinear[75] in the second argument'

Definition 3.4.3. Let V be a $\mathbb{C}$-vector space, $\mathcal{B} = (b_1, \dots, b_n) \subset V$ an ordered basis and H a Hermitian form on V. Define *the matrix of* H *with respect to* $\mathcal{B}$, to be the matrix

$$[H]_{\mathcal{B}} = [a_{ij}], \quad a_{ij} = H(b_i, b_j).$$

The Hermitian symmetric property of a Hermitian form implies that

$$[H]_{\mathcal{B}} = \overline{[H]}_{\mathcal{B}}^{t},$$

where, for any matrix $A = [a_{ij}] \in Mat_{m,n}(\mathbb{C})$, we define

$$\overline{A} = [b_{ij}], \quad b_{ij} = \overline{a_{ij}}.$$

A matrix $A \in Mat_n(\mathbb{C})$ is called a *Hermitian matrix* if

$$A = \overline{A}^{t}.$$

For any $A \in Mat_n(\mathbb{C})$, we will write

$$A^h \stackrel{def}{=} \overline{A}^{t};$$

hence, a matrix $A \in Mat_n(\mathbb{C})$ is Hermitian if $A^h = A$.

Lemma 3.4.4. *For any* $A, B \in Mat_n(\mathbb{C}), \eta \in \mathbb{C}$ *we have*

- $(A+B)^h = A^h + B^h$,
- $(AB)^h = B^h A^h$,
- $(\eta A)^h = \overline{\eta} A^h$.

Lemma 3.4.5. *Let* V *be a* $\mathbb{C}$*-vector space,* $\mathcal{B} \subset V$ *an ordered basis of* V *and* H *a Hermitian form on* V*. Then, for any* $u, v \in V$*, we have*

$$H(u,v) = [u]_{\mathcal{B}}^{t} [H]_{\mathcal{B}} \overline{[v]}_{\mathcal{B}}.$$

Moreover, if $A \in Mat_n(\mathbb{C})$ *is any matrix such that*

$$H(u,v) = [u]_{\mathcal{B}}^{t} A \overline{[v]}_{\mathcal{B}},$$

for every $u, v \in V$*, then* $A = [H]_{\mathcal{B}}$.

Example 3.4.6. 1. Consider the function

$$H : \mathbb{C}^2 \times \mathbb{C}^2 \to \mathbb{C} \; ; \; (\underline{z}, \underline{w}) \mapsto z_1\overline{w}_1 + \sqrt{-1}z_2\overline{w}_1 - \sqrt{-1}z_1\overline{w}_2.$$

H is a Hermitian form on $\mathbb{C}^2$.

2. The function

$$H : \mathbb{C}^2 \times \mathbb{C}^2 \to \mathbb{C} \; ; \; (\underline{z}, \underline{w}) \mapsto z_1w_1 + z_2w_2,$$

is NOT a Hermitian form on $\mathbb{C}^2$: it is easy to see that

$$H\left(\begin{bmatrix}1\\ \sqrt{-1}\end{bmatrix}, \begin{bmatrix}1\\ 1\end{bmatrix}\right) = 1 + \sqrt{-1} \neq 1 - \sqrt{-1} = \overline{H\left(\begin{bmatrix}1\\ 1\end{bmatrix}, \begin{bmatrix}1\\ \sqrt{-1}\end{bmatrix}\right)}.$$

3. The function

$$H : \mathbb{C} \times \mathbb{C} \to \mathbb{C} \; ; \; (z, w) \mapsto z\overline{w},$$

is a Hermitian form on $\mathbb{C}$.

4. Let $A = a_{ij} \in Mat_n(\mathbb{C})$ be a Hermitian matrix. Then, we define

$$H_A : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C} \; ; \; (\underline{z}, \underline{w}) \mapsto \underline{z}^t A\overline{\underline{w}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}z_i\overline{w}_j.$$

H_A is a Hermitian form on $\mathbb{C}^n$. Moreover, **any Hermitian form H on $\mathbb{C}^n$ is of the form $H = H_A$, for some Hermitian matrix $A \in Mat_n(\mathbb{C})$.**

Lemma 3.4.7. *Let $H \in \mathrm{Herm}(V)$, $\mathcal{B}, \mathcal{C} \subset V$ ordered bases on V. Then, if $P = P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{C}$, then*

$$P^h[H]_{\mathcal{C}}P = [H]_{\mathcal{B}}.$$

Definition 3.4.8. Let $H \in \mathrm{Herm}(V)$. We say that H is *nondegenerate* if $[H]_{\mathcal{B}}$ is invertible, for any basis $\mathcal{B} \subset V$. The previous lemma ensures that this notion of nondegeneracy is well-defined (ie, does not depend on the choice of basis $\mathcal{B}$).[76]

Theorem 3.4.9 (Classification of Hermitian forms)**.** *Let V be a $\mathbb{C}$-vector space, $n = \dim V$ and $H \in \mathrm{Herm}(V)$ be nondegenerate. Then, there is an ordered basis $\mathcal{B} \subset V$ such that*

$$[H]_{\mathcal{B}} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}, \quad d_i \in \{1, -1\}.$$

Hence, if $u, v \in V$ with

$$[u]_{\mathcal{B}} = \begin{bmatrix}\xi_1\\ \vdots\\ \xi_n\end{bmatrix}, \; [v]_{\mathcal{B}} = \begin{bmatrix}\eta_1\\ \vdots\\ \eta_n\end{bmatrix},$$

then we have

$$H(u, v) = \sum_{i=1}^{n} d_i\xi_i\overline{\eta}_i.$$

Proof: The proof is similar to the proof of Theorem 3.2.6 and uses the following facts: for any Hermitian form $H \in \mathrm{Herm}(V)$, there exists $v \in V$ such that $H(v, v) \neq 0$; if $H \in \mathrm{Herm}(V)$ is nondegenerate then for any subspace $U \subset V$ we have $V = U \oplus U^{\perp}$. The first fact follows from an analagous 'polarisation identity' for Hermitian forms. □

[76] Note that the determinant of A^h is equal to $\overline{\det A}$: indeed, we have

$$\det(A^h) = \det(\overline{A}^t) = \det\overline{A} = \overline{\det A}.$$

Definition 3.4.10. A *Hermitian (or unitary) space* is a pair (V, H), where V is a $\mathbb{C}$-vector space and H is a Hermitian form on V such that $[H]_{\mathcal{B}} = I_n$, for some basis $\mathcal{B}$. This condition implies that H is nondegenerate.

If (V, H) is a Hermitian space and $E \subset V$ is a nonempty subset then we define the *orthogonal complement of E (with respect to H)* to be the subspace

$$E^{\perp} = \{v \in V \mid H(v, u) = 0, \text{ for every } u \in E\}.$$

We say that $z, w \in V$ are *orthogonal (with respect to H)* if $H(z, w) = 0$. We say that $E \subset V$ is *orthogonal* if $H(s, t) = 0$, for every $s \neq t \in E$.

A basis $\mathcal{B} \subset V$ is an *orthogonal basis* if $\mathcal{B}$ is an orthogonal set. A basis $\mathcal{B} \subset V$ is an *orthonormal basis* if it is an orthogonal basis and $H(b, b) = 1$, for every $b \in \mathcal{B}$.

We define $\mathbb{H}^n = (\mathbb{C}^n, H_{I_n})$, where

$$H_{I_n}(\underline{z}, \underline{w}) = z_1\overline{w}_1 + ... + z_n\overline{w}_n.$$

As in the Euclidean case we obtain the notion of a 'Hermitian morphism': a *Hermitian morphism* $f : (V, H_V) \to (W, H_W)$ is a linear morphism such that

$$H_W(f(u), f(v)) = H_V(u, v), \text{ for any } u, v \in V.$$

In particular, if (V, H) is a Hermitian space then we denote the set of all Hermitian isomorphisms of (V, H) by $U(V, H)$, or simply $U(V)$ when there is no confusion. A Hermitian isomorphism is also called a *unitary transformation of V*. Thus,

$$U(V) = \{f : V \to V \mid H(u, v) = H(f(u), f(v)), \text{ for any } u, v \in V\}.$$

We denote $U(n) = U(\mathbb{H}^n)$ and it is straightforward to verify[77] that

$$U(n) = \{T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^n) \mid A \in Mat_n(\mathbb{C}) \text{ and } A^hA = I_n\}.$$

We say that $A \in Mat_n(\mathbb{C})$ is a *unitary matrix* if

$$A^hA = I_n.$$

Thus, we can identify the set of unitary transformations of $\mathbb{H}^n$ with the set of unitary matrices. Moreover, this association is an **isomorphism of groups**.

As a consequence of Theorem 3.4.9 we can show that there is essentially only one Hermitian space of any given dimension.

Theorem 3.4.11. *Let (V, H) be a Hermitian space, $n = \dim V$. Then, there is a Hermitian isomorphism*

$$f : (V, H) \to \mathbb{H}^n.$$

Remark 3.4.12. There are generalisations to Hermitian spaces of most of the results that apply to Euclidean spaces (section 3.3). In particular, we obtain notions of length and Cauchy-Schwarz/triangle inequalities. For details see [1], section 9.2.

[77] Every linear endomorphism f of $\mathbb{C}^n$ is of the form $f = T_A$, for some $A \in Mat_n(\mathbb{C})$. Then, for f to be a Hermitian morphism we must have

$$\underline{z}^t\overline{\underline{w}} = (A\underline{z})^t\overline{A\underline{w}} = \underline{z}^tA^t\overline{A}\overline{\underline{w}}, \text{ for every } \underline{z}, \underline{w} \in \mathbb{C}^n.$$

This implies that $A^t\overline{A} = I_n$, which is equivalent to the condition $A^hA = I_n$.

3.5 The spectral theorem

In this section we will discuss the diagonalisabliity properties of morphisms in Euclidean/Hermitian spaces. The culmination of this discussion is the **spectral theorem**: this states that self-adjoint morphisms are orthogonally/unitarily diagonalisable and have real eigenvalues. This means that such morphisms are diagonalisable and, moreover, there exists an orthonormal basis of eigenvectors.

Throughout section 3.5 we will only be considering Euclidean (resp. Hermitian) spaces $(V, \langle , \rangle)$ (resp. (V, H)) and, as such, will denote such a space by V, the inner product (resp. Hermitian form) being implicitly assumed given.

First we will consider f-invariant subspaces $U \subset V$ and their orthogonal complements, for an orthogonal/unitary transformation $f : V \to V$.

Proposition 3.5.1. *Let $f : V \to V$ be an orthogonal (resp. unitary) transformation of the Euclidean (resp. Hermitian) space V and $U \subset V$ be an f-invariant subspace. Then, $U^\perp$ is f^+-invariant, where $f^+ : V \to V$ is the adjoint of f (with respect to the corresponding inner product/Hermitian form).*[78]

Proof: To say that U is $f-$-invariant means that, for every $u \in U$, $f(u) \in U$. Consider the orthogonal complement of U in V, $U^\perp$ and let $w \in U^\perp$. Then, we want to show that $f^+(w) \in U^\perp$. Now, for each $u \in U$, we have

$$H(u, f^+(w)) = H(f(u), w) = 0,$$

as $f(u) \in U$. Hence, $f^+(w) \in U^\perp$ and $U^\perp$ is f^+-invariant. □

Lemma 3.5.2. *Let (V, H) be a Hermitian space and $U \subset V$ be a subspace. Then, the restriction of H to U is nondegenerate.*

Proof: Suppose that $v \in U$ is such that $H(u, v) = 0$, for every $u \in U$. Then, $V = U \oplus U^\perp$ (as H is nondegenerate). Hence, if $w \in V$ then $w = u + z$, with $u \in U, z \in U^\perp$ and

$$H(w, v) = H(u + z, v) = H(u, v) + H(z, v) = 0 + 0 = 0.$$

Hence, using nondegeneracy of H on V we have $v = 0_V$ and the restriction of H to U is nondegenerate. □

3.5.1 Normal morphisms

Throughout this section we will assume that V is a Hermitian space, equipped with the Hermitian form H. The results all hold for Euclidean spaces with appropriate modifications to statements of results and to proofs.[79]

Definition 3.5.3 (Normal morphism)**.** Let V be a Hermitian space. We say that $f : V \to V$ is a *normal morphism* if we have

$$f \circ f^+ = f^+ \circ f.$$

[78] Given a linear morphism $f : V \to V$, where (V, H) is a Hermitian space, we define the *adjoint of f* to be the morphism

$$f^+ = \sigma_H^{-1} \circ f^* \circ \sigma_H : V \to V,$$

where

$$\sigma_H : V \to V^* \ ; \ v \mapsto \sigma_H(v), \text{ so that } (\sigma_H(v))(u) = H(u, v).$$

It is important to note that σ_H **is NOT $\mathbb{C}$-linear**: we have $\sigma_H(\lambda v) = \overline{\lambda}\sigma_H(v)$, for any $\lambda \in \mathbb{C}$. However, the composition $\sigma_H^{-1} \circ f^* \circ \sigma_H$ **IS linear** (check this). The definition of f^+ implies that, for every $u, v \in V$, we have

$$H(f(u), v) = H(u, f^+(v));$$

moreover, f^+ is the unique morphism such that this property holds.

As a result of the nonlinearity of σ_H we **DO NOT** have a nice formula for the matrix of f^+ in general. However, if $V = \mathbb{H}^n$ and $f = T_A \in \mathrm{End}_{\mathbb{C}}(V)$, where $A \in Mat_n(\mathbb{C})$, then $f^+ = T_{A^h}$: indeed, for any $\underline{z}, \underline{w} \in \mathbb{C}^n$ we have

$$H(A\underline{z}, \underline{w}) = (A\underline{z})^t \overline{\underline{w}} = \underline{z}^t A^t \overline{\underline{w}} = \underline{z}^t \overline{A^h \underline{w}} = H(\underline{z}, A^h \underline{w}).$$

[79] We could consider a Euclidean space as being a **real Hermitian space**, since $x = \overline{x}$, for every $x \in \mathbb{R}$.

Example 3.5.4. Let V be a Hermitian (resp. Euclidean) space. Then, unitary (resp. orthogonal) transformations of V are normal.

However, not all normal morphisms are unitary/orthogonal transformations: for example, the morphism $T_A \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^3)$ defined by the matrix

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

is normal but does not define a unitary transformation of $\mathbb{H}^3$ (as $A^h A \neq I_3$).

Normal morphisms possess useful orthogonality properties of their eigenvectors.

Lemma 3.5.5. *Let $f : V \to V$ be a normal morphism of the Hermitian space (V, H), $f^+ : V \to V$ the adjoint of f (with respect to H). If $v \in V$ is an eigenvector of f with associated eigenvalue $\lambda \in \mathbb{C}$ then v is an eigenvector of f^+ with associated eigenvalue $\overline{\lambda} \in \mathbb{C}$.*

Proof: First, we claim that E_λ (the λ-eigenspace of f) is f^+-invariant: indeed, for any $u \in E_\lambda$ we want to show that $f^+(u) \in E_\lambda$. Then,

$$f(f^+(u)) = f^+(f(u)) = f^+(\lambda u) = \lambda f^+(u),$$

so that $f^+(u) \in E_\lambda$. Hence, f^+ defines an endomorphism of E_λ. Now, let $v \in E_\lambda$ be nonzero (so that $v \in V$ is an eigenvector of f with associated eigenvalue λ). Then, for any $u \in E_\lambda$ we have

$$H(u, f^+(v)) = H(f(u), v) = H(\lambda u, v) = H(u, \overline{\lambda} v) \implies H(u, f^+(v) - \overline{\lambda} v) = 0, \text{ for every } u \in E_\lambda.$$

Then, by Lemma 3.5.2 we see that

$$f^+(v) - \overline{\lambda} v = 0_V \implies f^+(v) = \overline{\lambda} v,$$

and the result follows. □

Lemma 3.5.6. *Let $f : V \to V$ be a normal morphism of the Hermitian space V. Then, if $v_1, \dots, v_k \in V$ are eigenvectors of f corresponding to* distinct *eigenvectors $\xi_1, \dots, \xi_k$ (so that $\xi_i \neq \xi_j$, $i \neq j$), then $\{v_1, \dots, v_k\}$ is orthogonal.*

Proof: Consider v_i, v_j with $i \neq j$. Then, we have $f(v_i) = \xi_i v_i$ and $f(v_j) = \xi_j v_j$ as v_i, v_j are eigenvectors. Then,

$$\xi_i H(v_i, v_j) = H(\xi_i v_i, v_j) = H(f(v_i), v_j) = H(v_i, f^+(v_j)) = H(v_i, \overline{\xi}_j v_j) = \xi_j H(v_i, v_j),$$

so that

$$(\xi_i - \xi_j) H(v_i, v_j) = 0 \implies H(v_i, v_j) = 0, \text{ since } \xi_i \neq \xi_j.$$

□

Theorem 3.5.7 (Normal morphisms are orthogonally diagonalisable)**.** *Let (V, H) be a Hermitian space, $f : V \to V$ a normal morphism. Then, there exists an orthonormal basis of V consisting of eigenvectors of f.*

Proof: Since V is a $\mathbb{C}$-vector space we can find an eigenvector $v \in V$ of f with associated eigenvalue $\lambda \in \mathbb{C}$ (as there is always a root of the characteristic polynomial χ_f). Let $E_\lambda \subset V$ be the corresponding λ-eigenspace (so that $E_\lambda \neq \{0_V\}$). Consider the orthogonal complement $E_\lambda^\perp$ of E_λ (with respect to H). Then, since H is nondegenerate we have

$$V = E_\lambda \oplus E_\lambda^\perp.^{80}$$

We are going to show that $E_\lambda^\perp$ is f-invariant: let $w \in E_\lambda^\perp$, so that for every $v \in E_\lambda$ we have

$$H(u, v) = 0.$$

[80] You can check that $E_\lambda \cap E_\lambda^\perp = \{0_V\}$.

We want to show that $f(w) \in E_\lambda^\perp$. Let $u \in E_\lambda$. Then, using Lemma 3.5.5, we obtain

$$H(f(w), u) = H(w, f^+(u)) = H(w, \overline{\lambda}u) = \lambda H(w, u) = 0.$$

Hence, $f(w) \in E_\lambda^\perp$ and $E_\lambda^\perp$ is f-invariant.

So, we have that $E_\lambda^\perp$ is both f-invariant and f^+-invariant (Proposition 3.5.1) and so f and f^+ define endomorphisms of $E_\lambda^\perp$. Moreover, we see that the restriction of f to $E_\lambda^\perp$ is normal. Hence, we can use an induction argument on $\dim V$ and assume that there exists an orthonormal basis of $E_\lambda^\perp$ consisting of eigenvectors of f, $\mathcal{B}_1$ say. Using the Gram-Schmidt process we can obtain an orthonormal basis of E_λ, $\mathcal{B}_2$ say. Then, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is an orthonormal basis (Lemma 3.5.6) and consists of eigenvectors of f. □

Corollary 3.5.8. *1. Let $A \in Mat_n(\mathbb{C})$ be such that*

$$AA^h = A^h A.$$

Then, there exists a unitary matrix $P \in U(n)$ (ie, $P^{-1} = P^h$) such that

$$P^h AP = D,$$

where D is a diagonal matrix.

Remark 3.5.9. Suppose that $A \in Mat_n(\mathbb{R})$. Then, we have

$$A^h = A^t,$$

so that the condition

$$A^h A = AA^h \implies A^t A = AA^t.$$

Thus, if $A^t A = AA^t$ then Corollary 3.5.8 implies that A is diagonalisable. However, it is not necessarily true that there exists $P \in \mathrm{GL}_n(\mathbb{R})$ such that

$$P^{-1}AP = D,$$

with $D \in Mat_n(\mathbb{R})$. For example, consider the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in Mat_2(\mathbb{R}).$$

Then,

$$A^t A = I_2 = AA^t,$$

so that A is normal. Then, Corollary 3.5.8 implies that we can diagonalise A. However, the eigenvalues of A are $\pm\sqrt{-1}$ so that we must have

$$P^{-1}AP = \pm \begin{bmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{bmatrix},$$

so that it is not possible that $P \in \mathrm{GL}_2(\mathbb{R})$.[81]

3.5.2 Self-adjoint operators and the spectral theorem

Definition 3.5.10. Let V be a Hermitian space. We say that a morphism $f \in \mathrm{End}_{\mathbb{C}}(V)$ is *self-adjoint* if $f = f^+$. **Self-adjoint morphisms are normal morphisms**.

Example 3.5.11. Let V be a Hermitian (resp. Euclidean) space. Then, $T_A \in \mathrm{End}(V)$ is self-adjoint if and only if A is Hermitian (resp. symmetric).

[81] Why?

Lemma 3.5.12. *Let V be a Hermitian space, $f \in \mathrm{End}_{\mathbb{C}}(V)$ a self-adjoint morphism. Then, all eigenvalues of f are real numbers.*

Proof: As f is self-adjoint then f is normal. Using Lemma 3.5.5 we know that if $v \in V$ is an eigenvector of f with associated eigenvalue $\lambda \in \mathbb{C}$, then $v \in V$ is an eigenvector of f^+ with associated eigenvalue $\overline{\lambda} \in \mathbb{C}$. As $f = f^+$ we must have that $\lambda = \overline{\lambda}$, which implies that $\lambda \in \mathbb{R}$. □

Since a self-adjoint morphism f is normal (indeed, we have $f \circ f^+ = f \circ f = f^+ \circ f$), then Theorem 3.5.7 implies that V admits an orthonormal basis consisting of eigenvectors of f. This result is commonly referred to as **The Spectral Theorem**.

Theorem 3.5.13 (Spectral theorem)**.** *Let V be a Hermitian space, $f \in \mathrm{End}_{\mathbb{C}}(V)$ a self-adjoint morphism. Then, there exists an orthonormal basis $\mathcal{B}$ of V consisting of eigenvectors of f and such that*

$$[f]_{\mathcal{B}} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \in Mat_n(\mathbb{R}).$$

Corollary 3.5.14. *1. Let $A \in Mat_n(\mathbb{C})$ be Hermitian ($A^h = A$). Then, there exists a unitary matrix $P \in U(n)$ such that*

$$P^h A P = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}, \text{ where } d_1, \dots, d_n \in \mathbb{R}.$$

2. Let $A \in Mat_n(\mathbb{R})$ be symmetric ($A^t = A$). Then, there exists an orthogonal matrix $P \in O(n)$ such that

$$P^t A P = D,$$

where D is diagonal.

Example 3.5.15. 1. Consider the matrix

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

Then, $A^t = A$ so that there exists $P \in O(3)$ such that $P^t A P$ is diagonal (Theorem 3.5.13).

How do we determine P? We know that A is diagonalisable so we proceed as usual: we find that

$$\chi_A(\lambda) = (1 - \lambda)(\lambda - \sqrt{3})(\lambda + \sqrt{3}).$$

Then, if we choose eigenvectors $v_1 \in E_1$, $v_2 \in E_{-\sqrt{3}}$, $v_3 \in E_{\sqrt{3}}$ such that $||v_i|| = 1$, then we have

$$P = [v_1 \ v_2 \ v_3] \in O(3).$$

For example, we can take

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6-2\sqrt{3}}} & \frac{1}{\sqrt{6+2\sqrt{3}}} \\ 0 & \frac{1-\sqrt{3}}{\sqrt{6-2\sqrt{3}}} & \frac{1+\sqrt{3}}{\sqrt{6+2\sqrt{3}}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6-2\sqrt{3}}} & \frac{-1}{\sqrt{6+2\sqrt{3}}} \end{bmatrix} \in O(3)$$

2. Consider the matrix

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & -1-\sqrt{-1} \\ 0 & -1+\sqrt{-1} & 1 \end{bmatrix}.$$

Then, $A = A^h$ so that A is Hermitian. Hence, there exists $P \in U(3)$ such that

$$P^h A P = \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix}, \ d_1, d_2, d_3 \in \mathbb{R}.$$

We first determine

$$\chi_A(\lambda) = -(1+\lambda)^2(\lambda - 2),$$

so that the eigenvalues are $\lambda_1 = -1, \lambda_2 = 2$. Then,

$$E_{-1} = \text{span}_{\mathbb{C}} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 - \sqrt{-1} \\ -2 \end{bmatrix} \right\}.$$

Since

$$H_{I_3}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 - \sqrt{-1} \\ -2 \end{bmatrix} \right) = 1.0 + 0.(-1 + \sqrt{-1}) + 0.(-2) = 0,$$

we have that

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 - \sqrt{-1} \\ -2 \end{bmatrix} \right) = (v_1, v_2)$$

is an orthogonal basis of E_{-1}. In order to obtain an orthonormal basis we must scale v_1, v_2 by $H_{I_3}(v_i, v_i)$. Hence, as

$$H_{I_3}(v_1, v_1) = 1, \ H_{I_3}(v_2, v_2) = 0.0 + (-1 - \sqrt{-1})(-1 + \sqrt{-1}) + (-2).(-2) = 2 + 4 = 6,$$

we have that

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 0 \\ -1 - \sqrt{-1} \\ -2 \end{bmatrix} \right)$$

is an orthonormal basis of E_{-1}.

Now, we need only determine a vector $v_3 \in E_2$ for which $H_{I_3}(v_3, v_3) = 1$: such an example is

$$v_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 0 \\ -1 - \sqrt{-1} \\ -1 \end{bmatrix}.$$

Hence, if we set

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{-1}{\sqrt{6}} - \sqrt{\frac{-1}{6}} & \frac{-1}{\sqrt{3}} - \sqrt{\frac{-1}{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \end{bmatrix},$$

then

$$P^h A P = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 2 \end{bmatrix}.$$

3. Consider the matrix

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

As $A = A^t$ we can find $P \in O(3)$ such that

$$P^t A P = D,$$

where D is diagonal. We have that

$$\chi_A(\lambda) = -(1-\lambda)^2(\lambda - 4),$$

so that the eigenvalues of A are $\lambda_1 = 1, \lambda_2 = 4$.

We have that

$$E_1 = \text{span}_{\mathbb{R}} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\},$$

where

$$\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right),$$

is a basis of E_1. Using the Gram-Schmidt process we can obtain an orthonormal basis

$$\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right) \subset E_1.$$

Now, we need to find $v_3 \in E_4$ such that $||v_3|| = 1$: we can take

$$v_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then, if we let

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix},$$

then $P \in O(3)$ and

$$P^t A P = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 4 \end{bmatrix}.$$

References

[1] Shilov, Georgi E., *Linear Algebra*, Dover Publications 1977.